

## ANNEX J: DATA QUALITY ASSESSMENT (DQA)

### Contents

1. Data Quality Assessment for Guatemala.....	2
1.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – GUATEMALA .....	2
1.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – GUATEMALA .....	7
1.3. HOUSEHOLD LABOR MARKET SURVEYS - GUATEMALA (ENCUESTA NACIONAL DE EMPLEO E INGRESOS - ENEI).....	11
2. Data Quality Assessment for El Salvador .....	14
2.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – EL SALVADOR.....	14
2.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – EL SALVADOR.....	18
2.3. HOUSEHOLD LABOR MARKET SURVEYS – EL SALVADOR (ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES - EHPM).....	22
3. Data Quality Assessment for Honduras.....	26
3.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – HONDURAS.....	26
3.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – HONDURAS.....	30
3.3. HOUSEHOLD LABOR MARKET SURVEYS – HONDURAS (ENCUESTA PERMANENTE DE HOGARES DE PROPÓSITOS MÚLTIPLES - EPHPM).....	34

# 1. Data Quality Assessment for Guatemala

## 1.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – GUATEMALA

Data collection for Wave 1 of the Pilot Establishment Surveys in Guatemala occurred between September and October 2019. For Wave 1, the hospitality industry was the target sector for data collection. Survey administration was tablet-based.

### SUMMARY

	OVERALL QUALITY	
What is the overall conclusion regarding the quality of the data?	VALIDITY	
<hr/> <i>Somewhat High Quality</i> <hr/>		
	RELIABILITY	
	TIMELINESS	
	PRECISION	
	INTEGRITY	

### DQA QUESTION      ADDITIONAL DETAILS

VALIDITY – Data should represent the intended result clearly and adequately.

**Was the sampling methodology appropriately followed?**

Yes	The sampling strategy established a sample frame from a directory of hotels from the Guatemalan Institute of Tourism (INGUAT). The frame was stratified by region and hotel price range. 100% of “high price” hotels, 96% of “mid-price”, and 80% of “low price” hotels were selected for a final sample size of 1,043. While it is clear this was followed, some aspects of the sampling methodology could be stated more clearly in the survey documents:
-----	---

- Says “sample is composed of all hotels classified as ‘recommended’ in the INGUAT directory and a portion of hotels classified as ‘not recommended’”. Unclear if this was explicitly used in sampling strategy though, or how.
- Would be helpful for the report to show the selected sample by strata (report only shows the sample frame and survey achievement by strata).

### Is the response rate high enough so that we can be confident the data is reflective of the selected sample?

Yes

- 2019 response rate: 77%

NORC suggests a minimum target of 60% response rate for establishment surveys. The response rate here compares favorably to existing establishment surveys in the United States, such as the OEWS (70.9% - June 2019), ARS (76.7% - June 2019), ORS (70.6% - Dec 2019), and the CESS (60.0% - Oct 2019).<sup>1, 2</sup>

### Do results collected fall within a plausible range?

No

NORC examined the dataset to ensure responses fall within valid and plausible ranges. Issues encountered were:

- *ciiu\_principal*: 4 observations have “0” as their main activity
- *salario\_trabajo*: Response option "Igual o mayor a 5.000 quetzales" is inconsistent with questionnaire and codebook.
- *vacantes\_cantidad*: 1 observation where number of vacancies is “0”, despite saying that there were vacancies.
- *subcontratados\_establecimiento*: responses are “NA” or missing for 77% of observations, but it appears the question should apply to all observations.
- *curso\_capacitacion\_copy*: a large share of responses should be missing, but contain values like “1”, “2”, “5”, or “7”, which are not valid. All come from *curso\_capacitacion\_copy\_6\_recode*.
- *horas\_habituales\_copy*: one observation contains an invalid response, “1”, when it should be missing.
- *inguat5*: responses are “NA” for 1% of observations, but appears the question should apply to all. Question also differs in format from questionnaire, which asks this question separately for high and low seasons.
- *remuneracion*: Some wages appear too low to be plausible. 6% of responses are 6 quetzales or less (approx. USD \$0.75). Some others also appear too low, and well below Guatemala’s minimum wage, but since the question does not specify a period for the wage (e.g., hourly, daily, monthly, annual), it is impossible to tell. Without specifying a time period for the question, the data is difficult or impossible for an outside party to use.

RELIABILITY – Data should reflect stable and consistent data-collection processes and analysis methods over time.

### Did the data collection produce responses that are internally consistent?

Yes

- NORC created 8 internal consistency checks. On average, there were 0.24 consistency problems per survey.
  - 77.8% had no problems raised.

<sup>1</sup> OEWS = Occupational Employment and Wage Statistics Survey, ARS = Annual Refiling Survey, ORS = Occupational Requirements Survey, CESS = Current Employment Statistics Survey

<sup>2</sup> <https://www.bls.gov/osmr/response-rates/>

- 20.9% had just one failed consistency check
- 1.2% had two or three failed consistency checks
- Checks with the greatest share of interviews with problems were:
  - Business has no kitchen staff reported, but restaurant services are reported as one of its main economic activities (11.3%)
  - Average salaries reported for operational staff are higher than salaries reported for managers (5.1%)
  - Total number of staff reported does not equal the number of staff in disaggregated categories (e.g., by gender) (2.2%)

Results suggest a high degree of internal consistency in the data.

**Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?**

- No
- Project materials contain summaries of the sampling methodology, but it is not presented in sufficient detail to run a new survey with the same methodology, and descriptions in different documents conflict with each other.
    - The Final Report says the sample was stratified by region and price range (21 strata), while the manual says the sample was stratified by price range only (3 strata).
    - Final Report and Manual say all hotels classified as “recommended” by INGUAT were included, along with a fraction of hotels “not recommended.” It is not clear if this represents a separate level of stratification, how the fraction of “not recommended” hotels were selected, and in what proportion these hotels were selected.
  - The organization of the dataset is insufficient to ensure it can be used reliably for reproducible analysis. Dataset must be compared against the codebook to find question and value labels, though the variables are out of order compared to the questionnaire, and the codebook is also out of order. Variable labels in the codebook are phrased differently from the questionnaire, and neither the dataset nor codebook include the question tags (e.g., “C01”). All of this makes the dataset difficult to work with reliably.

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.

**Is the data collected within the range of time anticipated by the survey methodology?**

- Yes
- Data was collected between September and October 2019. Data collection period is reasonable to capture a snapshot of a single moment in time, without worrying about the comparability of data collected at different times, and is consistent with the time period envisioned in the methodology.
  - The dataset itself does not include a variable to show the date of the interview.

**PRECISION** – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest “5” or “10”)?**

- No
- NORC looked for evidence of rounding total number of workers, number of guests, percentage of guests that are foreigners, number of workers in individual jobs, number of job vacancies, number of training hours, and employee pay. The level of rounding is generally appropriate for these variables, with two important exceptions:
- Evidence of rounding number of guests: **yes**

- 30.1% of responses are multiples of 100.
- Evidence of rounding employee pay: **yes**
  - 36.4% of responses are multiples of 1000 quetzales (approx. \$130 USD). Assuming rounding to 100 quetzales is acceptable, this is considerably higher than the expected 10%, and is likely too coarse to detect changes over time, given mean salaries of approximately 3000 quetzales.

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

- NA
- The sampling strategy established a sample frame from a directory of hotels from the Guatemalan Institute of Tourism (INGUAT), limited the geographic scope to 7 departments of the country, and included only hotels with at least 2 workers, for a total sample frame of 1,251 hotels.
  - The frame was stratified by region and hotel price range. 100% of “high price” hotels, 96% of “mid-price”, and 80% of “low price” hotels were selected for a final sample size of 1,043.
  - The sampling method didn’t include any power calculations.
  - Weights were applied to the final analysis to adjust for the representativeness of the effective sample.

Although no power calculations are discussed in the methodology, sample size was determined by the limits of the sample frame. Therefore, this point is not applicable.

**Does the collected data allow for disaggregated analysis for underserved populations?**

- No
- Collected data allows for disaggregating most information on employees by gender and age group, and disaggregating some by level of education.
  - It is important to note the data does not allow for disaggregation by ethnic or linguistic minority groups or rural/urban.

**INTEGRITY** – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**

- Yes
- Data was collected by tablet, minimizing the risk of transcription errors.
  - After completing an initial interview, the data collection included a verification step, where each survey was reviewed for completeness and accurate recording of answers, including consistency of responses.
  - According to the Methodology Protocol document, specific quality control checks performed included checks for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. This document also states that backcheck interviews were conducted on 25% of interviews, where original interview responses were compared against the responses in validation backcheck interviews to detect discrepancies. The checks described here are established best practices for quality control safeguards.

**RECOMMENDATIONS**

- The current CSV-formatted data must be compared against the codebook to understand the meaning of individual variables. Many users will read the questionnaire and then look for specific variables in the dataset – this is very difficult in the dataset’s current format, because variables are out of order from the questionnaire, the codebook uses different phrasing than appears in the questionnaire, and question tags (e.g., “C01”) are absent. At a minimum, variables in the datasets should be presented in the same order as they appear in the questionnaire, and the codebook should include the question tags from the questionnaire.
  - Findings around the internal consistency of survey responses suggest internal consistency was widely maintained. However, results also suggest that these checks were not programmed directly into the survey logic. Future surveys should consider internal consistency and program hard and soft checks into the survey logic (e.g., to generate an error message if the number of male and female employees is greater than the total number of reported employees).
  - The survey should employ similar programming to check for implausible responses, such as wages too low to be plausible. The number of questions with values outside the range of permitted answers is concerning and suggests appropriate checks were not programmed into the survey.
  - The survey must specify for a time period for questions sensitive to time, including wage payment (e.g., hourly, daily, weekly, annually), for these data to be usable.
  - The survey asks, for each level of employment (e.g., upper managers, middle managers, operations) how many of the employees speak English, providing some indication of the extent to which establishments have oriented their services to foreigners. As part of ILAB’s focus on underserved populations, the survey could also ask about how many employees speak an Indigenous Guatemalan language. This would provide some indication of the extent to which businesses are employing underserved Indigenous populations, as well as providing services that cater to these populations.
  - For questions at high risk for rounding answers (e.g., average salary), enumerators should be trained to identify the appropriate level of detail responses are expected at, and to probe for more specific answers if the respondent provides an answer that appears to be rounded to a higher level. These questions should be monitored by the data manager during data collection to detect rounding issues and follow up with enumerators.
-

## 1.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – GUATEMALA

The data collection period for Wave 2 of the Pilot Establishment Surveys in Guatemala is not specified in survey documents. For Wave 2, the food and beverage processing industry was the target sector for data collection. Survey administration was phone-based.

### SUMMARY

What is the overall conclusion regarding the quality of the data?

*Mostly High Quality*

OVERALL QUALITY



VALIDITY



RELIABILITY



TIMELINESS



PRECISION



INTEGRITY



### DQA QUESTION

### ADDITIONAL DETAILS

**VALIDITY** – Data should represent the intended result clearly and adequately.

**Was the sampling methodology appropriately followed?**

Yes

The sampling strategy established a sample frame from three directories of companies in the food and beverage processing sector, considered the entire country of Guatemala in its geographic scope, and included only food and beverage processing companies that operated between July 2020 to July 2021 with at least 2 workers, for a frame of 583 establishments. Enumerators attempted to contact each establishment via telephone with the purpose of confirming the commercial name and contact information. The records of 328 establishments could not be confirmed and were discarded. Enumerators went on to screen the remaining 255 establishments in the field on the requisites and found 149 did not pass eligibility requirements. The final sample of responding establishments who passed the screener was 106.

**Is the response rate high enough so that we can be confident the data is reflective of the selected sample?**

No

- 2021 response rate: 48%

NORC suggests a minimum target of 60% response rate for establishment surveys. The response rate here is significantly lower. Therefore, we cannot confidently say that the data is reflective of the selected sample.

**Do results collected fall within a plausible range?**

Yes

NORC examined the dataset to ensure responses fall within valid and plausible ranges. Issues encountered were:

- *ano\_inicial*: 1 observation where the year the company was founded is “NA”.
- *existe\_temporada*: 1 response was “NA” although it appears the question should apply to all observations

This suggests appropriate safeguards were in place to ensure response validity.

**RELIABILITY – Data should reflect stable and consistent data-collection processes and analysis methods over time.****Did the data collection produce responses that are internally consistent?**

Yes

- NORC created 7 internal consistency checks. On average, there were 0.08 consistency problems per survey.
  - 92.5% had no problems raised.
  - 6.6% had just one failed consistency check
  - 0.9% had two or three failed consistency checks
- Checks with the greatest share of interviews with problems were:
  - Business reported no machine equipment, but employs machine operators (6.6%)
  - The number of employees trained is greater than the number of employees (1.9%)

Results suggest a high degree of internal consistency in the data.

**Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?**

Yes

- Project materials present sufficient summaries of the sampling methodology to replicate a new survey with the same methodology. Other methods, including survey software as well as contact methods and protocols are clearly detailed.
- The organization of the dataset is sufficient to ensure it can be used reliably for reproducible analysis. The codebook includes question tags which ensure variables in the codebook can be appropriately linked to questions in the questionnaire. Variables in the Wave 2 dataset are ordered appropriately in relation to the questionnaire and codebook.



- One important exception to the above: survey weights are required to analyze the data but are not included in the dataset. This requires the user to find the weights in the report, and manually incorporate them into the dataset. This generates a high risk for error.

**TIMELINESS – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.**

**Is the data collected within the range of time anticipated by the survey methodology?**

No

- The Methodology Protocol document does not include the data collection period, and the data itself does not include a variable to show the date of the interview. As such, we are unable to estimate whether the data collection period is sufficient or whether it was the time period envisioned by the methodology. Information about the data collection period must be easily identifiable in survey documents for the data to be useful to outside users.

**PRECISION – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.**

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest “5” or “10”)?**

Yes

- NORC looked for evidence of rounding total number of workers, number of workers who completed the training courses, percentage of production workers, percentage of sales for the main product, and number of workers in individual jobs. NORC found only limited evidence of systematic rounding of responses:
- Evidence of rounding total number of workers: yes
    - 39.6% of responses are multiples of 10

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

NA

Although no power calculations are discussed in the methodology, sample size was determined mostly by the limits of the sample frame. Therefore, this point is not applicable.

**Does the collected data allow for disaggregated analysis for underserved populations?**

No

- Collected data allows for disaggregating most information on employees by gender and age group, and disaggregating some by level of education.
- It is important to note the data does not allow for disaggregation by ethnic or linguistic minority groups or rural/urban.

**INTEGRITY – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.**

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**

Yes

- Data was collected by tablet, minimizing the risk of transcription errors.

According to the Methodology Protocol document, there was a quality control check following survey implementation, where the survey team will review 100% of the questionnaires to verify the survey is filled out appropriately. In the cases where errors are detected, the surveyors called the informant to retrieve the information prior to updating the complete survey record.

**RECOMMENDATIONS**

- Survey weights must be included in the dataset. It is not sufficient to assume data users will find these weights in the methodology description and manually incorporate them.
- Survey documentation should include the dates during which data was collected.
- Although NORC considers the quality checks described in survey materials to be minimally sufficient, these materials do not describe other best practices. NORC recommends data collection include a backcheck of a random selection of interviews, and including checks on the main data for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate.
- Levels of rounding of responses are generally appropriate, with some exceptions. For open-ended numeric responses, enumerators should be trained to know the appropriate level of detail to look for, and to probe further if respondents appear to be rounding at a higher level than desired.
- The dataset could include a variable to show the date of each survey interview. At minimum, the survey methodology document should include the range of time in which the survey data was collected to better understand the timeliness of the data collection process.

In addition to the demographic variables offering the age range and gender of employees, it is also valuable to allow disaggregation by Indigenous status, as well as the demographic composition of the establishment's leadership and senior management.

### 1.3. HOUSEHOLD LABOR FORCE SURVEYS - GUATEMALA (ENCUESTA NACIONAL DE EMPLEO E INGRESOS - ENEI)

The data from the Encuesta Nacional de Empleo e Ingresos (ENEI) is publicly available on a semi-annual basis, conducted by Guatemala's national statistics agency, Instituto Nacional de Estadística (INE). The most recent iteration of the survey available at the time of writing was September/October 2021, with a sample size of 5,790 households across 593 primary sampling units.

#### SUMMARY<sup>3</sup>

What is the overall conclusion regarding the quality of the data?

*Somewhat High Quality*

OVERALL QUALITY 

TIMELINESS 

PRECISION 

USABILITY 

#### DQA QUESTION

#### ADDITIONAL DETAILS

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence policymaking decisions.

**Are data available frequently enough to inform policymaking decisions?**

- No
- ENEI data is publicly available at semi-annual frequency from 2013-2019, with the exception of 2016 and 2017, when data was collected with trimestral frequency. It is available annually for 2010-2012.
  - Data was not collected in 2020 or the first half of 2021, owing to COVID-19. The datasets and results from the September/October 2021 survey were not posted until the end of February 2022, suggesting a lag of at least 4 months between data collection and publication.

<sup>3</sup> While the DQA evaluation of the establishment surveys rates five data quality indicators, the evaluation of the national household labor market surveys rates three data quality indicators. This decision was made following ILAB's interests and learning priorities for understanding different aspects of the data quality for each survey.

- For comparable surveys among other countries in the region: Colombia (*Gran Encuesta Integrada de Hogares*), conducts surveys monthly, with data posted publicly with a lag of approximately 2 months; Costa Rica (*Encuesta Continua de Empleo*), conducts surveys quarterly, with data reported with a lag of approximately 2 months.
- Frequency of data availability for ENEI is sufficient to inform basic policy and investment decisions, particularly those that are not especially sensitive to small changes from month to month. The frequency may be insufficient to guide policymakers through turbulent events that result in large swings from one month to the next, or that require close monitoring of small changes.
- There has not been any sustained improvement in the frequency of data collection since 2012. Importantly, while the frequency of data collection has continued at a fairly consistent semi-annual pace since 2013, the months data is collected change from year to year. This makes year-over-year comparisons difficult, since employment figures in October may not be comparable to those in December, owing to seasonal changes.
- It is important to note that while the timeliness of data collection and reporting are outside high quality standards, these elements were outside of the NTLMI program's scope.

**PRECISION** – Data have a sufficient level of detail to permit informed policymaking decisions.

**Has the margin of error been reported along with the data?**

Yes      Although the margin of error is not provided directly in the reports or survey materials, this is acceptable given how the materials are presented. The methodological description of the survey and calculations for precision are sufficient to allow a user to calculate this on their own, given the data.

**Are the margins of error acceptable for program/policy decision making?**

No      
 

- The sample of the Oct 2021 survey included 5,790 households across 593 primary sampling units.
- For comparison, Costa Rica's Encuesta Continua de Empleo includes a sample of nearly 10,000 households across 800 primary sampling units.
- While Guatemala's sample size is adequately powered for most outcomes at the national or regional levels, it is likely to fall short for sub-populations of interest to policymakers, such as measuring changes in specific economic sectors, minority groups, or smaller geographic levels.
- It is important to note that while the project provided some assistance on sampling methods, increasing the sample size of the survey to ensure it was adequately powered for sub-populations of interest was outside the NTLMI project's scope.

**Does the collected data allow for disaggregated analysis for traditionally underserved populations?**

Yes      
 

- The data allow results to be disaggregated by household socioeconomic characteristics, age group, location (rural/urban), gender, level of education, and indigenous identity.

**USABILITY** – The ease with which an outside user can access, understand, and use the data to inform policy, business, recruitment, job training, and talent acquisition decisions.

**Is the data easy for outside users to access?**

- Yes
- Data is easy to find and download online and is publicly available. The website is well-organized, with functioning links and clear labels.

**Is the available data well-documented, labelled, properly formatted, and easy enough to understand to be usable for outside users?**

- Yes
- Since 2010, data is available with clear labels for variables and values. Accompanying codebooks, data dictionaries, and questionnaires facilitate the process of identifying survey questions in the data to fully understand question context and the meaning of responses.
  - There is some variation from wave-to-wave regarding exactly what is available. For example, questionnaires and codebooks are not available for all survey waves. However, because the structure of datasets does not vary much from one wave to the next, what is available is sufficient for use by outside users.

**Does the data have wide enough geographic coverage to inform policy making decisions?**

- Yes
- The sample frame for ENEI includes all rural and urban communities across all 334 municipalities in the country.
  - The design ensures a sample that is representative of the population, with sufficient allocation of the sample across socioeconomic sectors and location to permit calculating figures for the national capital area, other urban areas, and rural.

**RECOMMENDATIONS**

- Consistency in the months data is collected in will ensure year-over-year comparability, eliminating uncertainty over whether changes over time are due to structural or seasonal changes.
- The INE has already shown some capacity to conduct these surveys with trimestral frequency, as it did in 2016 and 2017, but it does not appear to have been able to sustain this. Resuming and sustaining data collection at least at trimestral frequency is a reasonable near-term goal, and would move Guatemala closer to Costa Rica, which collects this data at quarterly frequency.
- For this data to better reach its potential value, INE should aim for lags of no more than two months between the end of data collection and when results are reported and the data are published.
- Moving forward, INE should ensure a codebook, data dictionary, and questionnaire is provided for each round of the survey. These accompanying materials are currently available for some, but not all survey waves.

## 2. Data Quality Assessment for El Salvador

### 2.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – EL SALVADOR

The data collection period for Wave 1 of the Pilot Establishment Surveys in El Salvador was from October to December 2019. For Wave 1, the hospitality industry was the target sector for data collection. Survey administration was phone-based, with responses recorded via tablet.

#### SUMMARY

What is the overall conclusion regarding the quality of the data?

*Somewhat High Quality*

OVERALL QUALITY



VALIDITY



RELIABILITY



TIMELINESS



PRECISION



INTEGRITY



#### DQA QUESTION

#### ADDITIONAL DETAILS

VALIDITY – Data should represent the intended result clearly and adequately.

Was the sampling methodology appropriately followed?

No

The sampling frame was constructed using the Registry of Members of the Salvadoran Chamber of Tourism (CASASUR), with additional samples from the “Inventory 2018”, the “National Registry of Tourism (2016)”, as well as online search using Booking, Google Travel, and Kayak.com. As a result, the final sample included 407 hotels and hostels, 89 restaurants, and 27 tour operators, across 14 departments.

In the end, a total of 180 hotels, 18 tourist operators, and 35 restaurants were successfully interviewed. The table below summarizes the distribution of different types of establishments in the planned sample and effective sample.

Type	Planned Sample (n = 523)	Effective Sample (n = 233)
Hotels and hostels	77.8%	77.3%
Restaurants	17%	8%
Tour operators	5%	15%

The inadequacy of the originally intended sample frame (CASASUR) was outside the program's control. The decision to supplement this frame with the results of an online search, rather than resort to sampling geographic units and conducting a listing exercise, was likely due to budgetary constraints but does have implications for the survey's external validity. Likewise, the survey's internal validity is compromised because no adjustments have been made for non-response. In summary, the sampling strategy is clearly stated, but the effective sample is not strictly representative of the intended population.

#### Is the response rate high enough so that we can be confident the data is reflective of the selected sample?

No The overall response rate is 44.6%. Separated by the establishment type, the response rates are:

- Hotels and Hostels: 44.2%
- Restaurants: 39.3%
- Tourist operators: 66.7%

NORC suggests a minimum target of 60% response rate for establishment surveys. The response rates here are significantly lower for hotels and restaurants, and comparable in the tourist operator sector. Therefore, we can't confidently say that the data is reflective of the selected sample.

#### Do results collected fall within a plausible range?

Yes NORC examined the dataset to ensure responses fall within valid and plausible ranges. Only one issue was encountered:

- *curso\_capacitacion\_1*: one person selected "10", one person selected "12", and two people selected "11", while the expected range is from 1 to 9.

This suggests data quality procedures were sufficient and effective for ensuring collecting data fall within the appropriate ranges for each variable.

RELIABILITY – Data should reflect stable and consistent data-collection processes and analysis methods over time.

#### Did the data collection produce responses that are internally consistent?

Yes

- NORC created 7 internal consistency checks. On average, there were 0.07 consistency problems per survey.
  - 98.3% had no problems raised.
  - 1.7% had just one failed consistency check
- The issue identified was:
  - Four respondents noted their establishment type as "Restaurant" but didn't report any kitchen staff (*cantidad\_actual1*, *cantidad\_actual2*, *cantidad\_actual3*) (1.7%)

Results suggest a high degree of internal consistency in the data.

#### Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?

- |    |   |
|----|---|
| No | <ul style="list-style-type: none"> <li>In addition to the registry census, a likely convenience sampling step (i.e., online search) was included to supplement the original sample. This may not be easily replicated in future surveys, since it is dependent on specific sites, search terms, and inclusion criteria which vary depending on the individual conducting the search.</li> <li>The organization of the dataset is insufficient to ensure it can be used reliably for reproducible analysis. Dataset must be compared against the codebook to find question and value labels, though the variables are out of order compared to the questionnaire, and the codebook is also out of order. Variable labels in the codebook are phrased differently from the questionnaire, and neither the dataset nor codebook include the question tags (e.g., "C01"). All of this makes the dataset difficult to work with reliably.</li> </ul> |
|----|---|

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.

**Is the data collected within the range of time anticipated by the survey methodology?**

- |     |  |
|-----|--|
| Yes | <ul style="list-style-type: none"> <li>Data collection period is reasonable to capture a snapshot of a single moment in time, without worrying about the comparability of data collected at different times, and is the time period envisioned in the methodology.</li> <li>The dataset itself does not include a variable to show the date of the interview.</li> </ul> |
|-----|--|

**PRECISION** – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest "5" or "10")?**

- |     |   |
|-----|---|
| Yes | NORC looked for evidence of rounding total number of workers, number of guests, percentage of guests that are foreigners, number of workers in individual jobs, number of job vacancies, number of training hours, and employee pay. There was no significant evidence of rounding. |
|-----|---|

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

- |    |  |
|----|--|
| NA | The survey was carried out in a census form. Therefore, its sampling method didn't include any power calculation and this point is not applicable. In addition, the final analysis didn't include weights, and the results were analyzed based on unweighted data. |
|----|--|

**Does the collected data allow for disaggregated analysis for underserved populations?**

- |    |   |
|----|---|
| No | <ul style="list-style-type: none"> <li>Collected data allows for disaggregating most information on employees by gender and age group, and disaggregating some by level of education.</li> <li>It is important to note the data does not allow for disaggregation by ethnic or linguistic minority groups, rural/urban, or other variables that allow identification of underserved populations.</li> </ul> |
|----|---|

**INTEGRITY** – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**



Yes

- Data was collected by tablet, minimizing the risk of transcription errors.
- After completing an initial interview, the data collection included a verification step, where each survey was reviewed for completeness and accurate recording of answers, including consistency of responses.
- According to the Methodology Protocol document, specific quality control checks performed included checks for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. This document also states that backcheck interviews were conducted on 25% of interviews, where original interview responses were compared against the responses in validation backcheck interviews to detect discrepancies. The checks described here are established best practices for quality control safeguards.

## RECOMMENDATIONS

- Many users will read the questionnaire and then look for specific variables in the dataset – this is very difficult in the dataset's current format, because variables are out of order from the questionnaire, the codebook uses different phrasing than appears in the questionnaire, and question tags (e.g., "C01") are absent. At a minimum, variables in the datasets should be presented in the same order as they appear in the questionnaire, and the codebook should include the question tags from the questionnaire.
- In addition to the two demographic variables (gender and age range) used to understand the employment situation in this sector, it is also valuable to allow disaggregation by indigenous status, as well as the demographic composition of a hotel's leadership and senior management.

## 2.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – EL SALVADOR

The data collection period for Wave 2 of the Pilot Establishment Surveys in El Salvador was from September to October 2021. For Wave 2, the food and beverage processing sector was the target for data collection. Survey administration was phone-based with responses recorded via tablet.

### SUMMARY

What is the overall conclusion regarding the quality of the data?

*Mostly High Quality*

OVERALL QUALITY 

VALIDITY 

RELIABILITY 

TIMELINESS 

PRECISION 

INTEGRITY 

### DQA QUESTION

#### ADDITIONAL DETAILS

VALIDITY – Data should represent the intended result clearly and adequately.

**Was the sampling methodology appropriately followed?**

**Yes** The sampling strategy established a sample frame from three directories of companies in the food and beverage processing sector. The sample strategy considered the entire country of El Salvador in its geographic scope and included only food and beverage processing companies that operated between August 2020 to August 2021 with at least 2 workers, for an initial sample of 415 establishments. Following a verification process of confirming the commercial name and contact information via telephone, surveyors further reduced the list to the final consolidated survey sample of 298 establishments.

**Is the response rate high enough so that we can be confident the data is reflective of the selected sample?**

**No**

- 2021 response rate: 53%

NORC suggests a minimum target of 60% response rate for establishment surveys. The response rate here is significantly lower. Therefore, we cannot confidently say that the data is reflective of the selected sample.

**Do results collected fall within a plausible range?**

- Yes
- NORC examined the dataset to ensure responses fall within valid and plausible rangers. There were no issues encountered in the examination.  
This suggests data quality procedures were sufficient and effective for ensuring collected data falls within the appropriate ranges for each variable.

**RELIABILITY** – Data should reflect stable and consistent data-collection processes and analysis methods over time.

**Did the data collection produce responses that are internally consistent?**

- Yes
- NORC created 7 internal consistency checks. On average, there were 0.03 consistency problems per survey.
    - 96.99% had no problems raised.
    - 3.01% had just one failed consistency check
  - The issues identified were:
    - One respondent noted having trained more individuals than worked at their establishment (personas\_capacitadas, trabajadores\_total) (1.7%)
    - Three respondents noted having machine operators but no machines
- Results suggest a high degree of internal consistency in the data.

**Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?**

- Yes
- Project materials present sufficient summaries of the sampling methodology to replicate a new survey with the same methodology.
  - Other project methods, including survey software as well as contact methods and protocols are clearly detailed.
  - The organization of the dataset is sufficient to ensure it can be used reliably for reproducible analysis. Variables in the Wave 2 dataset are ordered appropriately in relation to the questionnaire and codebook, and variables in the codebook include question tags (e.g., “C01”).

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.

**Is the data collected within the range of time anticipated by the survey methodology?**

- Yes
- Data was collected between September and October 2021. The data collection time period is reasonable to capture a snapshot of a single moment in time, without worrying about the comparability of data collected at different times, and is the time period envisioned in the methodology.
  - The data itself does not include a variable to show the date of the interview.

**PRECISION** – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest “5” or “10”)?**

No NORC looked for evidence of rounding total number of workers, number of workers who completed training courses, percentage of production workers, percentage of sales for the main product, and number of workers in individual jobs. There is evidence of systematically rounding numeric responses, including the following:

- Evidence of rounding total number of workers: **yes**
  - 22.6% of responses are multiples of 10
- Evidence of rounding number of workers in individual jobs: **yes**
  - 38.5% of responses are multiples of 10
- Evidence of rounding percentage of production workers: **yes**
  - 63.9% of responses are multiples of 10

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

NA

- The sampling strategy established a sample frame from three directories of companies in the food and beverage processing sector. The sample strategy considered the entire country of El Salvador in its geographic scope and included only food and beverage processing companies that operated between August 2020 to August 2021 with at least 2 workers, for an initial sample of 415 establishments. The directories were sourced from the Directorate-General for Statistics and Censuses (DIGESTYC), the Salvadoran Association of Industrialists (ASI), and the Export and Investment Promoter of El Salvador (PROESA).
- The sampling method did not include power calculations, survey stratification, or survey weights.

Although no power calculations are discussed in the methodology, sample size was determined mostly by the limits of the sample frame. Therefore, this point is not applicable.

**Does the collected data allow for disaggregated analysis for underserved populations?**

No

- Collected data allows for disaggregating most information on employees by gender and age group, and disaggregating some by level of education.
- It is important to note the data does not allow for disaggregation by ethnic or linguistic minority groups or rural/urban.

**INTEGRITY** – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**

Yes

- Data was collected by tablet, minimizing the risk of transcription errors.
- According to the Methodology Protocol document, there was a quality control check following survey implementation, where the survey team will review 100% of the questionnaires to verify the survey is filled out appropriately. In the cases where errors are detected, the surveyors will call the informant to retrieve the information prior to updating the complete survey record. Other than

this, there were no specific quality control checks performed. Based on the information provided, there were limited quality control checks during survey implementation.

## RECOMMENDATIONS

- The current CSV-formatted data must be compared against the codebook to understand the meaning of individual variables. Many users will read the questionnaire and then look for specific variables in the dataset – this is very difficult in the dataset’s current format, because one would have to reference the codebook to find question and value labels. The question tags in the codebook and questionnaire should appear in the dataset.
  - Although NORC considers the quality checks described in survey materials to be minimally sufficient, these materials do not describe other best practices. NORC recommends data collection include a backcheck of a random selection of interviews, and including checks on the main data for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. These checks were in place for the Wave 1 Establishment Surveys, but are not described in the materials for Wave 2.
  - In addition to the demographic variables offering the age range and gender of employees, it is also valuable to allow disaggregation by Indigenous status, as well as the demographic composition of the establishment’s leadership and senior management.
-

### 2.3. NATIONAL STATISTICS AGENCY HOUSEHOLD LABOR FORCE SURVEYS – EL SALVADOR (ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES - EHPM)

The data from the Encuesta de Hogares de Propósitos Múltiples (EHPM) is publicly available at an annual frequency from 2010 to 2020, conducted by El Salvador’s national statistics agency, Dirección General de Estadística y Censos (DIGESTYC). The most recent iteration of the survey that is publicly available is 2020, with 1,664 primary sampling units.

#### SUMMARY<sup>4</sup>

<p>What is the overall conclusion regarding the quality of the data?</p> <hr/> <p style="text-align: center;"><i>Somewhat High Quality</i></p> <hr/>	OVERALL QUALITY	<span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: gray;">■</span> <span style="color: gray;">■</span>
	TIMELINESS	<span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: gray;">■</span> <span style="color: gray;">■</span>
	PRECISION	<span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: gray;">■</span> <span style="color: gray;">■</span>
	USABILITY	<span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: red;">■</span> <span style="color: gray;">■</span> <span style="color: gray;">■</span>

#### DQA QUESTION      ADDITIONAL DETAILS

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence policymaking decisions.

**Are data available frequently enough to inform policymaking decisions?**

- |          |  |
|----------|--|
| Somewhat | <ul style="list-style-type: none"> <li>• EHPM data is published publicly at annual frequency from 2010-2020. Within each annual dataset, data collection occurs throughout the year, and is intended to be representative with quarterly frequency.</li> <li>• Data is published all at once for the entire year, rather than publishing quarterly datasets throughout the year. Given that data for 2021 was not publicly available as of April 2022, the lag between the close of data collection and publication is at least 4 months, and at least 1 year from the end of data collection for Quarter 1.</li> <li>• For comparable surveys among other countries in the region: Colombia (<i>Gran Encuesta Integrada de Hogares</i>), conducts surveys monthly, with data posted publicly with a lag of approximately 2 months; Costa Rica (<i>Encuesta Continua de Empleo</i>), conducts surveys quarterly, with data reported with a lag of approximately 2 months.</li> </ul> |
|----------|--|

<sup>4</sup> While the DQA evaluation of the establishment surveys rates five data quality indicators, the evaluation of the national household labor market surveys rates three data quality indicators. This decision was made following ILAB’s interests and learning priorities for understanding different aspects of the data quality for each survey.

- Frequency of data availability for EHPM (quarterly frequency) is sufficient to guide policymakers through turbulent events that result in large swings from one month to the next, or that require close monitoring of small changes. However, because the data is published all at once after the end of the year, as opposed to publishing data throughout the year, the data is of more limited use to outside users for informing decisions in real time. In other words, the data exists at an appropriate frequency, but is published too infrequently for outside users to take advantage of this.
- It is important to note that while the timeliness for reporting and publishing data are outside high quality standards, these elements were outside of the NTLMI program's scope.

**PRECISION** – Data have a sufficient level of detail to permit informed policymaking decisions.

**Has the margin of error been reported along with the data?**

Yes      Although the margin of error is not provided directly in the reports or survey materials, this is acceptable given how the materials are presented. The methodological description of the survey and calculations for precision are sufficient to allow a user to calculate this on their own, given the data. Note however, that the methodological description does not clearly state the number of households included in the sample; thus, margin of error can only be calculated from the published dataset using the number of observations found in the data.

**Are the margins of error acceptable for program/policy decision making?**

No      

- The data included 1,664 primary sampling units (PSUs). These are distributed throughout the year, with approximately 139 selected per month (416 per quarter). Survey materials do not clearly state the number of households included in the sample. The 2019 dataset includes 21,331 households (approximately 5,333 per quarter).
- For comparison, Costa Rica's Encuesta Continua de Empleo includes a sample of nearly 10,000 households across 800 primary sampling units per quarter.
- While El Salvador's sample size is adequately powered for most outcomes at the national or regional levels, it is likely to fall short for sub-populations of interest to policymakers, such as measuring changes in specific economic sectors, minority groups, or smaller geographic levels.
- It is important to note that while the project provided some assistance on sampling methods, increasing the sample size of the survey to ensure it was adequately powered for sub-populations of interest was outside the NTLMI project's scope.
- Sampling is based on the outdated 2007 census, which may impact the accuracy of estimates.

**Does the collected data allow for disaggregated analysis for traditionally underserved populations?**

Yes      

- Data allows the analyst to disaggregate by gender, age group, poverty indicators, urban/rural, and other indicators typically of interest for identifying traditionally underserved populations.
- Importantly, however, the data does not allow the analyst to disaggregate by Indigenous, Afro-Salvadoran, or other racial identification.

**USABILITY** – The ease with which an outside user can access, understand, and use the data to inform policy, business, recruitment, job training, and talent acquisition decisions.

**Is the data easy for outside users to access?**

- Yes
- Data is easy to find and download online, and is publicly available. The website is well-organized, with functioning links and clear labels.
  - However, the published data is only a subset of the full survey, including approximately half of the total number of observations. The full dataset is not publicly available.

**Is the available data well-documented, labelled, properly formatted, and easy enough to understand to be usable for outside users?**

- No
- Organization of the data varies widely from year to year and requires significant effort to understand some of these differences.
  - Variables required for survey analysis, such as household ID and PSU, are not clearly labelled.
  - Structure of the data varies from the sampling methods described in the accompanying documentation for 2020, though there is no accompanying documentation to clarify this. Survey materials describe a sample of 1,664 primary sampling units (PSUs), though the data includes only 852. It is unclear if the discrepancy is due to not publishing the full version of the data, or sampling methods that were not carried out as described in the materials (perhaps due to the 2020 COVID-19 outbreak). This makes it even more challenging to identify variables that are not clearly labelled, such as household ID and PSU.
  - Variables that relate to survey questions are clearly named and labelled to relate back to the survey questionnaire, and the survey questionnaire is consistently published as an appendix to the annual report.
  - Documentation on sampling methodology often uses language that describes what “can” be done or what “is recommended,” but is not always clear about what was actually done in practice.

**Does the data have wide enough geographic coverage to inform policy making decisions?**

- Yes
- The sample frame for EHPM includes all rural and urban communities across all 14 departments of the country.
  - The design ensures a sample that is representative of the population, with sufficient allocation of the sample across to inform decisions at national or regional levels. However, certain key geographic variables, such as department, are missing from the dataset, which limits its usefulness for making decisions below the national level.

**RECOMMENDATIONS**

- Publishing datasets with quarterly frequency would prove more useful for making decisions in real time. With the current practice of publishing data single annual datasets with data from all 4 quarters of data collection throughout the year, public access users cannot see data from the first quarter until it is already well over a year old.
- Publishing the full dataset, as opposed to a subset, will allow users to make decisions based on more complete data, reducing the margin of error in the published dataset. If full data cannot be published, data should be accompanied by a README file clarifying that published data are a subset of the full dataset, and clarifying the process used for selecting the observations in the published dataset (e.g., random selection of observations, selection of PSUs, etc.).



- Sampling should be based on the most recently available census data.
  - The methodological description is important and must be clear and concise for outside users to be able to use the data, since an analyst must program the survey structure into the statistical software to obtain accurate estimates. The existing descriptions are difficult to read. This could be improved through a concise, one-paragraph summary at the beginning of the sampling methodology appendix of the annual report (or in the README file recommended above). The paragraph should clearly identify the variable(s) used for stratification, primary sampling units, and secondary sampling units. It should also identify the number of unique units for each of these sampling variables (e.g., 14 stratas).
  - Include department as a variable in the dataset to allow users to include finer levels of geographic information in their analysis. While it is understandable that very detailed geographic information, such as municipality or community name, may be undesirable to include in a public dataset, department name should be coarse enough to avoid concerns about data privacy, and would prove very useful for analysis.
-

### 3. Data Quality Assessment for Honduras

#### 3.1. PILOT ESTABLISHMENT SURVEYS, WAVE 1 – HONDURAS

Data collection for Wave 1 of the Pilot Establishment Surveys in Honduras occurred from October to December 2019. For Wave 1, the hospitality industry was the target for data collection. Survey administration was tablet-based.

#### SUMMARY

What is the overall conclusion regarding the quality of the data?

*Mostly High Quality*

OVERALL QUALITY



VALIDITY



RELIABILITY



TIMELINESS



PRECISION



INTEGRITY



#### DQA QUESTION

#### ADDITIONAL DETAILS

**VALIDITY** – Data should represent the intended result clearly and adequately.

**Was the sampling methodology appropriately followed?**

The sampling frame was constructed from the National Registry of Tourism and the Directory of Tourist Establishments (DET) in Honduras. The table below summarizes the distribution of sample (both planned and effective) by department. The distribution in the effective sample corresponds to the initial design.

Yes

Department	Planned Sample (n = 774)	Effective Sample (n = 437)
ATLANTIDA	16%	18%

COLON	4%	4%
COMPAYAGUA	5%	5%
COPAN	14%	15%
CORTAS	22%	23%
FRANCISCO MORAZAN	15%	13%
INTIBUCA	2%	1%
ISLAS DE LA BAHIA	15%	16%
LEMPIRA	5%	5%
YORO	1%	NA

### Is the response rate high enough so that we can be confident the data is reflective of the selected sample?

The survey response rate is 65%.

Yes NORC suggests a minimum target of 60% response rate for establishment surveys. The response rate here is comparable to other existing establishment surveys in the United States, such as the OEWS (70.9% - June 2019), ARS (76.7% - June 2019), ORS (70.6% - Dec 2019), and the CESS (60.0% - Oct 2019).<sup>5, 6</sup>

### Do results collected fall within a plausible range?

NORC examined the dataset to ensure responses fall within valid and plausible ranges. Most of the data appeared fine. A few issues encountered were:

Yes

- *educacion\_puesto\_*: Response option in the questionnaire is inconsistent with dataset and codebook.
- *salario\_trabajo*: Response option "Igual o mayor 10.000 lempiras" in the questionnaire is inconsistent with dataset and codebook.
- *vacantes\_cantidad*: 3 observations where number of vacancies is "0", despite saying that there were vacancies.

**RELIABILITY** – Data should reflect stable and consistent data-collection processes and analysis methods over time.

### Did the data collection produce responses that are internally consistent?

Yes

- NORC created 6 internal consistency checks for this dataset. On average, there were 0.23 consistency problems per survey.
  - 76.6% had no problems raised.
  - 22.7% had just one failed consistency check
  - 0.69% had two or three failed consistency checks
- Checks with the greatest share of interviews with problems were:

<sup>5</sup> OEWS = Occupational Employment and Wage Statistics Survey, ARS = Annual Refiling Survey, ORS = Occupational Requirements Survey, CESS = Current Employment Statistics Survey

<sup>6</sup> <https://www.bls.gov/osmr/response-rates/>

- Business has no kitchen staff reported, but restaurant services are reported as one of its main economic activities (14.0%)
- Business has kitchen staff reported but doesn't report restaurant services as one of their main activities (6.2%)

Results suggest a high degree of internal consistency in the data.

**Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?**

- No
- The project materials contain summaries of the sampling methodology. Given the nature of the sample (census based on available registries), it is possible to run a new survey with the same methodology. In other words, the data collection method is replicable over time.
  - The data collection method, including contact protocols, scripts, and survey software (SurveyCTO), is also clearly documented.
  - The methodology protocol describes the limitations on using available registries, and proposes approaches to mitigate these problems. However, there is no documentation on any approach adopted to mitigate these limitations in actual practice.
  - The organization of the dataset is insufficient to ensure it can be used reliably for reproducible analysis. Dataset must be compared against the codebook to find question and value labels, though the dataset variables are out of order compared to the questionnaire, and the codebook is also out of order. The variable labels in the codebook are phrased differently from the questionnaire, and neither the dataset nor codebook include question tags (e.g., "C01"). All of this makes the dataset difficult to work with reliably.

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.

**Is the data collected within the range of time anticipated by the survey methodology?**

- Yes
- The survey was implemented for 8 weeks between October and December 2019. This range of time falls within the timeline (10 weeks) envisioned in the methodology protocol.

**PRECISION** – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest "5" or "10")?**

- Yes
- NORC looked for evidence of rounding total number of workers, number of workers in individual jobs, number of job vacancies, and number of training hours. There was no significant evidence of rounding.

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

- NA
- The survey was carried out in a census form. Therefore, its sampling method didn't include any power calculation and this point is not applicable. In addition, the final analysis didn't include weights, and the results were analyzed based on the original input.

**Does the collected data allow for disaggregated analysis for underserved populations?**

No

The data doesn't allow for disaggregated analysis by demographic variables. The only analysis that accounts for underserved populations includes female participation rate and age distribution of the personnel.

**INTEGRITY** – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**

Yes

- Data was collected by tablet, minimizing the risk of transcription errors.
- After completing an initial interview, the data collection included a verification step, where each survey was reviewed for completeness and accurate recording of answers, including consistency of responses.
- According to the Methodology Protocol document, specific quality control checks performed included checks for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. This document also states that backcheck interviews were conducted on 25% of interviews, where original interview responses were compared against the responses in validation backcheck interviews to detect discrepancies. The checks described here are established best practices for quality control safeguards.

**RECOMMENDATIONS**

- Many users will read the questionnaire and then look for specific variables in the dataset – this is very difficult in the dataset's current format, because variables are out of order from the questionnaire, the codebook uses different phrasing than appears in the questionnaire, and question tags (e.g., "C01") are absent. At a minimum, variables in the datasets should be presented in the same order as they appear in the questionnaire, and the codebook should include the question tags from the questionnaire.
- Findings around the internal consistency of survey responses suggest internal consistency was widely maintained. However, results also suggest that these checks were not programmed directly into the survey logic. Future surveys should consider internal consistency and program hard and soft checks into the survey logic (e.g., to generate an error message if the number of male and female employees is greater than the total number of reported employees).
- In addition to the two demographic variables (gender and age range) used to understand the employment situation in this sector, it is also valuable to allow disaggregation by indigenous status, as well as the demographic composition of a hotel's leadership and senior management.

### 3.2. PILOT ESTABLISHMENT SURVEYS, WAVE 2 – HONDURAS

The data collection period for Wave 2 of the Pilot Establishment Surveys in Honduras is not specified in survey documents. For Wave 2, the food and beverage processing industry was the target for data collection. Survey administration was phone-based.

#### SUMMARY

What is the overall conclusion regarding the quality of the data?

*Mostly High Quality*

OVERALL QUALITY



VALIDITY



RELIABILITY



TIMELINESS



PRECISION



INTEGRITY



#### DQA QUESTION

#### ADDITIONAL DETAILS

**VALIDITY** – Data should represent the intended result clearly and adequately.

**Was the sampling methodology appropriately followed?**

- Yes
- The sampling strategy established a sample frame from three directories of companies in the food and beverage processing sector, considered the entire country of Honduras in its geographic scope, and included only food and beverage processing companies that operated between March 2020 to March 2021 with at least 2 workers for an initial sample of 662.
  - Surveyors attempted to contact each establishment via telephone with the purpose of confirming the commercial name and contact information. In this process, the records of 366 establishment were unable to be confirmed and were discarded. Surveyors went on to screen establishments in the field on the requisites and further reduced the list to the final consolidated survey sample of 344 establishments.

**Is the response rate high enough so that we can be confident the data is reflective of the selected sample?**

No

- 2021 response rate: 35%

NORC suggests a minimum target of 60% response rate for establishment surveys. The response rate here is significantly lower. Therefore, we cannot confidently say that the data is reflective of the selected sample.

#### Do results collected fall within a plausible range?

Yes

NORC examined the dataset to ensure responses fall within valid and plausible rangers. No issues were encountered.

**RELIABILITY** – Data should reflect stable and consistent data-collection processes and analysis methods over time.

#### Did the data collection produce responses that are internally consistent?

NORC created 7 internal consistency checks. On average, there were 0.04 consistency problems per the survey.

Yes

- 97% had no problems raised.
- 2% had just one failed consistency check
- 1% had failed more than one consistency check
- The issues identified were:
  - Two responses noted having trained more individuals than worked at their establishment (*personas\_capacitadas*, *trabajadores\_total*)
  - One response reported having IT staff (*cantidad\_actual14*) but no technical equipment (*equipo\_fabricacion*)
  - One response reported having machine operators (*cantidad\_actual2*) but no machine equipment (*equipo\_fabricacion*)

Results suggest a high degree of internal consistency in the data.

#### Are data collection and analysis methods documented in writing that can be used to ensure the same procedures are followed each time?

Yes

- Project materials present sufficient summaries of the sampling methodology to replicate a new survey with the same methodology.
- The data collection method, including contact protocols and survey software is clearly documented.
- The organization of the dataset is sufficient to ensure it can be used reliably for reproducible analysis. Variables in the Wave 2 dataset are ordered appropriately in relation to the questionnaire and codebook, and variables in the codebook include question tags (e.g., "C01").

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence decision-making.

#### Is the data collected within the range of time anticipated by the survey methodology?

No

- The Methodology Protocol document does not include the data collection period. As such, we are unable to determine whether the data collection period is sufficient or whether it was the time period envisioned by the methodology. The time period for data collection must be included in survey documentation for the data to be useful to data users.

**PRECISION** – Data have a sufficient level of detail to permit policy makers and program managers to make decisions based on clear information.

**Are responses as precise as intended by the question (or is there any evidence of systematically rounding or approximating numeric responses)? For example, in a question about age, are responses evenly distributed (or are they systemically rounded to the nearest “5” or “10”)?**

Yes

NORC looked for evidence of rounding total number of workers, number of workers who completed the training courses, percentage of production workers, percentage of sales for the main product, and number of workers in individual jobs. NORC found only limited evidence of systematic rounding of responses:

- Evidence of rounding total number of workers: **yes**
  - 32% of responses are multiples of 10

**How was sample size determined, and were power calculations conducted to ensure the survey would be adequately powered for informing policymaking decisions?**

NA

Although no power calculations are discussed in the methodology, sample size was determined mostly by the limits of the sample frame. Therefore, this point is not applicable.

**Does the collected data allow for disaggregated analysis for underserved populations?**

No

- Collected data does not allow for disaggregating most information on employees by gender and age group, and disaggregating some by level of education.
- It is important to note the data does not allow for disaggregation by ethnic or linguistic minority groups or rural/urban.

**INTEGRITY** – Data collected should have safeguards to minimize the risk of bias, data transcription, or manipulation.

**Are appropriate procedures or safeguards in place to minimize the risk of bias, or data transcription errors?**

Yes

- Data was collected by tablet, minimizing the risk of transcription errors.
- According to the Methodology Protocol document, there was a quality control check to review 100% of the questionnaires to verify the survey was filled out appropriately. In the cases where errors were detected, the surveyors called the informant to retrieve the correct information prior to updating the complete survey record.
- Based on the information provided, there were limited quality control checks throughout survey implementation.

**RECOMMENDATIONS**



- For questions at high risk for rounding answers (e.g., number of workers), enumerators should be trained to identify the appropriate level of detail responses are expected at, and to probe for more specific answers if the respondent provides an answer that appears to be rounded to a higher level. These questions should be monitored by the data manager during data collection to detect rounding issues and follow up with enumerators.
  - Although NORC considers the quality checks described in survey materials to be minimally sufficient, these materials do not describe other best practices. NORC recommends data collection include a backcheck of a random selection of interviews, and including checks on the main data for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. These checks were in place for the Wave 1 Establishment Surveys, but are not described in the materials for Wave 2.
  - Survey materials should include a date range for data collection for the data to be usable by outside users.
  - The dataset could include a variable to show the date of each survey interview. At minimum, the survey methodology document should include the range of time in which the survey data was collected to better understand the timeliness of the data collection process.
  - In addition to the demographic variables offering the age range and gender of employees, it may be valuable to allow disaggregation by ethnic group identification, as well as the demographic composition of the establishment's leadership and senior management.
  - Specific and consistent quality control checks should be performed throughout data collection, including checks for: duplicate responses, range of responses, invalid values, examining performance by enumerator, survey duration, completed interviews per day, and response rate. The current quality control checks are insufficient in limiting data transcription errors.
-

### 3.3. NATIONAL STATISTICS AGENCY HOUSEHOLD LABOR FORCE SURVEYS – HONDURAS (ENCUESTA PERMANENTE DE HOGARES DE PROPÓSITOS MÚLTIPLES - EPHPM)

The data from the Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) is conducted once per year by Honduras’s national statistics agency, Instituto Nacional de Estadística (INE). The most recent iteration of the survey for the year 2019 included 7,200 households and 1,200 primary sampling units.

#### SUMMARY<sup>7</sup>

What is the overall conclusion regarding the quality of the data?

*Not Very High Quality*

OVERALL QUALITY 

TIMELINESS 

PRECISION 

USABILITY 

#### DQA QUESTION

#### ADDITIONAL DETAILS

**TIMELINESS** – Data should be available at a useful frequency, should be current, and should be timely enough to influence policymaking decisions.

**Are data available frequently enough to inform policymaking decisions?**

No

- Public data availability for EPHPM is insufficient for even basic analysis incorporating anything more than 2018 and 2019.
- EPHPM is conducted once per year. Once collected, public availability of the data is inconsistent from year to year. No data is currently available for 2020 or 2021.
- For comparable surveys among other countries in the region: Colombia (*Gran Encuesta Integrada de Hogares*), conducts surveys monthly, with data posted publicly with a lag of approximately 2 months; Costa Rica (*Encuesta Continua de Empleo*), conducts surveys quarterly, with data reported with a lag of approximately 2 months.

**PRECISION** – Data have a sufficient level of detail to permit informed policymaking decisions.

<sup>7</sup> While the DQA evaluation of the establishment surveys rates five data quality indicators, the evaluation of the national household labor market surveys rates three data quality indicators. This decision was made following ILAB’s interests and learning priorities for understanding different aspects of the data quality for each survey.

**Has the margin of error been reported along with the data?**

Yes                    Although the margin of error is not provided directly in the reports or survey materials, this is acceptable given how the materials are presented. The methodological description of the survey and calculations for precision are sufficient to allow a user to calculate this on their own, given the data.

**Are the margins of error acceptable for program decision making?**

No                    

- The sample of the 2021 survey included 7,200 households across 1,200 primary sampling units. Statistical power is lower in publicly available datasets, since only a subset is available, which does not include all observations and PSUs.
- For comparison, Costa Rica’s Encuesta Continua de Empleo includes a sample of nearly 10,000 households across 800 primary sampling units per quarter (40,000 households and 3,200 PSUs per year).
- While Honduras’s sample size is adequately powered for most outcomes at the national or regional levels, it is likely to fall short for sub-populations of interest to policymakers, such as measuring changes in specific economic sectors, minority groups, or smaller geographic levels.
- It is important to note that while the project provided some assistance on sampling methods, increasing the sample size of the survey to ensure it was adequately powered for sub-populations of interest was outside the NTLMI project’s scope.

**Does the collected data allow for disaggregated analysis for traditionally underserved populations?**

No                    

- Data allows the analyst to disaggregate by gender, age group, poverty indicators, urban/rural, and other indicators typically of interest for identifying traditionally underserved populations.
- Importantly, however, the data does not allow the analyst to disaggregate by Indigenous, Afro-Honduran, or other racial or ethnic identification.
- Further, the two departments not included in the survey, Gracias a Dios and Islas de Bahia, have much larger concentrations of ethnic minority populations than other parts of the country. This implies a degree of explicit underrepresentation of these populations in the survey.

**USABILITY** – The ease with which an outside user can access, understand, and use the data to inform policy, business, recruitment, job training, and talent acquisition decisions.

**Is the data easy for outside users to access?**

No                    

- Data is easy to find and download online, and is publicly available. However, datasets are only available for 2018 and 2019. The structure of the linked pages for each survey year varies greatly from year to year, with 2018 and 2019 leading to a direct download

of the survey data, and earlier years leading to a unique page for each survey wave with accompanying survey documentation and summary statistics, but no downloadable data.

#### Is the available data well-documented, labelled, properly formatted, and easy enough to understand to be usable for outside users?

- No
- Variables required for survey analysis, such as PSU, are not clearly labelled.
  - The structure of the dataset varies from the sampling methods described in the accompanying documentation because the published data is not the complete dataset, though there is no accompanying documentation to clarify this. This makes it even more challenging to identify variables that are not clearly labelled, such as PSU.
  - Variables that relate to survey questions are clearly named and labelled. However, the survey questionnaire is only available for download for years prior to 2018, which limits the analyst's ability to refer back to the documentation if they have questions about specific variables in the dataset. No data dictionary or codebook is available for any survey year.
  - Documentation on sampling methodology often uses language that describes what "can" be done or what "is recommended", but is not always clear about what was actually done in practice.

#### Does the data have wide enough geographic coverage to inform policy making decisions?

- Somewhat
- The sample frame for EPHPM includes all rural and urban communities of the country, with the exception of those in the Gracias a Dios and Islas de Bahia departments.
  - Sampling is based on the 2013 census, which is the most recent available.
  - The design ensures a sample that is representative of the population, with sufficient allocation of the sample across location to permit calculating figures for the national capital area, San Pedro Sula, other urban areas, and rural.
  - However, the exclusion of Gracias a Dios and Islas de Bahia has implications for the geographic and sociodemographic representativity of the survey. Additionally, higher proportions of ethnic minority and underserved communities in these departments limits the usefulness of these surveys for making informed policy decisions related to these geographies and communities.

### RECOMMENDATIONS

- Publishing the full datasets for each survey year, as opposed to either a subset of the complete data or summary statistics and no data, will allow users to make decisions based on more complete data, reducing the margin of error in the published dataset and allowing more flexibility in the analysis they can conduct. At a minimum, a partial dataset should be publicly available for each survey wave, as the current availability of data for surveys prior to 2018 is insufficient for even basic analyses.
- Budget is surely a major consideration for not including all departments in the survey. However, the survey should consider how Gracias a Dios and Islas de Bahia departments might be included. The exclusion of these departments limits the usefulness of the survey for understanding the situation of some of the most vulnerable populations in the country.
- If only a partial set of the full data can be published, data should be accompanied by a README file clarifying that published data are a subset of the full dataset, and clarifying the process used for selecting the observations in the published dataset (e.g., random selection of observations, selection of PSUs, etc.).

- Consistently make the survey questionnaire available for download for all survey years.
-



## ANNEX K: DATA ANALYTICS

This annex presents labor market statistics calculated from the national household labor force surveys and pilot establishment surveys for Guatemala, El Salvador, and Honduras. National household labor force surveys are conducted by each country's national statistics office, with frequency of data availability and variables identifying population sub-groups that vary by country. Pilot establishment surveys were conducted by AIR as part of the NTLMI program. Results are presented for key population sub-groups, and include labor market growth sectors, labor force participation rates, average pay in the top labor market sectors, and key barriers to labor force participation.

Labor force participation in all three countries is characterized by low levels of formal employment, high levels of informal employment, and low levels of unemployed individuals looking for work. However, labor force participation rates (i.e., those either working or looking for work) vary widely by different sub-populations, with rates that are much higher for men than women, lower for youth than older individuals, and higher for rural areas than urban. Similarly, the composition of labor force participation varies by sub-group in each country, with levels of formal employment that are much higher in urban areas than rural, and lower for women and youth. In general, the top reasons for not participating in the labor force were similar among the three countries, with women most likely to say they were not participating in the labor force due to household responsibilities, and men most likely to say they were not participating in the labor force due to their pursuit of educational or training opportunities.

While gender, age and location correlate with formal employment and labor force participation in all three countries, they do so to different degrees. Notably, the largest differences in formal employment in Guatemala are seen by gender, and when comparing urban men and women to those who identify as indigenous Guatemalans. This distinguishes Guatemala from Honduras, where the largest differences are seen comparing those in urban and rural areas. Below, we pursue a country-by-country analysis of labor market trends.

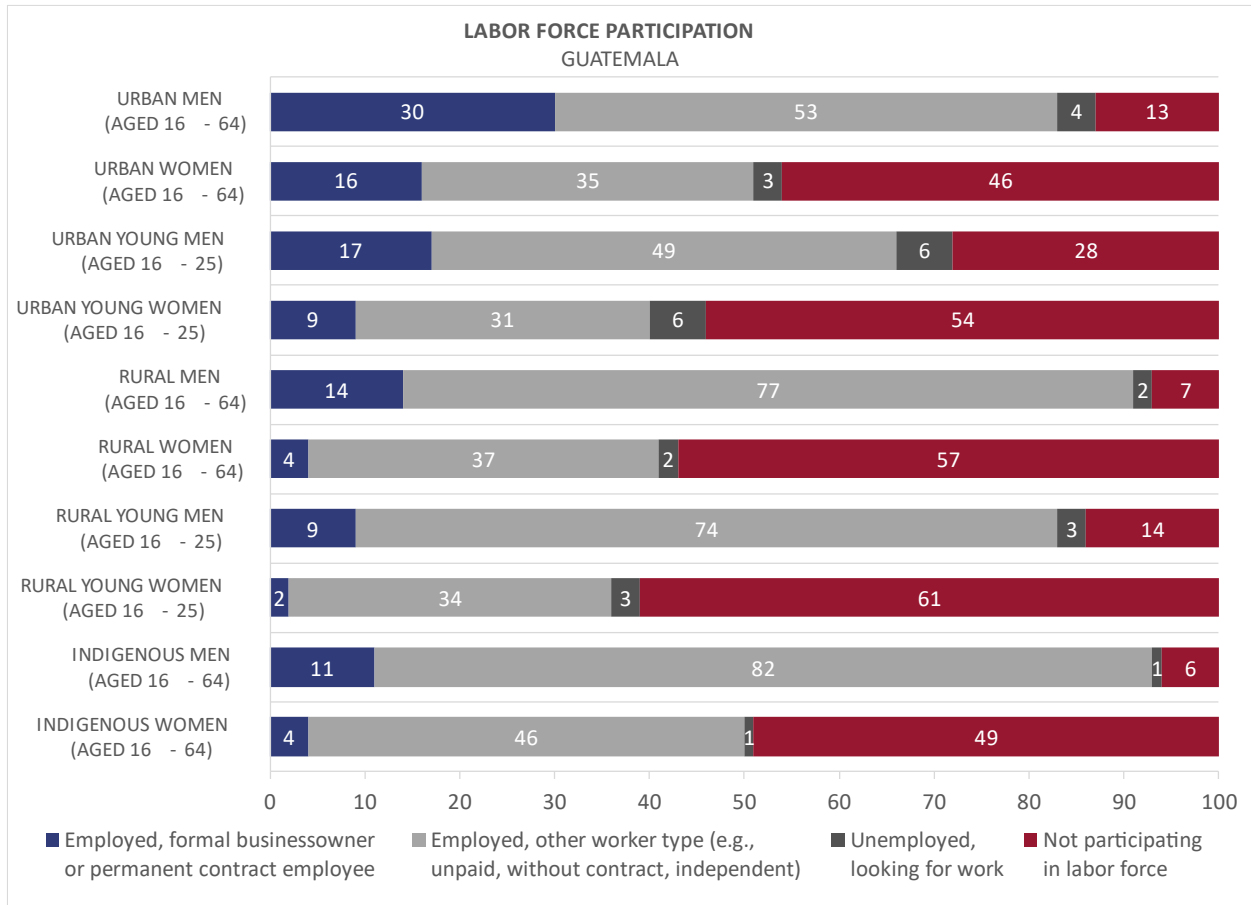
### 1. Guatemala

Labor statistics for Guatemala are calculated from the October 2021 and October 2010 National Survey on Employment and Income (ENEI).

#### 1.1. LABOR FORCE PARTICIPATION

The labor force participation rate is calculated for different sub-populations of interest for survey respondents in Guatemala between the ages of 16 and 64. Labor force participation is defined as the percentage of the population who were either working or looking for work during the week before the survey.

Figure A3.1: Guatemalan Labor Force Participation Rates, October 2021



Source: October 2021 ENEI.

Given the importance of the informal labor market in Central America, it is important to distinguish between formal and informal employment. With the data available in the ENEI, we proxy formal employment as those who are either formal business owners or private or public sector employees with contracts. Our proxy for informal employment is a category that includes public and private sector employees without contracts, unpaid workers, and independent workers.

Figure A3.1 shows the results. Of the Guatemalan population between the ages of 16 and 64, Indigenous men had the highest labor force participation rate in 2021, at 94 percent either employed or unemployed and looking for work. Like other sub-populations, however, this high labor force participation rate is characterized by a low level of unemployment and high levels of informal employment: just 11 percent of working age men who identify as indigenous Guatemalans are formally employed. Levels of formal employment are low for all sub-groups, as urban men have the highest levels of formal employment at just 30 percent, while rural young women have the lowest levels at only 2 percent.

The data show women in Guatemala experience lower levels of labor force participation compared to their male counterparts. Urban women have the highest labor force participation rate among women in Guatemala, with formal and informal employment rates of 16 percent and 35 percent, respectively, and three percent unemployed and looking for work, for a total labor force participation rate of 54



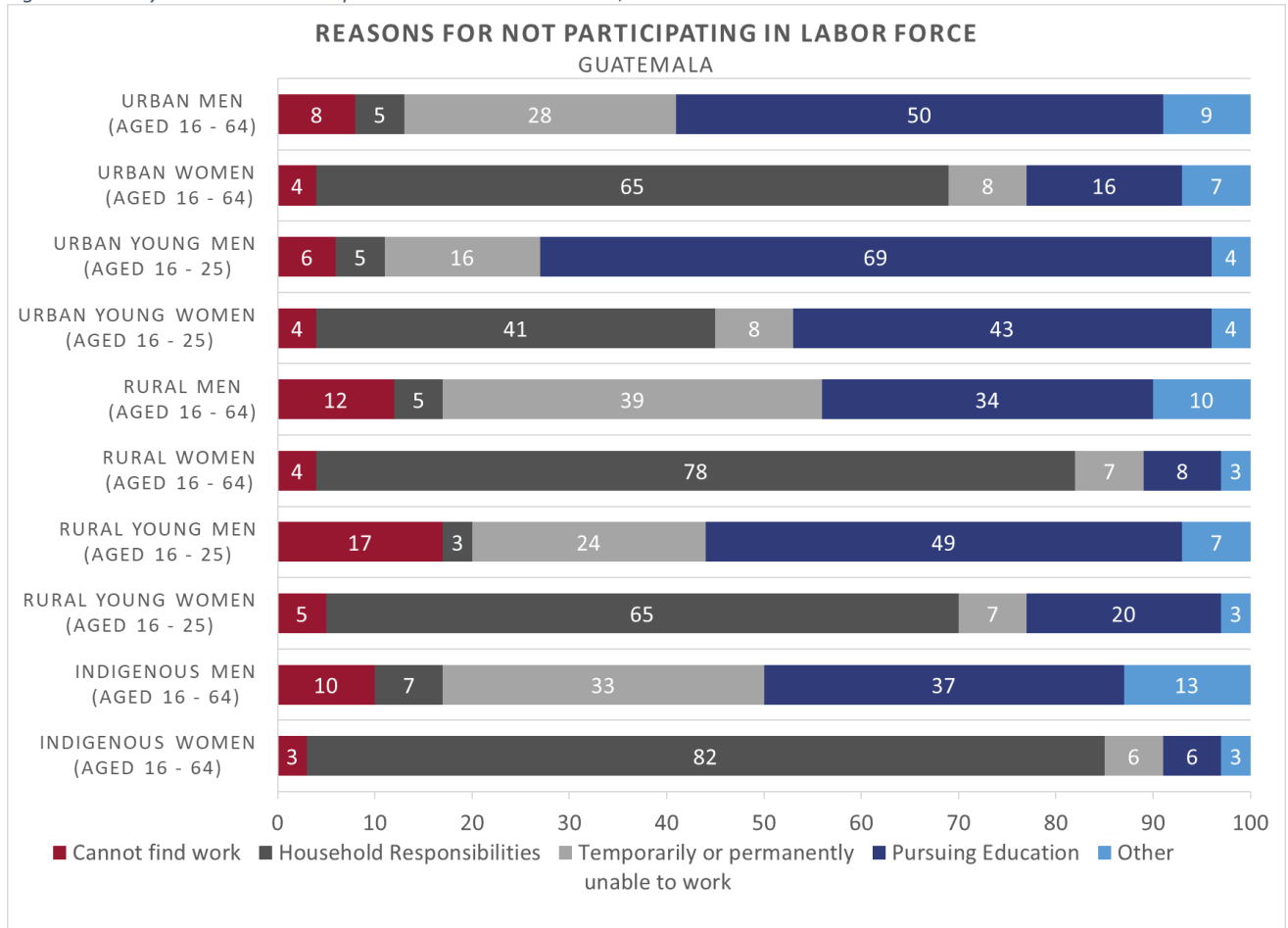
percent. Labor force participation rates among women are significantly lower among women aged 16-25, and among rural women.

## 1.2. BARRIERS TO LABOR FORCE PARTICIPATION

For those who are not working and not looking for work, we identify the main reasons for not participating in the labor force. Results are presented for key sub-groups, allowing us to compare barriers to labor force participation across different populations.

Overall, the main barriers to labor force participation for Guatemalans include being unable to find work; household responsibilities (including housework, family responsibilities, or having no one else to care for young children); education; and being temporarily or permanently unable to work (e.g., due to sickness, accident, disability, pregnancy, or age). With the exception of urban young women, at all ages and in both rural and urban areas, the main reason women did not look for work was due to household responsibilities, corresponding to between approximately 65 and 80 percent of women outside the labor force. The percentage is notably higher for rural and indigenous women than for urban women, perhaps suggesting a lack of childcare options or stricter, more traditional gender roles. For young urban women, household responsibilities (41 percent) follow behind education (43 percent) as the top reason for not participating in the labor force.

Figure A3.2: Key Labor Force Participation Barriers in Guatemala, October 2021



Source: October 2021 ENEI

### 1.3. LABOR MARKET SECTORS WITH GREATEST WAGE AND JOB GROWTH

By combining data from the October 2010 and 2021 ENEI surveys, we are able to calculate wage and job growth by economic sector. The first panel of Table A3.1 shows the top three industries for job and wage growth among those with formal employment, defined as in Figure A3.1. The top industry for wage growth was public administration, defense, teaching, healthcare, and social services, where wages grew by 59.6 percent over the period. This sector is dominated by public sector employment, though also includes private sector teachers and healthcare workers<sup>8</sup>. The top industry for job formal employment growth was agriculture and related activities, with growth of nearly 200 percent between 2010 and 2021. The next two panels in the table show the top industries for job and wage growth among those with informal employment, for urban and rural areas, respectively. In urban areas, professional, scientific, technical, and administrative services represented the top industry for both wage growth (60.4 percent) and job growth (68.6 percent) among those with informal employment. In rural areas, the top industry in the informal sector for wage growth was public administration, defense, teaching, healthcare, and social services (94.5 percent), while construction (203.3 percent) saw the highest job growth. Though it may appear odd that an industry dominated by public sector

<sup>8</sup> Guatemala’s ENEI survey includes these activities in a single economic category, regardless of whether the employer is public or private.

employment had the highest wage growth in the *informal* economy in rural areas, the available data suggests this growth was mainly driven by teachers in private schools without employment contracts.

The remaining panels of Table A3.1 show the top three sectors for wage and job growth, for all of Guatemala and by level of education, including both formal and informal employment together. We define secondary education as having completed *diversificado*, the Guatemalan equivalent of senior high school, typically completed at age 18. The data show that approximately 90 percent of the rural population has less than a secondary education, compared to approximately 60 percent of the urban population. Including both formal and informal employment together, the industries with the highest wage growth between 2010 and 2021 are public administration, defense, teaching, healthcare, and social services (55.4 percent), other services<sup>9</sup> (48.1 percent), and professional, scientific, technical, and administrative services (37.1 percent). The industries with the highest job growth between 2010 and 2021 are professional, scientific, technical, and administrative services (117 percent), construction (105 percent), and wholesale and retail commerce, storage and transport, and hotel and food services (55.8 percent).

The results are further broken down by education level. In terms of wage growth, public administration, defense, teaching, healthcare, and social services is the top sector for Guatemalans of all education levels, with the exception of urban Guatemalans with less than secondary education (where it is the industry with the third highest wage growth). The results suggest wage growth in this industry was relatively even, regardless of education level, at between 45-60 percent.

Construction saw the highest job growth over the period for Guatemalans of all education levels, with the exception of those with tertiary education. Construction jobs grew by 22.8 percent for urban Guatemalans with less than secondary education, 173 percent for rural Guatemalans with less than secondary education, and by 256 percent for Guatemalans with a secondary education. Further analysis shows that while construction was consistently among the top sectors contributing to the employment of both rural and urban men of all education levels, the industry was not represented among the top five employing sectors for rural or urban women of any education level.

Table A3.1: Guatemalan Labor Market Sectors with Greatest Wage and Job Growth (2010-2021), by Education Level

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2010 and 2021				Top Economic Sectors with Highest <u>Job</u> Growth Between 2010 and 2021			
Sector	Average Wage 2010, Quetzales (USD)	Average Wage 2021, Quetzales (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2010	Number Employed in 2021	Average Job Growth
Respondents with Formal Employment							
Public administration, defense, teaching, healthcare, and social services	Q3,002 (US\$373)	Q4,792 (US\$620)	59.6%	Agriculture, livestock, hunting, and related activities	50,849	151,076	197.1%

<sup>9</sup> For Guatemala, the “other services” category includes artistic activities and entertainment; libraries, museums, and cultural activities; gambling and betting; sports and recreation; repair of electronics and household goods; domestic services; and other personal services.

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2010 and 2021				Top Economic Sectors with Highest <u>Job</u> Growth Between 2010 and 2021			
Sector	Average Wage 2010, Quetzales (USD)	Average Wage 2021, Quetzales (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2010	Number Employed in 2021	Average Job Growth
Financial and insurance activities	Q2,602 (US\$323)	Q3,667 (US\$474)	40.9%	Professional, scientific, technical, and administrative services	53,656	139,172	159.4%
Manufacturing, mining, and quarrying	Q3,096 (US\$384)	Q4,209 (US\$544)	35.9%	Other services (excluding financial, hotel, and food services)	21,328	48,655	128.1%
Respondents with Informal Employment - Urban							
Professional, scientific, technical, and administrative services	Q2,045 (US\$254)	Q3,280 (US\$424)	60.4%	Professional, scientific, technical, and administrative services	40,905	68,952	68.6%
Other services (excluding financial, hotel, and food services)	Q888 (US\$110)	Q1,348 (US\$171)	51.8%	Construction	132,448	185,765	40.3%
Information and communication	Q1,951 (US\$242)	Q2,926 (US\$378)	50.0%	Wholesale and retail commerce, storage and transport, hotel and food service	737,464	852,184	15.6%
Respondents with Informal Employment - Rural							
Public administration, defense, teaching, healthcare, and social services	Q1,412 (US\$175)	Q2,746 (US\$355)	94.5%	Construction	110,910	336,405	203.3%
Real Estate	Q1,500 (US\$186)	Q2,400 (US\$310)	60.0%	Other services (excluding financial, hotel, and food services)	144,875	252,915	74.6%

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2010 and 2021				Top Economic Sectors with Highest <u>Job</u> Growth Between 2010 and 2021			
Sector	Average Wage 2010, Quetzales (USD)	Average Wage 2021, Quetzales (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2010	Number Employed in 2021	Average Job Growth
Other services (excluding financial, hotel, and food services)	Q789 (US\$98)	Q1,110 (US\$144)	40.7%	Wholesale and retail commerce, storage and transport, hotel and food service	389,791	859,830	120.6%
All Respondents (including formal and informal employment)							
Public administration, defense, teaching, healthcare, and social services	Q2,809 (US\$349)	Q4,365 (US\$564)	55.4%	Professional, scientific, technical, and administrative services	112,294	244,239	117%
Other services (excluding financial, hotel, and food services)	Q964 (US\$120)	Q1,427 (US\$184)	48.1%	Construction	289,936	594,079	105%
Professional, scientific, technical, and administrative services	Q2,650 (US\$329)	Q3,633 (US\$470)	37.1%	Wholesale and retail commerce, storage and transport, hotel and food service	1,373,181	2,139,901	55.8%
Respondents with PhD, Masters, and Other Advanced Degrees (including formal and informal employment)							
Public administration, defense, teaching, healthcare, and social services	Q3,583 (US\$445)	Q5,398 (US\$697)	50.6%	Professional, scientific, technical, and administrative services	25,378	78,977	211%
Other services (excluding financial, hotel, and food services)	Q3,385 (US\$420)	Q4,429 (US\$572)	30.8%	Other services (excluding financial, hotel, and food services)	11,049	29,216	164%
Financial and insurance activities	Q4,413 (US\$548)	Q4,871 (US\$630)	10.4%	Agriculture, livestock, hunting, and related activities	8,529	20,552	141%
Respondents with Secondary Education (including formal and informal employment)							

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2010 and 2021				Top Economic Sectors with Highest <u>Job</u> Growth Between 2010 and 2021			
Sector	Average Wage 2010, Quetzales (USD)	Average Wage 2021, Quetzales (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2010	Number Employed in 2021	Average Job Growth
Public administration, defense, teaching, healthcare, and social services	Q2,580 (US\$320)	Q4,016 (US\$519)	55.7%	Construction	22,886	81,545	256%
Professional, scientific, technical, and administrative services	Q2,293 (US\$285)	Q3,561 (US\$460)	55.4%	Wholesale and retail trade, transport and storage, accommodation, and food service activities	278,949	637,652	128%
Financial and insurance activities	Q2,850 (US\$354)	Q3,620 (US\$468)	27.0%	Professional, scientific, technical, and administrative services	41,225	87,210	112%
Respondents with Less than Secondary Education (Urban - including formal and informal employment)							
Manufacturing, mining, and quarrying	Q1,524 (US\$189)	Q2,271 (US\$294)	49.0%	Construction	139,907	171,923	22.8%
Other services (excluding financial, hotel, and food services)	Q858 (US\$106)	Q1,263 (US\$163)	47.1%	Professional, scientific, technical, and administrative services	30,207	34,440	14%
Public administration, defense, teaching, healthcare, and social services	Q2,139 (US\$265)	Q3,139 (US\$405)	46.7%	Wholesale and retail trade, transport and storage, accommodation, and food service activities	627,090	640,289	2%
Respondents with Less than Secondary Education (Rural - including formal and informal employment)							
Public administration, defense, teaching, healthcare, and social services	Q1,526 (US\$189)	Q2,428 (US\$314)	59.1%	Construction	119,410	326,399	173%

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2010 and 2021				Top Economic Sectors with Highest <u>Job</u> Growth Between 2010 and 2021			
Sector	Average Wage 2010, Quetzales (USD)	Average Wage 2021, Quetzales (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2010	Number Employed in 2021	Average Job Growth
Other services (excluding financial, hotel, and food services)	Q786 (US\$98)	Q1,120 (US\$145)	42.5%	Wholesale and retail trade, transport and storage, accommodation, and food service activities	389,367	737,592	89.4%
Agriculture, livestock, hunting, and related activities	Q892 (US\$111)	Q1,267 (US\$164)	42.0%	Other services (excluding financial, hotel, and food services)	145,901	233,900	60.3%

Source: October 2010 and October 2021 ENEI

Note: Exchange rates come from the International Monetary Fund. Average exchange rate for 2010 = 8.06 Quetzales per US Dollar. Average exchange rate for 2021 = 7.73 Quetzales per US Dollar.

#### 1.4. KEY LABOR MARKET SECTORS

For each population sub-group, we calculate the top economic sectors employing the greatest share of the population, along with the average monthly salary. Figure A3.3 shows the results, with sub-groups for men shown in the left panel, and sub-groups for women shown in the right.

Among employed men in 2021, the main economic activity for all sub-groups was activities related to agriculture, livestock, and hunting<sup>10</sup>. The sole exception is urban young men, whose top sector was retail trade (18 percent). The main economic activity for women sub-groups was retail trade, with the exception of indigenous women, whose main economic activity was agriculture (24 percent). Those living in urban areas earned higher salaries compared to Guatemalans in rural areas working within the same industry. Similarly, young urban men earned higher average salaries compared to their female counterparts in the retail trade industry, while indigenous men earned higher average salaries than their female counterparts in agriculture.

<sup>10</sup> While it may be surprising that agriculture is a top employer in urban areas, the data shows this comes almost entirely from urban areas outside of the department of Guatemala. These towns and smaller cities, while technically urban, still have a large agricultural economy, particularly in the surrounding countryside.

Figure A3.3: Top Employing Labor Market Sectors in Guatemala and Associated Average Salaries, by Sub-group

MEN				WOMEN			
GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)	GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)
Urban	Agriculture, livestock, hunting, and related activities	16%	\$190	Urban	Retail trade, excluding cars and motorcycles	26%	\$301
Urban Youth*	Retail trade, excluding cars and motorcycles	18%	\$274	Urban Youth*	Retail trade, excluding cars and motorcycles	28%	\$227
Rural	Agriculture, livestock, hunting, and related activities	53%	\$173	Rural	Retail trade, excluding cars and motorcycles	23%	\$183
Rural Youth*	Agriculture, livestock, hunting, and related activities	50%	\$172	Rural Youth*	Retail trade, excluding cars and motorcycles	26%	\$170
Indigenous	Agriculture, livestock, hunting, and related activities	52%	\$144	Indigenous	Agriculture, livestock, hunting, and related activities	24%	\$116

\*Aged 16-25

Source: October 2021 ENEI.

Note: Exchange rates come from the International Monetary Fund. Average exchange rate for 2021 = 7.73 Quetzales per US Dollar.

### 1.5. EMPLOYER NEEDS AND BARRIERS TO FINDING TALENT

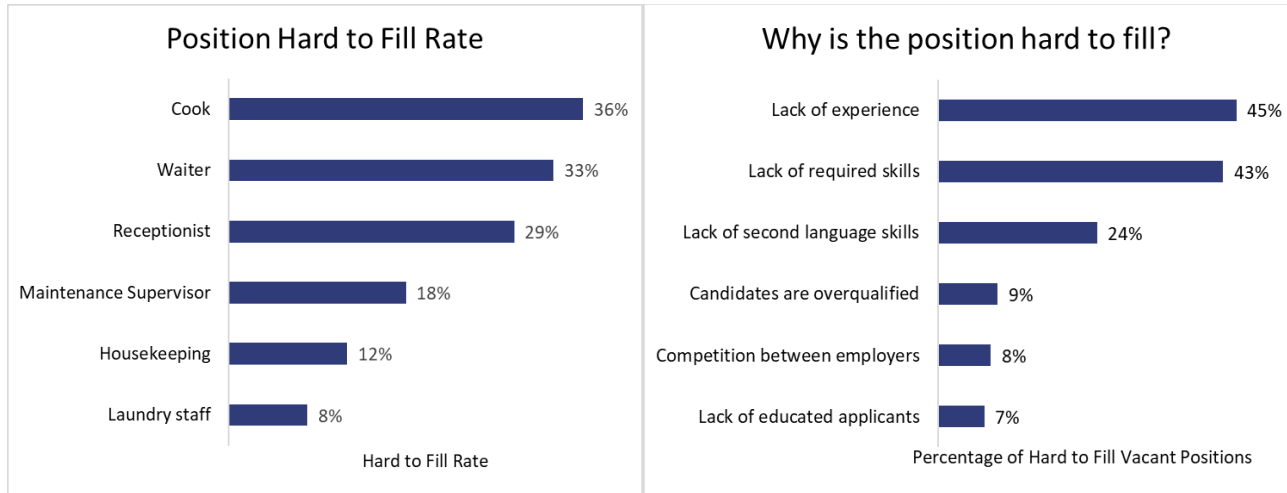
We use the Hospitality Sector Pilot Establishment Survey conducted by AIR to identify the positions in the hospitality industry that are hardest to fill, along with the reasons these positions are difficult to fill, as identified by managers.

Among the 682 businesses responding to the survey, the most common positions among respondents to the Hospitality Sector Pilot Establishment Survey in Guatemala were: housekeeping ( $n=984$  total positions), receptionist ( $n=551$ ), cook ( $n=191$ ), waiter ( $n=119$ ), laundry staff ( $n=74$ ), and maintenance supervisor ( $n=61$ ). For each position, we calculate a hard to fill rate, defined as the sum of employees in the position among employers that consider the position “hard to fill”, divided by the sum of employees in the position among all employers in the survey. Figure A3.4 shows that the positions that are the hardest to fill are cooks (36 percent), waiters (33 percent), and receptionists (29 percent).

Employers were also asked why positions were hard to fill. Among all hard to fill positions, the most common reasons given were a lack of candidates with experience in the position (45 percent), applicants that lack the required skills for the position (43 percent), and a lack of candidates with second language skills (43 percent).



Figure A3.4: Employer Needs and Barriers to Finding Talent in the Guatemala Hospitality Sector



Source: 2019 Guatemala Hospitality Sector Pilot Establishment Survey (AIR).

## 2. El Salvador

Labor statistics for El Salvador are calculated from the 2020 and 2015 Multipurpose Household Survey (EHPM)<sup>11</sup>. Data collection for the second quarter of 2020 was impacted by the COVID-19 pandemic; as a result, all calculations include data from quarters one, three, and four only, to ensure data is comparable across the two years.

### 2.1. LABOR FORCE PARTICIPATION

The labor force participation rate is calculated for different sub-populations of interest for survey respondents in El Salvador between the ages of 16 and 64.<sup>12</sup> Labor force participation is defined as the percentage of the population who were either working or looking for work during the week before the survey.

With the data available in the 2020 EHPM, we attempt to distinguish between formal and informal employment among the Salvadoran labor force. We define formal employment as those who are either formal business owners or private or public sector employees with contracts. Our proxy for informal employment is a category that includes public and private sector employees without contracts, unpaid workers, and independent workers.

Figure A3.5 shows rates of labor force participation among selected groups. Among respondents aged 16 to 64, rural men had the highest labor force participation rate, at 86 percent either employed or unemployed and looking for work. Like other sub-populations, this high labor force participation rate is characterized by a low level of unemployment and high levels of informal employment: just 18 percent of working age rural men are formally employed. As in Guatemala, levels of formal employment in El Salvador are low for all sub-groups, and urban men have the highest levels of formal employment at 33 percent, while rural young women have the lowest levels at only 5 percent.

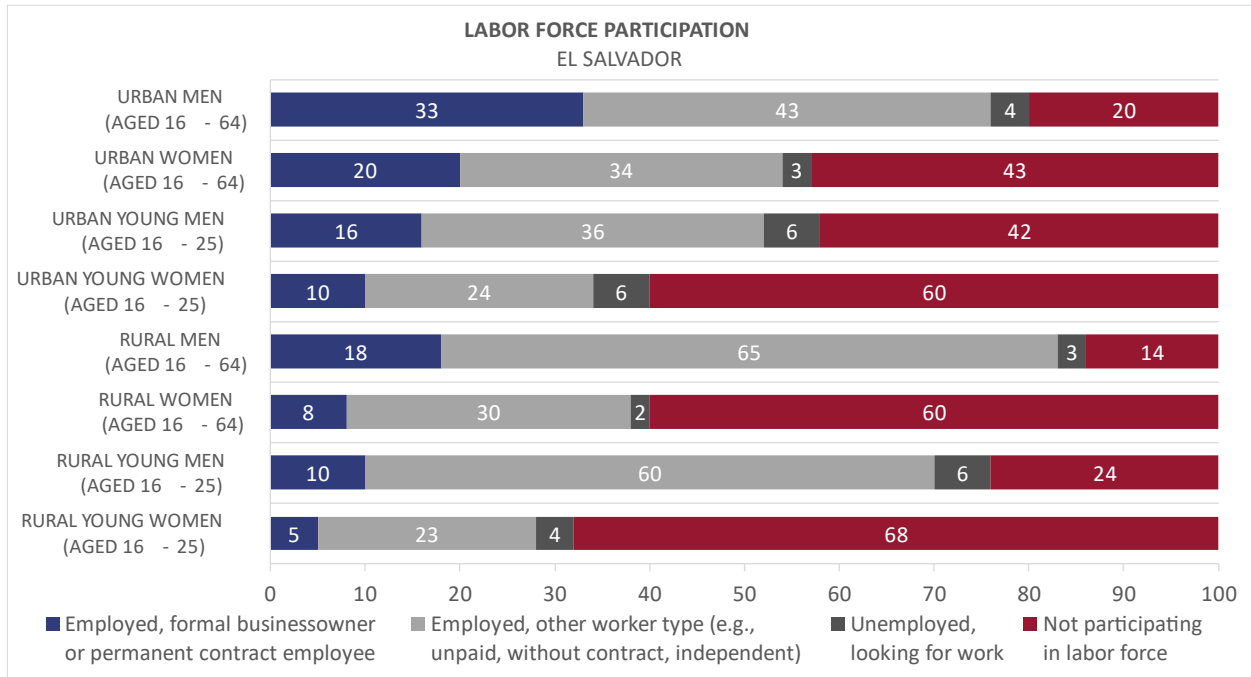
<sup>11</sup> EHPM data is available before 2015, but has a very different format and database structure from what is used in later years, making it difficult to compare against the 2020 data and calculate growth rates.

<sup>12</sup> Unlike Guatemala’s national labor force survey, El Salvador’s EHPM does not ask respondents for their racial or ethnic identification. Thus, results are not presented for ethnic minority groups.

Unlike in Guatemala, El Salvador’s EHPM does not allow data to be disaggregated by ethnic minority groups.

The data show women in El Salvador experience lower levels of labor force participation compared to their male counterparts. As in Guatemala, urban women in El Salvador had the highest labor force participation rate among women, with formal and informal employment rates of 20 percent and 34 percent, respectively, and three percent unemployed and looking for work, for a total labor force participation rate of 57 percent. Rural young women had the lowest rates of labor force participation, at just 32 percent.

Figure A3.5: Salvadoran Labor Force Participation Rates, 2020



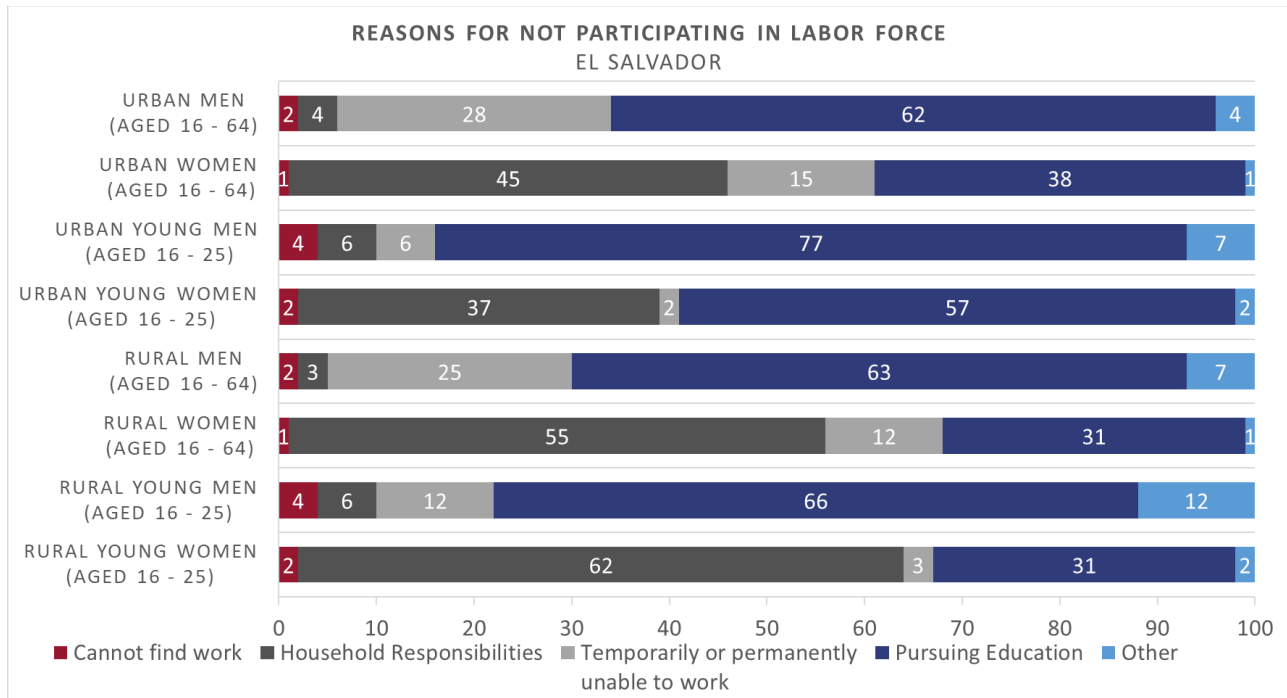
Source: 2020 EHPM, Quarters 1, 3, and 4.

## 2.2. BARRIERS TO LABOR FORCE PARTICIPATION

For those who are not working and not looking for work, we identify the main reasons for not participating in the labor force. Results are presented for key sub-groups, allowing us to compare barriers to labor force participation across different populations.

Overall, the main reasons for not participating in the labor force for respondents in El Salvador include household responsibilities (including household chores, care for dependents, and family and personal obligations); pursuing education; and temporarily or permanently unable to work due to reasons such as accident, advanced age, or disability. Differences by gender are also informative: while the primary reason for not participating in the labor force for all groups of men was the pursuit of educational opportunities, suggesting they will enter the labor force eventually, the main reason for Salvadoran women was household responsibilities, with the sole exception of urban young women, for whom household responsibilities (37 percent) follows behind education (57 percent).

Figure A3.6: Key Labor Force Participation Barriers in El Salvador, 2020



Source: 2020 EHPM, Quarters 1,3, and 4.

### 2.3. LABOR MARKET SECTORS WITH GREATEST WAGE AND JOB GROWTH

By combining data from the 2015 and 2020 EHPM surveys, we are able to calculate wage and job growth by economic sector. The first panel of Table A3.2 shows the top three industries for job and wage growth among those with formal employment, defined as in Figure A3.4. The top industry for formal employment wage growth was transportation, storage, and communications, where wages grew by 58.6 percent. The top industry for formal job growth was construction, with growth of 61 percent between 2015 and 2020. The next two panels in the table show the top industries for job and wage growth among those with informal employment, for urban and rural areas, respectively. In urban areas, manufacturing represented the top industry for informal employment wage growth (30.6 percent), while the top industry for informal job growth was construction (57.5 percent). In rural areas, the top informal sector for wage growth was commerce, hotels, and restaurants (31.8 percent), while construction (89.9 percent) saw the highest job growth.

The remaining panels of Table A3.2 show the top three sectors for wage and job growth, for all respondents and by level of education, including both formal and informal sectors of the economy together. We define secondary education as having reached *media*, the Salvadoran equivalent of senior high school. The data show that approximately 80 percent of the rural population has less than a secondary education, compared to approximately 58 percent of the urban population. About 80 percent of rural women have less than a secondary education, compared to 59 percent of urban women. Similarly, about 79 percent of rural men have less than a secondary education, compared to about 54 percent of urban men.

The economic sectors with the highest job growth between 2015 and 2020 are construction (68.5 percent), domestic services (50.3 percent), and finance and real estate (44.1 percent). Construction represents one of the top three sectors for job growth for Salvadorans of all education levels, ranging

from 40.4 percent among urban Salvadorans with less than secondary education, to 122.5 percent among all Salvadorans with a secondary education. Although it does not appear as one of the top three sectors among all respondents in El Salvador, the commerce, hotels, and restaurants sector is notable in that it appears as one of the top three sectors for wage growth among Guatemalans of all education levels, with the exception of those with a secondary education. *Table A3.2: El Salvador Labor Market Sectors with Greatest Wage and Job Growth (2015-2020), by Education Level*

Top Economic Sectors with Highest Wage Growth Between 2015 and 2020				Top Economic Sectors with Highest Job Growth Between 2015 and 2020			
Sector	Average Wage 2015, USD	Average Wage 2020, USD	Average Nominal Wage Growth	Sector	Number Employed in 2015	Number Employed in 2020	Average Job Growth
Respondents with Formal Employment							
Transportation, storage, and communications	\$412	\$653	58.6%	Construction	18,243	29,366	61.0%
Commerce, hotels, and restaurants	\$299	\$418	40.1%	Transportation, storage, and communications	29,034	45,391	56.3%
Education	\$558	\$726	30.2%	Education	45,516	70,733	55.4%
Respondents with Informal Employment - Urban							
Manufacturing	\$209	\$273	30.6%	Construction	53,795	84,723	57.5%
Education	\$270	\$350	29.6%	Domestic Services	59,042	86,886	47.2%
Domestic Services	\$148	\$190	28.2%	Finance and Real Estate	36,647	48,986	33.7%
Respondents with Informal Employment - Rural							
Commerce, hotels, and restaurants	\$177	\$234	31.8%	Construction	34,222	64,973	89.9%
Finance and Real Estate	\$250	\$329	31.7%	Manufacturing	57,689	90,648	57.1%
Domestic Services	\$146	\$175	19.3%	Transportation, storage, and communications	16,118	25,216	56.4%

Top Economic Sectors with Highest Wage Growth Between 2015 and 2020				Top Economic Sectors with Highest Job Growth Between 2015 and 2020			
Sector	Average Wage 2015, USD	Average Wage 2020, USD	Average Nominal Wage Growth	Sector	Number Employed in 2015	Number Employed in 2020	Average Job Growth
All Respondents (including formal and informal employment)							
Transportation, storage, and communications	\$337	\$475	41.1%	Construction	106,299	179,061	68.5%
Education	\$499	\$686	37.5%	Domestic Services	98,698	148,330	50.3%
Domestic Services	\$149	\$194	30.4%	Finance and Real Estate	113,978	164,230	44.1%
Respondents with PhD, Masters, and Other Advanced Degrees (including formal and informal employment)							
Transportation, storage, and communications	\$509	\$850	66.9%	Agriculture, livestock, hunting, and related activities	4,475	7,946	77.6%
Commerce, hotels, and restaurants	\$397	\$439	10.7%	Transportation, storage, and communications	14,789	25,671	73.6%
Manufacturing	\$478	\$527	10.2%	Construction	8,261	13,695	65.8%
Respondents with Secondary Education (including formal and informal employment)							
Education	\$311	\$457	47.1%	Construction	19,443	43,265	122.5%
Domestic Services	\$151	\$214	41.7%	Domestic services	16,873	31,566	87.1%
Finance and Real Estate	\$293	\$392	34.1%	Commerce, hotels, and restaurants	186,485	315,624	69.2%
Respondents with Less than Secondary Education (Urban - including formal and informal employment)							
Commerce, hotels, and restaurants	\$187	\$254	35.8%	Construction	47,378	66,535	40.4%
Transportation, storage, and communications	\$306	\$407	33.2%	Domestic services	47,908	65,532	36.8%
Health and Social Services	\$237	\$312	31.6%	Transportation, storage, and communications	35,028	45,124	28.8%
Respondents with Less than Secondary Education (Rural - including formal and informal employment)							

Top Economic Sectors with Highest Wage Growth Between 2015 and 2020				Top Economic Sectors with Highest Job Growth Between 2015 and 2020			
Sector	Average Wage 2015, USD	Average Wage 2020, USD	Average Nominal Wage Growth	Sector	Number Employed in 2015	Number Employed in 2020	Average Job Growth
Commerce, hotels, and restaurants	\$159	\$232	45.7%	Transportation, storage, and communications	12,337	22,270	80.5%
Finance and Real Estate	\$240	\$318	32.3%	Construction	31,217	55,567	78.0%
Manufacturing	\$199	\$251	26.3%	Finance and Real Estate	13,213	20,035	51.6%

Source: 2015 and 2020 EHPM, Quarters 1, 3, and 4

Note: Since 2001, El Salvador has used the US dollar as its local currency.

## 2.4. KEY LABOR MARKET SECTORS

For each population sub-group, we calculate the top economic sectors employing the greatest share of the population, along with the average monthly salary.

In El Salvador, the main economic activity among rural men of any age group was agriculture, livestock, hunting and fishing. The industry employed 46.5 percent of rural men and 44.4 percent of rural young men, with an average monthly salary of \$174 and \$171 USD, respectively. For rural women and rural young women, the main employment sector was commerce, hotels, and restaurants. The industry employed 42.6 percent of rural women and 48.5 percent of rural young women, with average monthly salaries of \$228 and \$229, respectively. Commerce, hotels, and restaurants also represents the main sector of employment for all urban sub-populations, with salaries that are higher, on average, than their rural counterparts.

Figure A3.7: Highest Employing Labor Market Sectors Per Group in El Salvador and Associated Average Salaries (USD)

MEN				WOMEN			
GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)	GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)
Urban	Commerce, hotels, and restaurants	28%	\$349	Urban	Commerce, hotels, and restaurants	46%	\$281
Urban Youth*	Commerce, hotels, and restaurants	38%	\$288	Urban Youth*	Commerce, hotels, and restaurants	49%	\$255
Rural	Agriculture, livestock, hunting, and forestry	47%	\$174	Rural	Commerce, hotels, and restaurants	43%	\$228
Rural Youth*	Agriculture, livestock, hunting, and forestry	44%	\$171	Rural Youth*	Commerce, hotels, and restaurants	49%	\$229

\*Aged 16-25

Source: 2020 EHPM, Quarters 1, 3, and 4

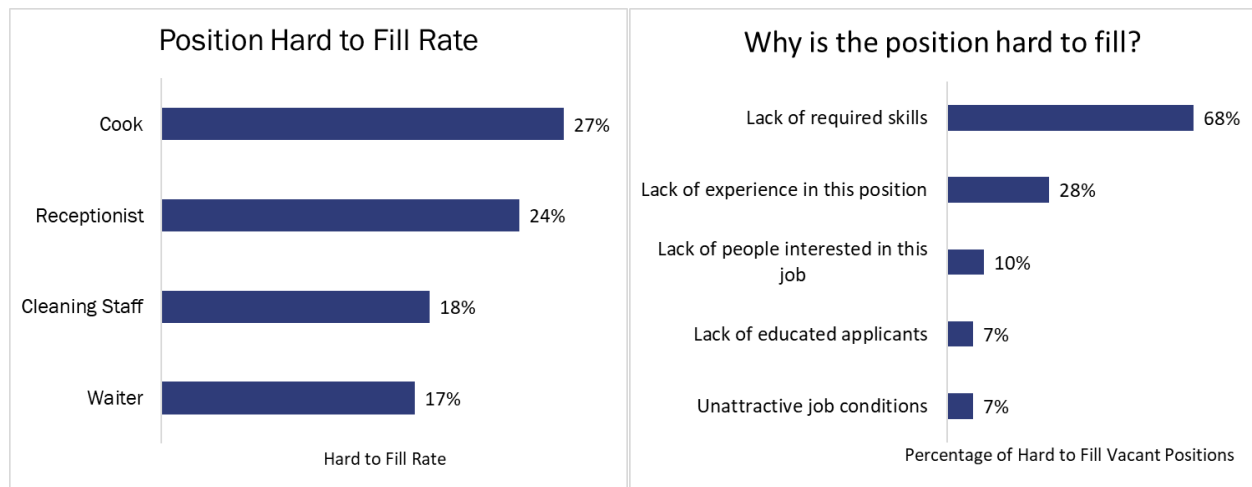
## 2.5. EMPLOYER NEEDS AND BARRIERS TO FINDING TALENT

We use the Hospitality Sector Pilot Establishment Survey conducted by AIR to identify the positions in the hospitality industry that are hardest to fill, along with the reasons these positions are difficult to fill, as identified by managers.

Among the 233 businesses responding to the survey, the most common positions among respondents to the Hospitality Sector Pilot Establishment Survey in El Salvador were: housekeeping ( $n=58$  total positions), cleaning staff ( $n=56$ ), waiter ( $n=48$ ), and cook ( $n=37$ ). For each position, we calculate a hard to fill rate, defined as the sum of employees in the position among employers that consider the position “hard to fill”, divided by the sum of employees in the position among all employers in the survey. Figure A3.8 shows that the positions that are the hardest to fill are cooks (27 percent), receptionists (24 percent), and cleaning staff (18 percent).

Employers were also asked why positions were hard to fill. Among all hard to fill positions, the most common reasons given were a lack of candidates with required skills for the position (68 percent), applicants that lack experience in the position (28 percent), and a lack of candidates interested in the job (10 percent).

Figure A3.8: Employer Needs and Barriers to Finding Talent in the El Salvador Hospitality Sector



Source: 2019 El Salvador Hospitality Sector Pilot Establishment Survey (AIR)

## 3. Honduras

Labor statistics for Honduras are calculated from the October 2019 and October 2018 Permanent Multipurpose Household Surveys (EPHPM).<sup>13</sup>

<sup>13</sup> As of the time of writing, downloadable EPHPM datasets were only publicly available for 2018 and 2019. NORC made attempts to obtain 2010 and 2015 data from Honduran officials, though the timing of the requests coincided with Honduran elections and a change in administration, and the requests were never fulfilled.

### 3.1. LABOR FORCE PARTICIPATION

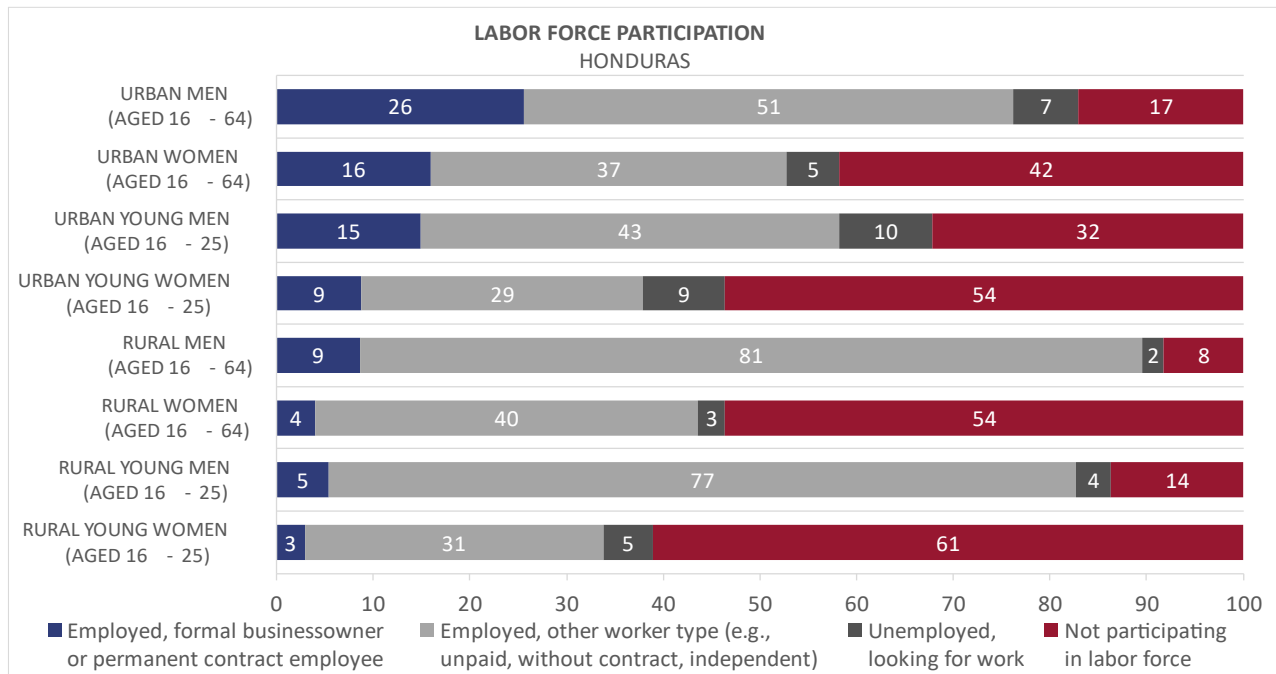
The labor force participation rate is calculated for different sub-populations of interest for Hondurans between the ages of 16 and 64.<sup>14</sup> Labor force participation is defined as the percentage of the population who were either working or looking for work during the week before the survey.

With data available in the 2019 EPHPM, we distinguish between formal and informal employment in the Honduran labor force. We define formal employment as either formal business owners or private or public sector employees with contracts. Our proxy for informal employment is a category that includes public and private sector employees without contracts, unpaid workers, and independent workers.

Figure A3.9 shows rates of labor force participation in Honduras among selected groups. Rural men experienced the highest labor force participation rates, at 92 percent either employed or unemployed and looking for work. Like other sub-populations, this high labor force participation rate is characterized by a low level of unemployment and high levels of informal employment: just 9 percent of working age rural men are formally employed. Urban men have the highest levels of formal employment, still at only 26 percent, while rural young women aged 16-25 have the lowest formal employment levels, at just 3 percent.

The data show women in Honduras have lower labor force participation compared to their male counterparts. Urban women had the highest labor force participation rate among women, with formal and informal employment rates of 16 percent and 37 percent, respectively, and seven percent unemployed and looking for work, for a total participation rate of 58 percent. Rural young women had the lowest rates of participation, at just 39 percent.

Figure A3.9: Honduran Labor Force Participation Rates, 2019



Source: 2019 EPHPM

<sup>14</sup> Unlike Guatemala’s national labor force survey, Honduras’s EPHPM does not ask respondents for their racial or ethnic identification. Thus, results are not presented for ethnic minority groups.

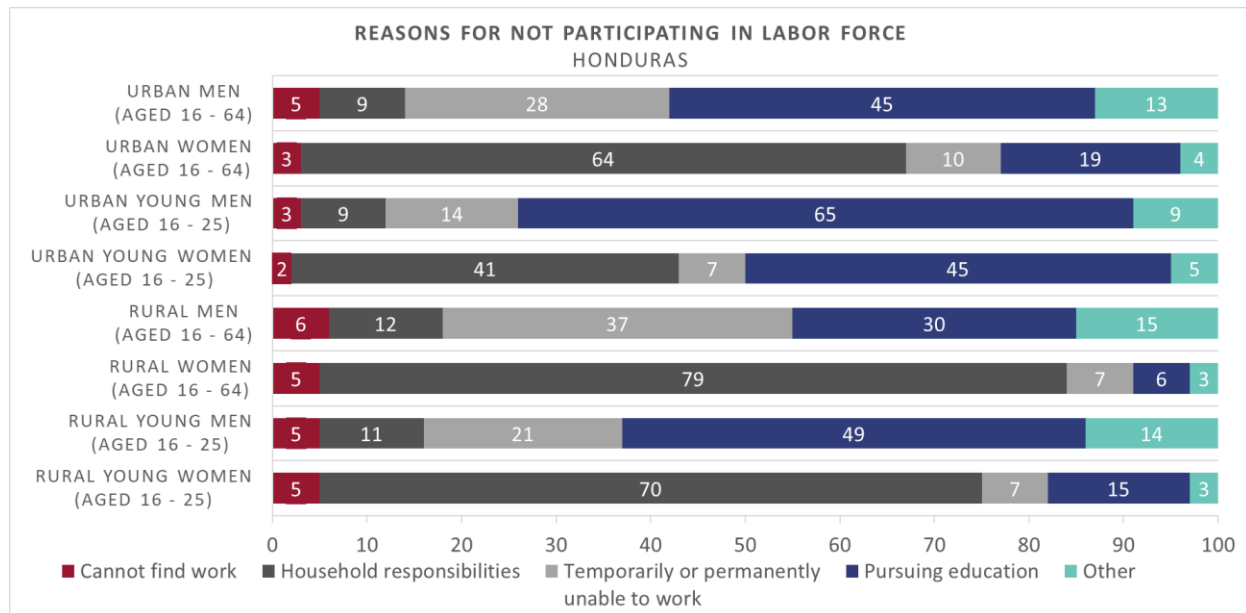


### 3.2. BARRIERS TO LABOR FORCE PARTICIPATION

For those who are not working and not looking for work, we identify the main reasons for not participating in the labor force. Results are presented for key sub-groups, allowing us to compare barriers to labor force participation across different populations.

Overall, the main barriers to labor force participation for respondents from Honduras include household responsibilities (including household chores, care for dependents, and family responsibilities); education; and being temporarily or permanently unable to work due to reasons such as advanced age or disability. Differences by gender are also informative: while the primary reason for not participating in the labor force for most groups of men was education, suggesting they will enter the labor force eventually, the main reason for Honduran women was household responsibilities, with the sole exception of urban young women, for whom household responsibilities (41 percent) follows behind education (45 percent).

Figure A3.10: Key Labor Force Participation Barriers in Honduras, 2019



Source: 2019 EPHPM

### 3.3. LABOR MARKET SECTORS WITH GREATEST WAGE AND JOB GROWTH

By combining data from multiple EPHPM surveys, we are able to calculate wage and job growth by economic sector. Unfortunately, EPHPM data could only be obtained for 2018 and 2019, meaning the growth trend can only be calculated over a one-year period. Given the short time period, readers should exercise caution in the interpretation of the results, given that year-to-year variation can be noisy, and is not necessarily reflective of longer-term trends.

The first panel of Table A3.3 shows the top three industries for job and wage growth among those with formal employment, defined as in Figure A3.9. The top industry for formal employment wage growth was finance and insurance, where wages grew by 25.0 percent. The top industry for formal job growth was health care and social services, with growth of 36.6 percent. The next two panels in the table show the top industries for job and wage growth among those with informal employment, for urban and rural areas, respectively. In urban areas, water and sanitation represented the top industry for informal employment wage growth (30.6 percent), while the top sector for informal job growth was

health care and social services (9.0 percent). In rural areas, the top informal sector for wage growth was transportation and storage (48.8 percent), while construction (10.3 percent) saw the highest job growth.

The remaining panels of Table A3.3 show the top three sectors for wage and job growth, for all respondents and by level of education, including both formal and informal sectors of the economy together. We define secondary education as having completed *secundaria*, the Honduran equivalent of senior high school, typically completed at age 18. The data show that approximately 85 percent of the rural population has less than a secondary education, compared to approximately 59 percent of the urban population.

Top economic sectors for wage growth vary by level of education, with education and wholesale or retail trade figuring among the top spots for more educated Hondurans, with either a secondary or tertiary education, and hotel and food service among the top spots for Hondurans with less than a secondary education in both urban and rural areas. Although the finance-and-insurance sector takes the top spot for overall wage growth, the results also suggest that employees in this industry are almost entirely Hondurans with tertiary levels of education, and it is this group that is driving the wage growth in the sector.

The economic sectors with the highest job growth between 2018 and 2019 are water and sanitation (30.5 percent), and health care and social services (16.9 percent). In fact, these are the only two sectors that showed job overall growth over the period in Honduras; the third best sector for job growth between 2018 and 2019, Administrative Services, maintained an approximately even number of jobs, with a small estimated loss of 0.6 percent.

Table A3.3: Honduran Labor Market Sectors with Greatest Wage and Job Growth (2018-2019), by Education Level

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2018 and 2019				Top Economic Sectors with Highest <u>Job</u> Growth Between 2018 and 2019			
Sector	Average Wage 2018, Lempira (USD)	Average Wage 2019, Lempira (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2018	Number Employed in 2019	Average Job Growth
Respondents with Formal Employment							
Finance and Insurance	L11,395 (US\$476)	L14,240 (US\$581)	25.0%	Health Care and Social Services	34,060	46,535	36.6%
Agriculture, livestock, forestry, and fishing	L8,720 (US\$364)	L10,457 (US\$427)	20.0%	Professional, scientific, and technical services	9,602	13,082	36.2%
Information and Communications	L11,918 (US\$498)	L13,991 (US\$571)	17.4%	Construction	18,128	24,029	32.6%
Respondents with Informal Employment - Urban							
Water and Sanitation	L2,843 (US\$119)	L4,235 (US\$173)	49.0%	Health Care and Social Services	25,332	27,606	9.0%

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2018 and 2019				Top Economic Sectors with Highest <u>Job</u> Growth Between 2018 and 2019			
Sector	Average Wage 2018, Lempira (USD)	Average Wage 2019, Lempira (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2018	Number Employed in 2019	Average Job Growth
Agriculture, livestock, forestry, and fishing	L3,349 (US\$140)	L4,788 (US\$195)	43.0%	Administrative Services	32,242	34,946	8.4%
Information and Communications	L6,450 (US\$270)	L8,232 (US\$336)	27.6%	Agriculture, livestock, forestry, and fishing	126,909	133,674	5.3%
Respondents with Informal Employment - Rural							
Transportation and Storage	L4,643 (US\$194)	L6,905 (US\$282)	48.8%	Construction	65,081	71,765	10.3%
Manufacturing	L2,705 (US\$113)	L3,346 (US\$137)	23.7%	Wholesale or retail trade, including vehicle and motorcycle repair	184,682	190,473	3.1%
Agriculture, livestock, forestry, and fishing	L1,942 (US\$81)	L2,333 (US\$95)	20.1%	Hotel and Food Service	36,973	37,981	2.7%
All Respondents (including formal and informal employment)							
Finance and Insurance	L10,370 (US\$434)	L13,806 (US\$563)	33.1%	Water and Sanitation	17,205	22,455	30.5%
Information and Communications	L8,565 (US\$358)	L11,123 (US\$454)	29.9%	Health Care and Social Services	70,728	82,676	16.9%
Agriculture, livestock, forestry, and fishing	L2,559 (US\$107)	L3,319 (US\$135)	29.7%	Administrative Services	68,710	68,271	-0.6%
Respondents with PhD, Masters, and Other Advanced Degrees (including formal and informal employment)							
Finance and Insurance	L11,604 (US\$485)	L16,048 (US\$655)	38.3%	Health Care and Social Services	22,405	27,647	23.4%
Wholesale or retail trade, including vehicle and motorcycle repair	L10,493 (US\$439)	L13,339 (US\$544)	27.1%	Manufacturing	36,092	37,987	5.3%

Top Economic Sectors with Highest <u>Wage</u> Growth Between 2018 and 2019				Top Economic Sectors with Highest <u>Job</u> Growth Between 2018 and 2019			
Sector	Average Wage 2018, Lempira (USD)	Average Wage 2019, Lempira (USD)	Average Nominal Wage Growth	Sector	Number Employed in 2018	Number Employed in 2019	Average Job Growth
Education	L12,808 (US\$536)	L14,056 (US\$574)	9.7%	Construction	11,348	11,724	3.3%
Respondents with Secondary Education (including formal and informal employment)							
Agriculture, livestock, forestry, and fishing	L3,990 (US\$167)	L5,491 (US\$224)	37.6%	Agriculture, livestock, forestry, and fishing	104,182	124,535	19.5%
Education	L8,649 (US\$361)	L11,158 (US\$455)	29.0%	Health Care and Social Services	35,761	41,134	15.0%
Wholesale or retail trade, including vehicle and motorcycle repair	L6,310 (US\$264)	L7,431 (US\$303)	17.8%	Transportation and Storage	42,347	47,943	13.2%
Respondents with Less than Secondary Education (Urban - including formal and informal employment)							
Agriculture, livestock, forestry, and fishing	L3,381 (US\$141)	L4,902 (US\$200)	44.9%	Public Administration, Defense, and Social Security	11,003	15,746	43.1%
Hotel and Food Service	L5,262 (US\$220)	L6,344 (US\$259)	20.6%	Administrative Services	26,080	27,707	6.2%
Administrative Services	L6,237 (US\$261)	L7,250 (US\$296)	16.2%	Water and Sanitation	10,249	10,807	5.4%
Respondents with Less than Secondary Education (Rural - including formal and informal employment)							
Transportation and Storage	L4,523 (US\$189)	L7,602 (US\$310)	68.1%	Wholesale or retail trade, including vehicle and motorcycle repair	147,032	152,366	3.6%
Hotel and Food Service	L2,690 (US\$113)	L3,273 (US\$134)	21.7%	Construction	58,646	59,255	1.0%
Wholesale or retail trade, including vehicle and motorcycle repair	L3,684 (US\$154)	L4,417 (US\$180)	19.9%	Manufacturing	105,163	99,477	-5.4%

Source: 2018 and 2019 EPHPM

Note: Exchange rates come from the International Monetary Fund. Average exchange rate for 2018 = 23.9 Lempira per US Dollar. Average exchange rate for 2019 = 24.51 Lempira per US Dollar.

### 3.4. KEY ECONOMIC SECTORS

In Figure A3.11, we show the top economic sectors employing the greatest share of each population sub-group, along with the average monthly salary.

In 2019, the main economic activity among rural men and rural young men in Honduras was agriculture, livestock, forestry, and fishing. The industry employed 70.5 percent of rural men and 69.5 percent of rural young men, with average monthly salaries of US\$114 and US\$105 lempira, respectively. For rural women and rural young women in Honduras, wholesale or retail trade constituted the most main sector for employment. The industry employed 27.7 percent of rural women and 21.4 percent of rural young women, with average monthly salaries of US\$168 and US\$153, respectively. The main economic activity for all urban sub-groups was wholesale or retail trade.

### 3.5. EMPLOYER NEEDS AND BARRIERS TO FINDING TALENT

We use the Hospitality Sector Pilot Establishment Survey conducted by AIR to identify the positions that are hardest to fill, along with the reasons these positions are difficult to fill, as identified by managers.

Among the 437 businesses responding to the survey, the most common positions among businesses responding to the Hospitality Sector Pilot Establishment Survey in Honduras were: Receptionist ( $n=253$ ), housekeeping ( $n=204$ ), cleaning staff ( $n=139$ ), cook ( $n=88$ ), waiter ( $n=67$ ), and kitchen assistant ( $n=39$ ). For each position, we calculate a hard to fill rate, defined as the sum of employees in the position among employers that consider the position “hard to fill”, divided by the sum of employees in the position among all employers in the survey. Figure A3.12 shows that the positions that are the hardest to fill are cooks (53 percent), receptionists (46 percent), and waiters (46 percent).

Figure A3.11: Highest Employing Labor Market Sectors in Honduras and Associated Average Salaries

MEN				WOMEN			
GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)	GROUP	TOP EMPLOYING INDUSTRY	Percent Employed in Industry	Average salary (USD)
Urban	Wholesale or retail trade, including vehicle and motorcycle repair	21%	\$348	Urban	Wholesale or retail trade, including vehicle and motorcycle repair	25%	\$275
Urban Youth*	Wholesale or retail trade, including vehicle and motorcycle repair	22%	\$277	Urban Youth*	Wholesale or retail trade, including vehicle and motorcycle repair	23%	\$280
Rural	Agriculture, livestock, forestry, and fishing	71%	\$112	Rural	Wholesale or retail trade, including vehicle and motorcycle repair	28%	\$164
Rural Youth*	Agriculture, livestock, forestry, and fishing	70%	\$102	Rural Youth*	Wholesale or retail trade, including vehicle and motorcycle repair	21%	\$149

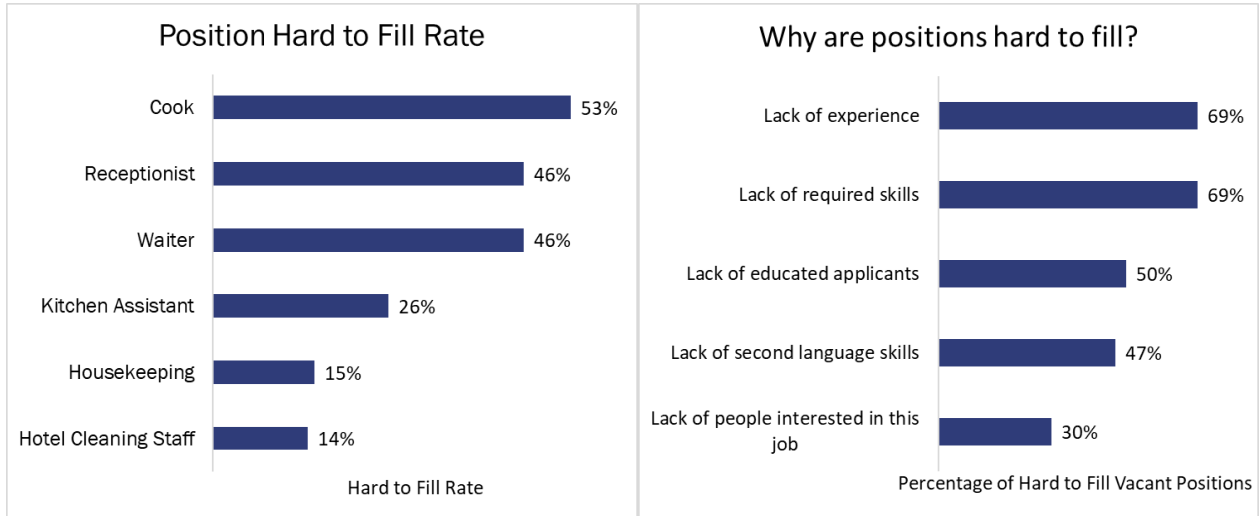
\*Aged 16-25

Source: 2019 EPHPM.

Note: Exchange rates come from the International Monetary Fund. Average exchange rate for 2019 = 24.51 Lempira per US Dollar.

Employers were also asked why positions were hard to fill. Among all hard to fill positions, the most common reasons given were a lack of candidates with experience in the position (69 percent), applicants that lack the required skills for the position (69 percent), and a lack of educated applicants (50 percent), followed by a lack of applicants with second language skills (47 percent).

Figure A3.12: Employer Needs and Barriers to Finding Talent in the Honduras Hospitality Sector



Source: 2019 Honduras Hospitality Sector Pilot Establishment Survey (AIR).