Data Description:

Gender Equity Program (GEP) Endline Survey

## Description of Activity

The Gender Equity Program (GEP), implemented by the Aurat Foundation between August 2010 and August 2017, was the United States Agency for International Development (USAID) flagship gender activity in Pakistan. GEP aimed to close the gender gap in Pakistan by actively supporting women’s economic, political, and social development. The program provided grants to government and non-government entities to implement gender-focused activities at the district level and supported work on national and provincial-level pro-women legislation.

## Purpose of GEP Endline Survey

The 2018 endline survey used the same instrument and sampling approach as the 2011 baseline. It collected data on knowledge, attitudes, and behaviors on a range of gender equity topics (i.e., women’s rights, democratic rights, inheritance rights, institutions of justice, women’s empowerment and decision making, women’s mobility, domestic violence, and gender-based violence). Comparing endline survey results to those of the 2011 baseline served to assess if and how attitudes and behaviors have changed since 2011.

## Survey Data

Variable names in the datasets and codebooks align with question numbers in the questionnaires.

## Sampling Design

The 2018 endline surveyed 6,048 men and women in 504 localities selected from Pakistan’s four provinces and Gilgit-Baltistan (GB). To ensure comparability with the 2011 baseline, the 2018 survey used the same data collection instrument as the 2011 survey and applied a similar sampling approach.

Samples for both the 2011 and 2018 surveys were selected through multi-stage stratified random sampling. Strata included province, urban and rural regions, and gender. In the provinces, the study team selected 28 total districts: nine districts from Punjab; six districts each from Sindh, Khyber Pakhtunkhwa (KP), and Balochistan; and one district (Gilgit) from GB. The sample was distributed equally across the selected districts, representing districts with high, medium, and low education inequality index values. In each selected district, the survey sampled urban and rural areas in proportion to the population distribution and was evenly distributed between men and women (Table 1).

Table 1: Sample Distribution by District

| **Province/**  **Region** | **Education Inequality Index Group** | **Education Inequality Index Value in 2011** | **Districts** | **Number of Respondents** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **2011** | | **2018** | |
| **Rural** | **Urban** | **Rural** | **Urban** |
| Punjab | Low | 0.74 | Rawalpindi | 256 | 144 | 84 | 132 |
| 0.68 | Lahore | 256 | 144 | 36 | 180 |
| 0.58 | Faisalabad | 257 | 143 | 96 | 120 |
| Medium | 0.53 | Sheikhupura | 256 | 144 | 120 | 96 |
| 0.49 | Khushab | 258 | 142 | 156 | 60 |
| 0.47 | Multan | 256 | 144 | 96 | 120 |
| High | 0.44 | Jhang | 256 | 144 | 156 | 60 |
| 0.40 | Pakpattan | 256 | 144 | 168 | 48 |
| 0.40 | Muzaffargarh | 256 | 148 | 168 | 48 |
| KP | Low | 0.60 | Abbottabad | 347 | 55 | 120 | 96 |
| 0.39 | Mardan | 270 | 131 | 132 | 84 |
| Medium | 0.38 | Peshawar | 161 | 239 | 96 | 120 |
| 0.35 | Bannu | 384 | 32 | 156 | 60 |
| High | 0.26 | D.I. Khan | 232 | 167 | 156 | 60 |
| 0.23 | Bonair | 364 | 38 | 216 | 0 |
| Sindh | Low | 0.48 | Hyderabad | 255 | 145 | 36 | 180 |
| 0.35 | Khairpur | 253 | 147 | 144 | 72 |
| Medium | 0.33 | Sanghar | 256 | 144 | 144 | 72 |
| 0.32 | Larkana | 254 | 145 | 120 | 96 |
| High | 0.28 | Badin | 266 | 134 | 144 | 72 |
| 0.16 | Jacobabad | 256 | 145 | 144 | 72 |
| Balochistan | Low | 0.54 | Quetta | 269 | 132 | 108 | 108 |
| 0.28 | Panjgur | 408 | 0 | 156 | 60 |
| Medium | 0.27 | Chagi | 352 | 48 | 180 | 36 |
| 0.23 | Lasbela | 331 | 84 | 120 | 96 |
| High | 0.19 | Loarlai | 350 | 60 | 156 | 60 |
| 0.10 | Naseerabad | 271 | 140 | 156 | 60 |
| GB | -- |  | Gilgit | 361 | 42 | 120 | 96 |
| **Total Number of Respondents** | | | | 7,947 | 3,325 | 3,684 | 2,364 |
| **Total Number of Clusters** | | | | 498 | 278 | 307 | 197 |

The first stage of sampling used districts as the primary sampling unit.[[1]](#footnote-2) In the second stage of sampling, within each district, the enumeration blocks listed in the population census served as secondary sampling units. In the third stage of sampling, within each sampled enumeration block, enumerators selected a predetermined number of households through a random walk. Finally, from each sampled household, enumerators then used a Kish Grid to randomly select one individual who met the screening criteria.

## Sampling Stage 1: Selection of Districts

The 2011 survey used an education-inequality index, based on data from the 2006-07 Pakistan Social and Living Standard Measurement Survey as a proxy for female underdevelopment. To calculate the education inequality index, analysts first calculated separate education indices for men and women in each district. These district- and sex-specific index represented the sum of 0.67 times the percentage of the male/female population that was literate and 0.33 times the male/female enrollment rate for the 5-24 years age cohort. The district education inequality index was then calculated according to the following formula[[2]](#footnote-4):

Analysts classified districts into three strata representing low, medium, and high levels of education inequality, respectively. The 2011 survey selected two districts from each education group in all provinces except Punjab, where it selected three districts from each group to control for greater heterogeneity across districts. To make the results comparable across the two surveys, the 2018 survey selected the same 28 districts surveyed in 2011 (listed in Table 1).

## Sampling Stage 2: Selection of Clusters

The second stage of sampling selected geographic clusters within each sampled district using the probability proportional to size (PPS) method, based on the number of households within each cluster.[[3]](#footnote-6) The 2011 survey used the list of “circles” in urban areas and “villages/mouzas/dehs” in rural areas, compiled for the 1998 population census, as a sampling frame for selecting clusters through PPS. The 2018 survey used the list of enumeration blocks compiled for the 2017 population census, in both urban and rural areas, as a sampling frame for selecting clusters through PPS.

## Sampling Stage 3: Selection of Households and Individuals

In the third sampling stage, enumerators selected a predetermined number of households within each cluster through a random walk. To select a starting point for the random walk, the enumeration team developed a list of important landmarks (schools, mosques, etc.) and chose one randomly. Starting from the third house on the right side of the chosen landmark, enumerators selected sample households using the right-hand rule*.*[[4]](#footnote-7)

In the 2011 survey, enumeration teams selected 12 household from each cluster in urban areas and 16 households from each cluster in rural areas. In the 2018 survey, enumeration teams selected 12 households from each sampled cluster in both urban and rural areas. In both surveys, in each sampled cluster, the enumeration team interviewed an adult man in half the households and an adult woman in half the households, using a Kish Grid to select individuals within each household. Female enumerators interviewed female respondents and male enumerators interviewed male respondents.

## Sampling Weights

Samples, even when well designed, are not generally a perfect a reflection of the reference population. The departure of the sample distribution from the population distribution leads to bias in sample results. To correct the results for deviation between sample and population distributions, the analysts applied stratification weights to account for over- or under-sampling of provinces and urban and rural areas, according to the following process:

1. To calculate the distributional share of each stratum (10 combinations of 5 provinces and urban or rural areas) in the *sample*, analysts divided the total number of households in each stratum by the total sample of households (i.e., 6048).
2. To calculate the distributional share of each stratum in the *population*, analysts divided the total number of households in each stratum by the total number of households listed in the 2017 population census.
3. To calculate the stratification weights, analysts then calculated the ratio of the population share and sample share of each stratum (i.e., divided the distributional share of each stratum in the population by its distributional share in the sample).

1. One exception in the sampling design is Gilgit, where the sample was drawn at the tehsil level. [↑](#footnote-ref-2)
2. Due to the non-availability of relevant data, the index is not calculated for Gilgit. [↑](#footnote-ref-4)
3. The PPS method implies that the probability of an individual household being selected is proportional to the number of households in the cluster. [↑](#footnote-ref-6)
4. The ‘Right Hand Rule’ is a method for selecting sample households by which enumerators contact every household falling to the right side of the street until a successful interview can be conducted. The enumeration team then skips a specific number of households before interviewing the next household. This rule negates interviewer bias in household selection. [↑](#footnote-ref-7)