Education Data Quality Assessment: Household Survey Data

Training workshop material

June 2017

Overview of Household Survey Data Quality Assessment (DQA) Tool

This section of the Training Workshop Manual provides an overview of the tool for assessing data from household surveys that include data on education. The overall training workshop guide provides background on CapEd and detailed information on the administrative data assessment, as well as a glossary of terms (relevant to the household survey tool as well). In contrast, this section of the manual focuses on the utility of household survey data on education, and how those data can be used in education statistics, including for reporting on SDG 4.

A number of household survey programmes include education data that can be used to produce key education statistics, including those on educational attainment among adults, literacy rates, school participation and completion (attendance rates, out-of-school (OOS) rates, dropout, and so on). In addition, some household survey programmes collect data on reading and math skills – notably the current round of UNICEF's Multiple Indicator Cluster Surveys (MICS). Some of these household surveys are focused on education, such as the DHS EdData Household Surveys (which ended in 2004), and the Nigeria Education Data Survey (conducted in 2004, 2010, and 2015). More common, however, are household surveys covering multiple topics that include education as a variable of interest.

In the last decade or so, education data from household surveys have been used to complement and supplement country administrative data. There has been a gradual increase in the use of household survey data on participation, literacy, and educational attainment, and these data are now used widely in intra- and cross-country comparisons made by the UIS, UNICEF, the World Bank, and many other providers and consumers of education statistics. Country-level use of household survey education data, however, has been more limited.

The two main multiple-topic international household survey programmes with education data are the Demographic and Health Surveys (DHS) and the MICS. DHS and MICS both collect data on population, health, nutrition, and HIV in developing countries. Data are representative at the national level and sub-national levels (by district or region). MICS and DHS cover many of the same topics, although there are variations. For instance, MICS collects birth date data, as well as age in completed years, for all household members, while in the household questionnaire, the DHS collects only age in completed years. DHS does obtain birth date data, and reconcile them with age in completed years, for selected groups (women age 15-49, sometimes men in a similar age range, and children up to age 4). Both the DHS and MICS data can be disaggregated by geographic area (at varying levels of systems), urban/rural residence, household wealth, sex, age, and other characteristics. MICS 6, the latest round of surveys, also includes data on disability among children and youth, allowing for the disaggregation of education indicators by disability status.

In addition, other household survey programmes collect data that can be used for producing education statistics. Household income and expenditure studies, labour surveys, and other multiple-topic surveys may include information on education that can be used to report on education matters, including the indicators for SDG 4.

As with any data source, it is important to assess the suitability of these household survey data on education for education statistics purposes. Questions need to be addressed on whether the right data are collected, whether these data are sufficiently reliable and whether they are representative, and so on. The Household Survey DQA tool is designed to help country teams assess the utility of various household data sources for use in education statistics.

Use of the Household Survey DQA tool

The tool is designed for use in assessing, separately, each recent household survey with data on education. In many instances, there will be the possibility of looking at time series data, given that many household survey programmes are undertaken every 3 to 5 years, in a given country.

At the country level, various government and parastatal groups may be involved in conducting household surveys with some data on education. These agencies may include the National Statistics Office, the Ministry of Education, and other line ministries. In some instances, non-government groups may be the country lead on survey efforts.

Each household survey will need to be reviewed using the tool. In some instances, review teams may include members familiar with the data source, while in other instances, consultations with the primary implementing agency will be required.

The structure of the Household Survey DQA tool

The Household Survey DQA tool is formatted in Excel, according to the Code of Practice (CoP) for household survey data (see attached CoP) principles and measures. The tabs cover the following topics:

- Background information
- Statistical processes
- Statistical outputs

The background information tab captures descriptive information on the household survey under review. Information is captured on: the name of the survey, agencies responsible for its implementation, the data collection period, source of the sampling frame, geographic coverage, and any additional information (beyond what is captured under principle 3) on data limitations. There are no scores attached to the information in the background information tab.

The remaining tabs in the spreadsheet are structured in the same way as in the DQA tool for administrative data, and the principles are quite similar to those for administrative data, as well, but are adapted for the household survey context. One major difference between the types of questions asked is in principle 3, accuracy and reliability, where the Household Survey DQA tool includes a number of questions on sampling and on the utility of the education data for producing education statistics.

The CoP (attached) provides further detail on the 6 principles elaborated on in the Household Survey DQA tool.

The training workshop and next steps

The workshop is focused on familiarizing participants with the types of information needed to assess the utility of household survey data for producing education statistics. One potential advantage of household survey data on education is the ability to disaggregate data by individual and household characteristics. School participation rates and other indicators often can be examined by household wealth; ethnic, linguistic, or religious group; and other characteristics – in contrast to data from administrative sources.

The workshop session on the DQA tool for household surveys focuses on how to use the tool to assess household survey data quality, and the usefulness of the data for producing education indicators. Emphasis will be on reviewing the tool and testing it with a sample household survey. As a group, participants will go through the questions in the tool, identifying the sources of information to be referenced in order to answer each question, and talking about how to use the scoring system (levels 4, 3, 2, and 1) for each item. Once a level or score has been agreed upon, it should be entered in the 'score' column, and notes should be added on the source of information, further explanation (as needed), and comments to guide further use of the data quality assessment information.

At the end of the workshop, participants will be able to assess household survey data on education for its usefulness in education statistics, and in knowing the criteria for producing quality data, they will be in a position to advocate for the improvement of household survey data consistency and quality.

As with the assessment of administrative data, workshop participants will convene a group with representatives from the agencies involved in household surveys with education data. These surveys may include MICS, DHS, labour force surveys, income and expenditure surveys, and so on. Once the range of recent surveys has been captured, the teams will pull together documentation and background information (publications, metadata, etc.), and access datasets, to have information at the ready for completion of the Household Survey DQA tool. Methods used will be similar to those used for the collection of information on administrative data.

Scoring guidelines for household survey data are similar to the approach used for administrative data. As with the administrative data, the Household Survey DQA tool does not produce a single numerical measure. Rather, each criterion and its assessment provides information on quality related to that measure. Scores are not meant for international comparison, but rather for highlighting areas of strong performance and areas for improvement. In the case of household surveys with data on education, a comparison of the measures in the Household Survey DQA tool may suggest which household surveys are best placed to provide education statistics for reporting and use.

Each criterion has levels of assessment attached, ranging from level 4 (meets quality standards), to level 1 (poor statistics), with level 3 (acceptable statistics) and level 2 (questionable statistics) in between. Most criteria include measures at all four levels, although some have fewer levels, according to the nature of the measure.

Once a household survey has been rated across all measures (27 questions over the 6 principles), scores can be summed up overall – assigning a score equal to the level ranking, such that level 4 = 4 points, level 3 = 3 points, level 2 = 2 points, and level 1 = 1 point. Scores can be summed across all measures, and for each principle and sub-section within a principle.

The governing training manual for administrative data provides an overview of next steps in the CapEd process; please refer to that manual for information on the process, and for the accompanying glossary.

Note: National Statistics Office (NSO) Institutional Environment Assessment tool

In addition to the household survey DQA tool, there is another tool designed to assess NSO capacity for household survey work. This NSO tool has much in common with the types of questions asked in the administrative data tool institutional environment tab. Note that this NSO (or appropriate country organization) tool is separate from the household survey DQA tool because it is intended to be used only once, to assess NSO capacity for household survey work, rather than for each household survey – as is the case with the household survey DQA tool.

The NSO institutional assessment is structured in the same way as is the household survey DQA tool, with scores from levels 1 to 4, to identify areas for improvement. There are two principles – policy and legal framework; and adequacy of resources – with a total of 11 measures.