

CAPI data collection with Survey Solutions

Sergiy Radyakin
sradyakin@worldbank.org

May, 2018

General instructions:

For each exercise name your questionnaire: **Gr20ExNN-YourFirstName-YourLastName**, where NN is the number of the exercise. For example, Max Mustermann may name the questionnaire for his third exercise as **Gr20Ex03-Max-Mustermann**. Some assignments do not create or modify questionnaires.

Read the whole exercise before starting, not just the first paragraph!

Verify your questionnaire in the Tester application. All exercises require Internet connectivity.

Keep track of the sites you are using, accounts you create and their respective passwords.

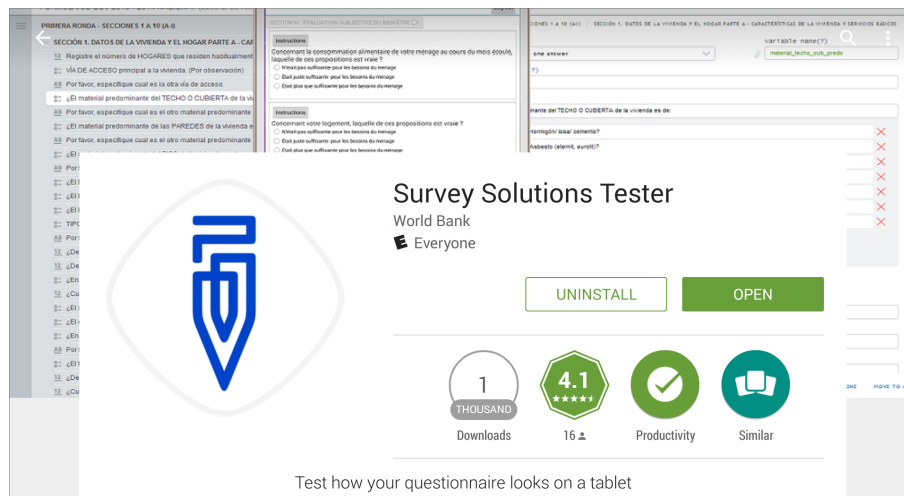
Refer to your notes and try to work independently. After finishing each exercise, discuss your solution with a colleague.

Exercise 0: Setup.

Create an account at the Designer website: <https://designer.mysurvey.solutions>. Note that registration requires an email account. Use your existing account or create a new one from any free email service provider popular in your area. Upon registration you will receive a link that you need to click before proceeding.

Your password must be at least 10 characters long, include 1 digit, 1 small letter, and 1 capital letter.

Install the Survey Solutions Tester application to your tablet. Use the same credentials to login.



Exercise 1: Practice different question types.

You are in charge of designing a questionnaire for a survey of housing conditions. The survey will be collecting information on the dwelling and available infrastructure of 3,000 households randomly selected from the whole population. Each household is identified by its address, and sample number.

Develop a questionnaire in the Survey Solutions Designer to acquire at least the following information:

- Date of the visit;
- Number of residents; number of males (including children); number of females (including children); number of children 0..15; number of children 0..5;
- Ownership status of the dwelling;
- Year the dwelling was built;
- Year of last major renovation;
- Construction material of the walls, the roof, etc.;
- Sources of drinking water that are available to the residents;
- Most commonly used source of drinking water;
- Expenditures on the following utilities: electricity, water supply, heating, security, maintenance;
- Distance to the nearest available service: post office, bank, primary school, hospital, bus station;
- Up to four major problems in the area;
- Which reforms or interventions the residents see as most efficient in solving the named problems.

When developing the questionnaire:

- Decide on the suitable structure of the questionnaire: create sections and subsections. Your questionnaire should contain more than one section.
- Decide yourself on the suitable categories for categorical questions. For example, most commonly used construction materials in your area.
- Decide on the units of measurement for continuous variables.
- In which questions you anticipate a high rate of “don’t know” answers? How should the interviewers handle this situation?
- Write interviewer instructions for questions where you anticipate the interviewers may have difficulties in the field.
- On a separate sheet of paper track all the difficulties you’ve encountered in the process.

Exercise 2: Questionnaire logic: Skips

Create a copy of the questionnaire you created in exercise 1. Rename it to *Gr20Ex02-YourFirstName-YourLastName*.

In the second exercise introduce the following changes:

- The question on the year the dwelling was built should be asked only if the dwelling is owned by the residents;
- The question on the year of the last major renovation should be asked only if the building is older than 20 years;
- Most commonly used source of drinking water should be asked only if more than one source was mentioned (use `q.Length` to get the number of selections in a multiple choice question `q`);
- Distance to school should be asked only if there are children aged 0..15 in the household.

1. Test your questionnaire in the tester.
2. Explore the different paths of going through the interview.
3. Draw these paths as a flowchart on a separate sheet.

Exercise 3: Questionnaire sharing

Share your questionnaire with a colleague (you will need to know your colleague's email) and ask them to review your questionnaire.

Decide if you want to give your colleague a read only (view) permissions or you want them to be able to modify your questionnaire.

Exercise 4: Questionnaire logic: Validation

Modify the questionnaire adding validation rules where necessary. Signal as errors the following situations:

- house with walls made of wood and a roof made of concrete;
- house with roof made of concrete and older than 100 years;
- house with roof made of straw or similar material with available running water;
- year of construction or year of renovation is in the future;
- year of renovation is earlier than the year the house was built;
- negative distance to any of the infrastructure targets;
- number of members, males, females or children being negative;
- number of members being too large (decide on the appropriate limit).

Look closely at the questions on household composition (males, females, children). Which relationships exist between these variables? Suggest corresponding validations and implement them in the questionnaire.

Add validations to utility expenditures. What are the reasonable ranges for utilities to be paid by typical households on average per month in your area? Is there any seasonality (heating season, irrigation season, etc). How can you allow for seasonality in validations?

It is possible, but unlikely that the house was renovated in the same year when it was built. How would you alert the interviewer about this?

Exercise 5: Rosters

Add demographic section to your questionnaire. Collect the following information about every household member:

- name;
- relationship to the head of the household: head (self), spouse, child, etc.
- age;
- sex;
- marital status (for persons 15 and older);
- disability status;
- literacy status (for persons of age 15 and older);
- school attendance status (for persons of age 5-23);
- highest completed level of education (for persons of age 16 and older);

- employment status (for persons 15 and older);
- industry, occupation, public sector status, union status, typical weekly hours worked, and salary or wages (for workers);

Which skip conditions are needed for this questionnaire? Which variables are crucial for the skips? What should be the behavior if these questions are not answered?

Exercise 5a: Rosters - validation

Add validations:

- age must be between 0 and 120;
- age of the person with status head may not be less than 15;
- person with higher degrees of education may not be illiterate;
- person employed in public sector is bound by a certain minimal wage (decide on appropriate value);
- hours worked must be feasible;
- unionized workers should not report more than 42 hours of typical weekly work time;
- no unions are available for employees in the military and police (industries);

What should be the minimum age, at which you expect the person may complete each selectable level of education? Add validation that verifies how the age agrees with education.

What should be the marital status of the person with relationship “spouse” in the household? Which validation or enabling condition do you need to add to your questionnaire?

Exercise 5b: Rosters - routing

Suppose it is essential to acquire all of the demographic information (name ... education) about all the household members before the interviewers proceed to the questions on employment. How would you modify the questionnaire to guide the interviewers on this new trajectory?

Exercise 6. Nested rosters

Add agricultural section to your questionnaire. Ask for each household how many land plots the household operates. Then for every plot:

- since when (or how many years);
- status (owned, rented, etc);
- area of the plot;
- irrigation status of the plot;
- whether the plot was used for crops cultivation last season (y/n);
- if yes, which of the following crops were grown: (wheat, corn, kassava, rice, potatoes, other);

For every crop inquire about the following:

- proportion of the plot occupied by this crop;
- amount harvested last season;
- crop disposition: how much was sold, consumed, lost, etc;
- source of seeds: own, local source, imported seeds;
- whether any of the following chemicals were used: herbicides, pesticides.

Add validations where necessary. Pay attention to numeric variables and relationships between them.
Should the sum of proportions occupied by all crops at the plot equal to 1?
Indicate as an error the situation where rice is grown on a non-irrigated plot.