

## DATA QUALITY REPORT

1. DATA IDENTIFICATION	
Data File:	Teachers.dta
Dictionary:	Teachers_codebook.smcl
Data Cleaning Do-File:	Teachers.do

2. STRUCTURAL STABILITY			
	Yes		No
2.1 Do all variables in the questionnaire match those in the data set?			X
2.2 Do all value labels in the questionnaire match those in the data set?	X		
2.3 What is the name of the dataset's unique identifier?	teacherid		

*If 2.1 or 2.2 is false, please append additional sheets describing the discrepancies.*

3. EDITING AND IMPUTATION			
	Yes		No
3.1 Did any variable exhibit errors in more than 5% of observations?	X		
3.2 Was any editing or imputation applied to the data set <i>due to illegal or incoherent responses?</i> (see attached do-file)	X		
3.3 Were any outliers rejected?			X

*If either 3.1 or 3.3 are true, please append additional sheets describing the errors identified. 3.2 should appear in the attached data cleaning do-file.*

This dataset corresponds to sections I, II and III of the questionnaire « Questionnaire for Teachers of school (Grades I-VI only) Expenditure and Service Delivery Survey ZAMBIA : Educational Sector The World Bank». This questionnaire was named Teacher\_Questionnaire.doc in the file.

## **2.1 Data set and questionnaire matching**

There is no data set for the section IV of the questionnaire.

## **3.1 Variable exhibit errors in more than 5% of observations**

S1q12, s1q16, s3q3a-s3q3g, s3q4a-s3q4g, s3q7a-7e

## **3.2 Editing**

The following editing work has been accomplished.

There are total 10 skip patterns in this data set that have been tested.

- (1) s1q6 is replaced by missing value if s1q5 is one and s1q6 is “.n”
- (2) s1q9 is replaced by missing value is s1q8 equals to 1,3,4 or 5 and s1q9 is “.n”
- (3) if s1q12 equals 2 then s1q13-15 should not be answered and s1q13-15 were replaced by missing values.
- (4) s2q7 is replaced by missing value is s2q6 is no.
- (5) each value of s3q3a-3g was replaced by missing value if each of s3q2a-2g equals to 2 and each of value of s3q3a-3g is zero.
- (6) each value of s3q4a-4g was replaced by missing value if each of s3q2a-2g equals to 2 and each of value of s3q4a-4g is zero.

Two skip pattern errors have been detected but not corrected due to the size of error correction is huge

- (1) For s1q16 skip pattern, over 50 percent of the survey answered the question 17-20, which should not be answered. The errors are huge and have not been corrected.
- (2) \*s3q7a-e and s3q8 questions, there are many people answered s3q8 questions though they confirmed no in s3q7. This skip pattern is not correct due to too many errors.

Other Editing

- (1) “-99” in the cell is replaced by “.r”, which means “refuse to answer”
- (2) S2q7, s3q8, s2q5, s1q6, s1q9, s1q13, s1q14, s1q15 and s1q18 initial value is “10” and was replaced by “.n”, which means “do not know”

