Quality of baseline data: The case of PRIME project

Md. Abul Kalam
Senior Staff Sociologist

December 2002

Table of Contents

269

Abstact

This study attempted to assess the quality of baseline data of the Primary Initiative in Mainstream Education project of BRAC Education Programme collected through census done by locally recruited interviewers. A sample survey of 900 households under census was done. Of these 900 households 100 were re-surveyed. Inter-observer agreement or concordance was used in assessing the quality of data. The measurement of inter-observer agreement indicated that there were less than 10% cases in each variable in which a disagreement between observers of sample survey and re-survey was found. This further indicated an acceptable level of the quality of data in sample survey as well as re-survey. However, the study indicated that there were more than 10% cases even more than 30% cases in each variable in which a disagreement between observers of census and sample survey was evident. This further indicated an unacceptable level of the quality of data in census. Thus, the study raises a question about the quality of data in census. Any interpretation of the census data with this unacceptable level of quality might be misleading. It is therefore, suggested that one should be careful in using the census data for a more realistic interpretation otherwise it would give misleading picture of the situation which further would mislead the intervention. The study also suggests, in any census in future it is to be ensured that proper, adequate and sufficient training is provided to interviewers with a well-explained manual. It further suggests that proper and adequate monitoring/supervision by the supervisors has to be done to ensure the quality of data and that the data are properly collected.

Introduction

BRAC Education Programme (BEP) has initiated an experimental project called Primary Initiative in Mainstream Education (PRIME). The main purpose of PRIME is to seek ways of working with the government to transfer BRAC's knowledge and practices to the mainstream education system of Bangladesh and thereby improve its quality. In July 2001 BEP started initial work of selecting upazillas, unions, and schools; continuous contact with different stakeholders of primary education sub-sector at field level as well as central level. Initially one union from each of the 30 upazillas were selected where the government's IDEAL project is working. Later on, another union from each of these upazllas were selected. IDEAL stands for Intensive District Approach to Education for All.

Few initiatives have initially been identified to implement through PRIME subject to government's approval. These include regularizing the irregular students in primary schools; controlling dropout of the students; involving local community in school activities; providing part-time refreshers training for the teachers; making all the stakeholders at field level more aware about the primary education sub-sector; and operating baby class (pre-primary) for children aged five years to prepare them physically and mentally for going to formal primary schools.

To gain information on enrolment status of school-age children and to assess the feasibility of opening pre-primary schools, PRIME has done a census in catchment area of all the formal primary schools in the project area. The data were collected by locally recruited and trained field investigators who are BRAC school graduates or ex- teachers of BRAC schools. BEP requested the Research and Evaluation Division (RED) of BRAC to assess the quality of data collected through the census. This study is an attempt to assess the quality of the data collected through the census done by the programme.

Objective of the study

The objective of the study was to assess the quality of data collected through the census which was done by locally recruited field investigators.

Methodology

Area and sample

All the 30 upazillas under PRIME were selected for the purpose. Two school-catchment areas from each upazilla were selected randomly. Fifteen households from the sample school-catchment area were randomly selected. Thus a sample of 900 (2 X 30 X 15) households were selected for the study. One hundred households among these were randomly selected for a re-survey to assess the quality of data in the sample survey.

Methods

Inter-observer reliability or inter-observer agreement was used to assess the quality of data in this study. Inter-observer agreement is the extent to which two or more observers obtain the same results when measuring the same behaviour (i.e when independently coding the same tape or event or behaviour). The measurement of inter-observer agreement involves the calculation either of the degree of correlation between the two sets of measurements, or of the agreement, sometimes called concordance between them (Robson, 1993:221). Here we have used concordance or the agreement between two sets of measurements. The index of agreement (or concordance) is simply the proportion expressed as a percentage. Sample household survey and re-survey were done.

Instrument and data collection

A simple instrument was used for collecting data on child's age, sex, education, enrolment and school type, parents' education and occupation, household's land holding, etc. Thirteen field investigators were recruited and trained for data collection. Of them, 10 did the sample survey and copied the information from the census sheets of the same households maintaining the same identity (line no) of the household members in two sets

172

of data sheets. The remaining three did the re-survey in 100 households among the households under sample survey.

Data analysis

Data were analyzed in two stages. In first stage, the agreement and disagreement between the data sets of sample survey and re-survey were measured directly without considering any error. Then we made ranges with data such as age, education, current class and land holding considering minor disagreement, and then measured the agreement and disagreement between these two data sets. In making ranges we ignored the difference of one year or class in age and education. In the case of land holding, we ignored the difference of 10 decimals. In the second stage the two above-mentioned approaches were repeated in measuring agreement and disagreement between the data sets of sample survey and census. Through the first stage, it was tried to establish or set a standard of the quality of data with sample survey against which the quality of census data has been assessed in terms of agreement and disagreement in the second stage.

Results

Setting the standard

Quality of the data in sample survey has been assessed to set a standard for assessing quality of data in census. A re-survey with 100 household among sample survey households was done. Table 1 presents the summary of agreement between observers of sample survey and re-survey. A disagreement between the observers of two surveys in few cases (2%) of children was found. Data show that the observers in two surveys reached agreement in over 90% of the cases of each variable. This indicated a minor disagreement of below 10% between two data sets.

Table 1. Proportion of the cases of agreement between observers in sample survey and re-survey by variables

Variable	Proportion of the cases of agreement between observers	
	Direct agreement N = 217	Agreement after making ranges N = 217
Land	90.3	96.8
Sex of child	98.6	--
Age of child	93.1	95.4
Education of child	92.6	93.5
Enrolment status	97.2	--
Current class	94.5	96.4
School type	95.9	--
Mother's education	92.2	93.1
Mother's occupation	100	--
Father's education	96.3	97.7
Father's occupation	97.2	--

Variable-wise analysis indicated that the observers reached agreement with highest proportion in mother's occupation (100%), followed by sex of child (98.6%), enrolment status of children, father's occupation (97.2%) and education (96.3%). The lowest proportion of agreement was found in household land holding (90.3%), followed by mother's education (92.2%), education of child (92.6%), age of child (93.1%), current class (94.5%), and school type (95.9%) (Table 1).

The summary of the proportion of cases of agreement between observers of sample survey and re-survey after arranging value of some variables into ranges is also shown in Table 1. Arranging the value of variables increased the proportion of agreement between the observers of two surveys. This resulted in an increase of the proportion of agreement between the observers of two surveys from 90.3% to 96.8% in land holding, 93.1% to 95.4% in age of child, 92.6% to 93.5% in education of child, 94.5% to 96.4% in current class, 92.2% to 93.1% in mother's education, and 96.3% to 97.7% in father's education. If we consider a difference of 20 or 30 decimals in case of land holding then the agreement between the observers increased to 97.8%.

Level of the agreement and disagreement between observers of sample survey and re-survey on value of different variables indicated an acceptable level of the quality of data of these two surveys with minimum error of less than 10%. In this case it was more accurate as the proportion of disagreement was found to be 5% or less in all the variables except child and mother's education.

Quality of census data

Table 2 resents the summary of agreement between observers of census and sample survey. A disagreement in 4.3% cases of children was found between observers of census and sample survey, which was about 2% in the cases of sample survey and re-survey. Table 2 shows that the observers of census and sample survey reached agreement in over 30% of the cases of each variable. These indicated a huge disagreement between the observers of census and sample survey in the value of different variables.

Table 2. Proportion of the cases of agreement between observers in sample survey and census by variables

Variable	Proportion of the cases of agreement between observers	
	Direct agreement N = 1802	Agreement after making ranges N = 1802
Land	33.2	62.1
Sex of child	96.1	--
Age of child	47.8	81.2
Education of child	75.2	84.6
Enrolment status	89.2	--
Current class	74.0	82.9
School type	81.3	--
Mother's education	58.0	63.7
Mother's occupation	97.0	--
Father's education	59.2	65.1
Father's occupation	71.4	--

Table 2 shows that the observers reached agreement with the highest proportion in mother's occupation (97.0%), followed by sex of child (96.1%), enrolment status of the children (89.2%), and school type (81.3%). The lowest proportion of agreement was found in household land holding (33.2%), followed by age of child (47.8%), mother's education (58.0%), father's education (59.2%), father's occupation (71.4%), current class (74.0%), and education of child (75.2%).

Table 2 also provides the summary of the proportion of agreement between observers of census and sample survey after arranging value of some variables into ranges. Making ranges with the value of some variables increased the proportion of agreement between the observers in these variables which ranged from about six to 32 percentage points. Arrangement of the value of variables into ranges increased the proportion of agreement between the observers of census and sample survey from 33.2% to 62.1% in land holding, 47.8% to 81.2% in age of child, 75.2% to 84.6% in education of child, 74.0% to

82.9% in current class, 58.0% to 63.7% in mother's education, and 59.2% to 65.1% in father's education. If we consider a difference of 20 and/or 30 decimals in the case of land holding then the proportion of the agreement between the observers increased to 72.1% and 77.9% respectively. However, even then, the proportion of disagreement in these cases remained very high.

Level of agreement and disagreement between the observers of census and sample survey on the value of different variables indicated a discouraging level of the quality of data in the census. A maximum of 10% error or disagreement might be acceptable in investigation to be scientifically valid. In the case of the data in census and sample survey disagreement between the observers was found to be more than 10% in almost all variables except sex of child and mother's occupation. The proportion of disagreement was found to be more than 30% in many cases. These findings raise questions about the quality of data collected through the census.

The reason for such a level of the quality of data in census might be different. One of them might be that the interviewers were not properly and adequately trained. This might be resulted in interviewers' misconception in variables and method/techniques for collecting information on different variables. The interviewers might not try to probe the information again by using proper method/techniques. Another reason might be due to information was not collected from appropriate respondents i.e there might be mistakes in selecting respondents for collecting data. The other important reason might lack of proper and adequate monitoring/supervision by others while data collection was underway.

Conclusion

This study assessed the quality of baseline data of the PRIME project of BEP collected through census done by locally recruited interviewers. Inter-observer agreement or concordance was used in assessing the quality of data. The measurement of inter-observer agreement between the observers of sample survey and re-survey showed that the observers/interviewers of both the surveys reached agreement in over 90% of the cases of all variables. This indicates that there were less than 10% cases in each variable in which a disagreement between observers of sample survey and re-survey was found. Less than 10% disagreement between the observers might lead to a less than 10% error in the data which might be an acceptable level for any scientific investigation as highest 10% of error can be taken acceptable. This further indicated an acceptable level of the quality of data in sample survey as well as re-survey.

The study revealed that the observers/interviewers of census and sample survey reached agreement in less than 90% cases of all the variables except two. This indicates that there were more than 10% cases even more than 30% cases in each variable in which a disagreement between observers of census and sample survey was evident. This further indicated an unacceptable level of the quality of data in census. Thus, the study raises a question about the quality of data in the census. Any interpretation of the census data with this unacceptable level of quality might be misleading. It is therefore, suggested that one should be careful in using the census data for a more realistic interpretation, otherwise it would give misleading picture of the situation which further would mislead the intervention. The study also suggests, in any census in future it is to be ensured that proper, adequate and sufficient training is provided to interviewers with a well-explained manual. It further suggests that proper and adequate monitoring/supervision by the supervisors has to be done to ensure that the data collection is properly done as well as to ensure the quality of data, while data collection is underway.

Reference

Robson Colin (1993) Real World Research: A resource for social scientists and practitioners- researchers, BLACKWELL, Oxford UK & Cambridge USA.