India - Young Lives: School Survey 2010-2011, Round 1

Boyden, J. - University of Oxford. Department of International Development

Report generated on: September 4, 2014

Visit our data catalog at:

Sampling

Sampling Procedure

Sampling Procedures: Multi-stage stratified random sample
Number of Units: 953 Young Lives children across 249 schools

Weighting

No weighting used.

Questionnaires

Overview

The survey instruments included data collection at the school, class and pupil level, and involved the head teacher, class teacher, and pupil. The instruments comprises of the following components:
• School roster
• Child questionnaire answer sheet 1
• Child questionnaire answer sheet 2
• Child Maths test
• Child Telugu test
• Child English test
• Child language learning experience
• Child observation 1 (Maths)
• Child observation 2 (Maths)
• Child observation 3 (Maths)
• Teacher questionnaire
• Teacher content knowledge test (Maths)
• Maths teacher observation 1
• Maths teacher observation 2
• Head teacher questionnaire
• School observation
• School observation - homework

Data Collection

Data Collection Dates

Start	End	Cycle
2010-12	2011-03	N/A

Data Collection Mode

Face-to-face interview; Self-completion; Educational measurements; Observation

DATA COLLECTION NOTES

Training
The fieldworker training took place during a period of three weeks in November 2010 at The Center for Economic and Social Studies (CESS). Each of the survey modules was introduced and discussed during training sessions. Each team of fieldworkers were then asked to conduct three practice surveys in schools of their choice under the supervision of their respective supervisors. This procedure allowed the fieldworkers to familiarize themselves with the instruments and enabled them to clarify any doubts before the final fieldwork.

Fieldwork
Fieldwork was conducted across the 7 districts and 20 mandals which constitute the Young Lives sample, between 1st December 2010 - 15th March 2011.

Fieldwork teams consisted of one supervisor and two pairs of fieldworkers. Teams travelled to sites as a group, and split into pairs to conduct the survey, with the supervisor responsible for overseeing the work of each pair, and for facilitating access to sample schools. The supervisor facilitiated discussion amongst the team each day to clarify any issues that arose during fieldwork. It was the role of this supervisor to check each completed instrument, clarify doubts, and report back to the Principal Investigator based in Hyderabad should they encounter difficulty. In addition to this day-to-day supervision, members of both the India and Oxford teams conducted visits to each fieldwork team throughout the fieldwork period. These visits involved spot checks of completed protocols to ensure uniformity of completion and understanding. Where problems were identified they were investigated, and the conclusions relayed to all field teams.

For the purposes of the Classroom Observation protocol, inter-observer reliability checks were conducted in the middle of the fieldwork period (January 2011) whereby two fieldworkers observed the same children and/or the same teachers, and their scripts were returned to Oxford for comparison. For the child observation, 87.0% of observations matched and for the teacher observation, 88.5% matched, demonstrating a high level of mutual understanding in the interpretation of the protocol between fieldworkers.

Data Collectors

Name	Abbreviation	Affiliation
Galab, S., Centre for Economic and Social Studies (India)		
Reddy, P., Centre for Economic and Social Studies (India)		

Data Processing

Other Processing

Data Archive Processing Standards
The data were processed to the UK Data Archive's A standard. A rigorous and comprehensive series of checks was carried out to ensure the quality of the data and documentation. Firstly, checks were made that the number of cases and variables matched the depositor's records. Secondly, checks were made that all variables had variable labels and all nominal (categorical) variables had value labels. Where possible, either with reference to the documentation and/or in communication with the depositor, absent labels were created. Thirdly, logical checks were performed to ensure that nominal (categorical) variables had values within the range defined (either by value labels or in the depositor's documentation). Lastly, any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below.

Data and documentation problems
None.

Data conversion information
From January 2003 onwards, almost all data conversions have been performed using software developed by the UK Data Archive. This enables standardisation of the conversion methods and ensures optimal data quality. In addition to its own data processing/conversion code, this software uses the SPSS and StatTransfer command processors to perform certain format translations. Although data conversion is automated, all data files are also subject to visual inspection by a member of the Archives Data Services team.

With some format conversions, data, and more especially internal metadata (i.e. variable labels, value labels, missing value definitions, data type information), will inevitably be lost or truncated owing to the differential limits of the proprietary formats. A UK Data Archive Data Dictionary file (generally in Rich Text Format (RTF)) is usually provided for each data file, enabling viewing and searching of the internal metadata as it existed in the originating format. These files are called: [data file name]_UKDA_Data_Dictionary.rtf

Important information about the data format supplied
The following descriptions provide important information about the Archive's data supply formats. Some of this information is specific to the ingest format of the data, i.e. the format in which the Archive received the data from the depositor. The ingest format for this study was SPSS.

SPSS files (*.sav files)
If SPSS was not the ingest format, this format will generally either have been created via the SPSS command processor (e.g. if the ingest format is STATA, SAS, Excel, or dBase). If the ingest format was non-delimited or fixed-width text, SPSS files will have been created using SPSS command syntax.

Issues: There is very seldom any loss of data or internal metadata when importing data files into SPSS. Any problems will have been listed above in the Data and Documentation Problems section of this file.

STATA (*.dta files)
If STATA was not the ingest format, STATA files will generally have been created from SPSS via the StatTransfer command processor. Importantly, StatTransfer's optimisation routine is run so that variables with SPSS write formats narrower than the data (e.g. numeric variables with 10 decimal places of data formatted to FX.2) are not rounded upon conversion to STATA because they are converted to 'doubles ' rather than floats. Discrete user missing values are copied across into STATA (as opposed to being collapsed into a single system missing code).

Issues: There are a number of data and metadata handling mismatches between SPSS and STATA. Where any data or internal metadata has been lost or truncated, it will be logged in the study's SPSS_to_STATA_conversion RTF file. Note that the complete internal metadata has been supplied in the UKDA Data Dictionary file(s): [data file name]_UKDA_Data_Dictionary.rtf

Tab-delimited text (*.tab files)
If tab-delimited text was not the ingest format, tab-delimited files will have been created from via the SPSS command processor, and also from Excel and MS Access files. When exporting from Access data tables to tab-delimited text, the

potentially problematic special characters (tabs, carriage returns, line feeds, etc.) allowed by Access memo and text fields may have been removed by the Archive if necessary.

Issues: Date formats in SPSS are always exported to mm/dd/yyyy in tab-delimited text format. There may be a mismatch with the documentation on such variables. Variables that include both date and time such as dd-mm-yyyy hh:mm:ss (e.g. 18-JUN-2011 13:28:00), will lose the time information and become mm/dd/yyyy. All users of the data in tab-delimited format should consult the UK Data Archive Data Dictionary RTF file(s).

If the data was exported from MS Access, more limited 'data documenter' information is generally available in the RTF variable information files. These files may also contain SQL setup information.

MS Excel (*.xls/*xslx files)
If MS Excel was not the ingest format, Excel files may have been created via StatTransfer. The date and time issues noted under tab-delimited format may also apply here.

SAS (*.sas7bdat and *sas files)
If SAS was not the ingest format, SAS files will usually have been created via StatTransfer or SPSS. SAS is not one of the Archive's standard supply formats, and the files are likely to have been created in response to a user request. The usual format is *.sas7bdat files plus a .sas proc formats file. Note that the complete internal metadata has been supplied in the accompanying UK Data Archive Data Dictionary file(s).

Issues: The main loss of information when converting from SPSS to SAS is user-missing value definitions. By editing the .sas file, the user can choose whether to collapse all user-missing values into system missing or preserve the?value and lose the user-missing definition. To achieve the latter?the following section of the .sas file should be removed before running it:

/* User Missing Value Specifications */

Note that the complete internal metadata has been supplied?in the UKDA Data Dictionary file(s): [data file name]_UKDA_Data_Dictionary.rtf

MS Access (*.mdb/*.mdbx files)
Due to substantial incompatibilities between versions of MS Access, the Archive will only make data available in MS Access format if this is the ingest format and/or the database contains important information in addition to the data tables (coding information, forms, queries, etc.).

Data Appraisal

No content available