NIGERIA DATA COLLECTION

LESSONS ON AUTHORING INFRASTRUCTURE SURVEYS AT SCALE

SUSTAINABLE
ENGINEERING
LAB.

One seemingly small, but largely integral component of data collection is writing effective surveys. The following provides a few examples of the challenges encountered during the Nigeria Scale-Up project and the insights learned.

Designing a data collection effort at a national scale is very different than a smaller more focused survey effort. At scale, time, training efforts, and the number of facilities needing to be surveyed is far greater. To illustrate the scale of the collection in Nigeria, a total of over 34,000 health facilities, 98,000 education facilities and 132,000 water facilities have been gathered to date. In a lot of cases, budgets are not large enough to provide the attention to detail necessary; therefore data quality can decrease when dealing with larger scales if not appropriately accounted for. One way to improve data quality is to focus on survey writing.

Terminology of surveys

Because of budget restrictions, most enumerators do not have extensive training or prior knowledge in health, education or water sectors. For example, terms and definitions that are common among health professionals, like *Caesarean Section* and *CHEW* (Community Health Worker) may be very foreign to the enumerator asking the question. There is also the possibility that common Western terms are unfamiliar to the communities you are working with. It's important to enlist local help when writing questions, to ensure comprehension and cultural clarity.

Take this survey question for example, "Which of the following improved water points are used by the facility and located within 100 meters of the facility?"

The water and sanitation term, *improved* needs to be defined in order for an accurate response. The World Health Organization (source) defines *improved* drinking water sources as:

Household connection
Public standpipe
Borehole
Protected dug well
Protected spring
Rainwater collection

This can be addressed by using a multiple choice question to help enumerators understand the accepted sources, but even this remains a little complex and confusing. Water and sanitation knowledge is still necessary to correctly identify the type. And in this case, terms like *protected* can be just as vague as *improved*.

A solution that proved effective in the case of the Nigeria Scale-up project was to accompany the term with supporting photos, illustrating the options. It was also necessary to provide more thorough training of enumerators on difficult concepts such as this.

Lesson: Avoid jargon and terms that require prior knowledge. Consider cultural differences and enlist local help. If necessary, accompany with a descriptive photo or provide more detailed training.

Double-barreled questions

When we originally set out to write surveys for health, water and education facilities the team was excited by the opportunity to gather any and all of the data we wanted. The thought was, the more data we had, the better the planning would be. Specialists from each sector outlined critical indicators of MDG progress and created a list of questions to ask in order to obtain and determine these indicators. We were afraid to leave anything out considering this was an opportunity to gather nationwide data for every health and education facility in Nigeria. After reviewing the data from initial surveys, the quality wasn't at the level wanted or expected. Considering our surveys were well over 100 questions long, we hypothesized that we were experiencing issues of enumerator and respondent fatigue. We were asking overly detailed questions, like "How many forceps are at this facility?" at the expense of more important questions like "Do you have emergency transport at this facility?" The survey length was reducing data quality. Our goal after coming to this hypothesis was to shorten the survey.

In an effort to condense, yet retain as much valuable data as possible, we unintentionally began asking complex or double-barreled questions. Instead of cutting questions out of the survey, we began combining multiple into one. For example, we asked, "Does this facility have a functional improved water point?"

In this case we are asking two questions:
1. Is the water source functional?
2. Is the water source improved?

This is an issue because the respondent can't possibly answer both with one response. In practice, some respondents only respond to one part of the question, and the same situation may lead to different answers depending on the interpretation of the respondent. For us, the end goal was to know if both of these items were true, therefore it felt intuitive to ask them together. It's obvious now to separate them out and ask multiple clear questions rather than combine them into one, but it's an easy oversight when trying to condense a survey.

Lesson: Double-barreled questions should be avoided. They are easy to spot, once you know what to watch for, but commonly overlooked when not.

In conclusion, when writing surveys at scale, make sure your survey is as concise as possible, questions are simply constructed and void of complex terminology. In the end, we cut each survey nearly in half to about 30-40 questions a piece, reducing the response time to under 15 minutes and improving data accuracy in the process. Hopefully these insights will prove helpful to others working in similar situations and constraints. Our best piece of advice would be, allow time to iterate with piloting, back-checking and revising.