Notes about data collection and sampling

The Labor, Migration and Skills survey consists of three identical household surveys conducted in Central Asia during the summer of 2013 by GIZ and the World Bank. It is representative at the national, regional (Oblast), and urban/rural level.

The survey collects comprehensive information not typically captured by traditional household surveys. It includes two distinct instruments: a core questionnaire and a skills questionnaire. The core questionnaire contains modules on education, employment, migration, health expenditure, remittances, government transfers, financial services, subjective poverty, and housing conditions, as well as a complete household expenditure module. The core questionnaire concludes with the random selection of one or two household members aged 15 to 64 to whom the skill questionnaire is administered. The random selection is based on a random number table (Kish grid). The second part of the survey, the skills questionnaire, contains detailed modules on labor and work expectations, migration and preparation for migration, language skills, and technical skill training. It also includes a self-assessment of technical skills and knowledge, self-reported use of skills, as well as a battery of cognitive and non-cognitive test modules.

The cognitive and non-cognitive test modules of the skills questionnaire are based on World Bank STEP surveys, which were developed with the support of a multi-disciplinary panel of experts. This ensures data comparability with results from World Bank STEP surveys, in Armenia, Bolivia, Colombia, Georgia, Ghana, Laos, Sri Lanka, Vietnam, and the Yunnan Province in China. The skills assessment modules benefited from the input of experts in psychology, skills assessment, education, and labor markets. Qualitative testing and pre-pilots were carried out to test the instruments. They helped winnow out some questions, fine-tune others, and organize the modules so the questionnaire could be administered efficiently and consistently.

The sample size of the core questionnaire is 6,300 households with a total of 35,770 individuals. Given that either one or two individual per household was randomly selected to partake in the skills questionnaire, this sample consists of 7,929 individuals. The survey was conducted from July to September, 2013. Also, in Tajikistan, 500 households were randomly selected to be administered diaries instead of the TLSS expenditure questionnaire.

Questionnaire sections

Visit 1: (All) Household Members	Visit 2: Selected Household Member
1. Demographic Profile card	1. Labour Conditions
2. Education	2. Labour market expectations
3. Education Expenditure	3. Russian Language skills
4. Immigration	4. Return migrants pre-departure preparation
5. Employment	5. Future migrants' pre-departure preparation
6. Labor Market	6. Skills acquisition pre-departure Questions about skills acquisition for both future migrants and return migrants
7. Work Migration Cycle	7. Most recent Technical skill Training
8. Most recent migration event	8. Technical skills: reading and writing
9. Remittances and gifts from non household members	9. Workplace skills
10. Migration Intent	10. Non-Cognitive Skills: Part A
11. Health expenditure	11. Non-Cognitive Skills: Part B
12. Financial services	12. Cognitive Skills: Memory
13. Subjective Poverty	13. Cognitive Skills: Language
14. Habits and adaptation	14. Cognitive Skills: Text Comprehension A
15. Food Consumption	15. Cognitive Skills: Text Comprehension B
16. Non-Food Consumption	16. Cognitive Skills: Table Comprehension
17. Other Non-Food Consumption	17. Cognitive Skills: Publicity Comprehension
18. Large items of non-food consumption	18. Cognitive Skills: Graph Comprehension
19. Fuel	
20. Payment for utilities and electricity	
21. Dwelling	
22. Energy	
23. Availability of utility equipment	
24. Gifts	
25. Government transfers	
26. Subjective budget – remittances	
27. Kish List (selection of member for follow up survey)	

Uzbekistan

The survey on labour, skills and migration in Uzbekistan was conducted between the 30th of July and the 6th of September by a team of 75 fieldworkers from the firm Expert-fikri.

The sample consists of 1,500 households with 8,622 individuals, stratified at the Oblast (region) and Urban/Rural level. The sampling strategy is an adaptation of the sampling methodology commonly used in Uzbekistan.

Oblast / Strata	PSU Sample	PSU National	Households Sample	Households National	Population Sample	Population National
1 Karakalpakstan Rural	2	1128	44	150312	244	811687
1 Karakalpakstan Urban	2	197	42	151911	236	820320
2 Andijan Rural	5	456	97	220497	580	1190686
2 Andijan Urban	2	396	41	251548	208	1358361
3 Bukhara Rural	3	1471	61	183362	355	990156
3 Bukhara Urban	2	208	26	115257	135	622389
4 Jizzakh Rural	2	518	42	109250	304	589948
4 Jizzakh Urban	1	149	18	97564	99	526843
5 Kashkadarya Rural	5	1046	106	274143	642	1480373
5 Kashkadarya Urban	2	320	36	210310	198	1135675
6 Navoi Rural	2	586	26	79759	151	430696
6 Navoi Urban	1	124	18	77938	71	420863
7 Namangan Rural	4	403	77	148097	405	799724
7 Namangan Urban	2	424	46	270147	264	1458795
8 Samarkand Rural	6	1877	124	362714	745	1958655
8 Samarkand Urban	2	432	45	214888	286	1160397
9 Surkhandarya Rural	4	865	91	242059	537	1307121
9 Surkhandarya Urban	1	244	22	142199	130	767872
10 Syrdarya Rural	2	258	25	77716	164	419667
10 Syrdarya Urban	1	87	12	54580	80	294734
11 Tashkent Rural	4	887	81	239390	535	1292704
11 Tashkent Urban	3	487	54	239489	207	1293243
12 Fergana Rural	6	1020	118	235569	721	1272073
12 Fergana Urban	2	552	48	333809	297	1802569
13 Khorezm Rural	3	559	64	191115	399	1032023
13 Khorezm Urban	1	178	19	98067	121	529563
14 City of Tashkent Urban	5	474	117	413760	508	2234306
Total	**75**	**15346**	**1500**	**5185450**	**8622**	**28001443**

The sample is grouped into PSUs, which are geographical areas of a walkable size. The 75 PSUs sampled (among the 15,000+ in the whole country) are divided according to the population size of each of the 27 urban and rural regions. Then within each urban and rural region, each PSU is randomly selected with a probability proportional to its size. Exceptionally in the sample for Uzbekistan, the sizes of the PSUs have been adjusted ex post for each region to better represent its population importance; PSUs nevertheless contain on average 20 households.

Within each PSU, the households are selected using a geographical sampling procedure. This procedure consists of generating a random point using a numbered grid over a map. From this starting point within the PSU, one out of every 5 households is interviewed, following a systematic route designed for each PSU.

The total number of either refusals or absences noted after 3 attempts amounts to 1,067 households. Each missing and refusal was replaced with another household by extending the geographical sampling procedure within the PSU.

4 PSUs out of 75 were replaced for various accessibility reasons, and neighbouring PSUs showing similar characteristics were selected to replace them.

Within each household, two sections of the questionnaire were directed at two different categories of individuals within the household. Sometimes, the same person responded to both sections. First, the most knowledgeable person of the household was asked the main part of the questionnaire, which includes questions regarding each household member for their education, health spending and labour and migration. This main part also includes a complete household expenditure module, questions about remittances, government transfers, financial services, subjective poverty and questions about the housing conditions.

The second part of the questionnaire was asked to a randomly chosen adult between the age of 15 to 64 who is not currently a migrant, using a random number table (Kish grid) to ensure the randomness of the selection. Provided it was not possible to reach the person selected after 3 attempts, another person was selected using the same random procedure.

The second part of the questionnaire included detailed modules about labour and work expectations, migration and preparation for migration, language skills, and technical skill training. It also included a self-assessment of technical skills and knowledge, a non-cognitive, and a cognitive test. Unfortunately, the 7 language questions of the cognitive skills test are unusable because of translation.

Exceptionally in the sample for Uzbekistan, households in which there were no individuals between the age of 15 to 64 were completely excluded.

Finally, control visits by independent consultants ensured the quality of the data collection exercise. A total of 156 home visits and 225 phone calls, with at least 3 controls in each PSU allowed an early detection of various issues and lead to a successful resolution of the identified problems.

Tajikistan

The survey on labour, skills and migration in Tajikistan was conducted between June 22 to August 2 by a team of 150 fieldworkers from the firm Zerkola.

The sample consists in 2 independent segments. One is part of a panel with the TLSS survey of 2012 collected by the survey firm Panorama (2,000 houeholds), which was conducted as part of the impact assessment of the reform of social protection pilot of Yevon and Istaravshan. The second segment of the sample is the Dushanbe Booster (1,300 households in Dushanbe only), needed for a deeper investigation about energy usage.

In total, the sample consists of 3,300 households with 20,142 individuals, representative at the Oblast (region) and Urban/Rural level. The sampling strategy is an adaptation of the two stage proportional sampling methodology used in Tajikistan for the TLSS in 2007, 2009 and 2012.

The Dushanbe Booster consists in 1,300 housheolds, in 65 PSUs dvided among the 4 districts.

Sampling Strata	HH per strata*	Population per strata	Proportion of HHs per strata	Count of PSU to generate
Somoni district	26184	128300	17.7%	(11.506) 11
Sino district	56979	279200	38.5%	25
Firdavsi district	35469	173800	24.0%	16
Shohmansur district	29286	143500	19.8%	13
TOTAL	**147 918**	**724 800**	**100%**	**65**

* HH per strata is estimated using an average HH size of 4.9. Total for Dushanbe is actual count

The main 2012 sample consists in 2,000 houshelds:

Oblast / Strata	PSU Sample	Households Sample	Households National	Population Sample	Population National
Dushanbe Urban	11	220	138588	1004	623102
GBAO Urban	3	60	5111	306	25996
GBAO Rural	6	120	25851	827	174135
Sughd Urban	9	180	109941	980	552239
Sughd Rural	21	420	259942	2658	1634499
Khatlon Urban	6	120	75088	751	449642
Khatlon Rural	23	460	281498	3459	2079821
RRS Urban	5	100	40690	435	178001
RRS Rural	16	320	196055	2460	1489705
Total	**100**	**2000**	**1132764**	**12880**	**7207138**

* Count of individuals based on 2012 data.

These same 2000 households have been interviewed again. In the case a household was not available, the closest neighbour was interviewed. 84% of the households of 2012 have been re-interviewed.

All 165 PSUs consist of 20 households. The sample as a whole is stratified according to Oblasts and Urban/Rural, and RRS Urban has been oversampled to ensure a sufficient number of observations to compare results across regions.

The sample of PSU was selected with a probability proportional to the number of households using the census of 2011, with the assistance of the National Statistics agency, in both segments of the sample. In a second stage, the National Statistics Agency also used the census to select the 20 households in each of the PSUs.

For every PSU (in 2012 and in the Dushanbe Booster), a list of 20 households was generated for the main sample, and a list of 5 households was created for the replacements (10 replacemetns in 2013 for the Dushanbe Booster). Replacements were allowed following the sequential number only after 3 visits or firm refusal.

No PSUs needed to be replaced in 2013 for inaccessibility reasons, but

For a randomly selected subset of 500 questionnaires, the expenditure section was replaced with diaries, allowing to test and calibrate the new Houshehold Budget Survey, and allowing the comparison between the TLSS and the HBS in 2013.

Within each household, two sections of the questionnaire were directed at two different categories of individuals within the household. Sometimes, the same person responded to both sections. First, the most knowledgeable person of the household was asked the main part of the questionnaire, which includes questions regarding each household member for their education, health spending and labour and migration. This main part also includes a complete household expenditure module, questions about remittances, government transfers, financial services, subjective poverty and questions about the housing conditions.

The second part of the questionnaire was asked to a randomly chosen adult between the age of 15 to 64 who is not currently a migrant, using a random number table (Kish grid) to ensure the randomness of the selection. Provided it was not possible to reach the person selected after 3 attempts, another person was selected using the same random procedure.

The second part of the questionnaire included detailed modules about labour and work expectations, migration and preparation for migration, language skills, and technical skill training. It also included a self-assessment of technical skills and knowledge, a non-cognitive, and a cognitive test. Unfortunately, the 7 language questions of the cognitive skills test are unusable because of translation.

Kyrgyzstan

The survey on labour, skills and migration in Kyrgyzstan was conducted between the 30th of July and the 6th of September by a team from the firm El-Pikir.

The sample consists of 1,500 households with 7,005 individuals, stratified at the Oblast (region) and Urban/Rural level.

Urban or Rural	Geographical Units	Population	Number of households (HH)	No of PSU
Urban	BISHKEK city	835,743	228,657	10
Urban	OSH city	258,111	57,392	3
Urban	ISSYK-KUL OBLAST/REGION	119,694	33,453	2
Urban	JALAL-ABAD OBLAST/REGION	187,734	43,263	3
Urban	NARYN OBLAST/REGION	34,822	7,976	2
Urban	BATKEN OBLAST/REGION	84,587	17,817	2
Urban	OSH OBLAST/REGION	70,272	12,488	2
Urban	TALAS OBLAST/REGION	32,886	8,240	2
Urban	CHUI OBLAST/REGION	121,403	41,285	3
Rural	ISSYK-KUL OBLAST/REGION	329,220	72,560	5
Rural	JALAL-ABAD OBLAST/REGION	832,606	144,907	9
Rural	NARYN OBLAST/REGION	222,946	43,112	3
Rural	BATKEN OBLAST/REGION	344,049	62,190	4
Rural	OSH OBLAST/REGION	1,104,248	187,805	12
Rural	TALAS OBLAST/REGION	193,893	35,973	2
Rural	CHUI OBLAST/REGION	749,999	188,805	11
	ALL Kyrgyzstan	5,522,213	1,185,923	75

The sample is grouped into 75 PSUs of 20 households each. Each PSU is randomly selected with a probability proportional to its size.

Within each PSU, the households are selected using a geographical sampling procedure. 4 random point wrre generated using a numbered grid over a map. From this starting point within the PSU, one out of every 4 households is interviewed, following a systematic route.

Within each household, two sections of the questionnaire were directed at two different categories of individuals. Sometimes, the same person responded to both sections. First, the most knowledgeable person of the household was asked the main part of the questionnaire. The second part of the questionnaire was asked to a randomly chosen adult between the age of 15 to 64 who is not currently a migrant, using a random number table (Kish grid) to ensure the randomness of the selection. Provided it was not possible to reach the person selected after 3 attempts, another person was selected using the same random procedure.

The second part of the questionnaire included detailed modules about labour and work expectations, migration and preparation for migration, language skills, and technical skill training. It also included a self-assessment of technical skills and knowledge, a non-cognitive, and a cognitive test. Unfortunately, the 7 language questions of the cognitive skills test are unusable because of translation.