DATA QUALITY REPORT

1. DATA IDENTIFICATION	
Data File:	generalschool.dta
Dictionary:	generalschool _codebook.smcl
Data Cleaning Do-File:	generalschool.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?	schoolco		

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 *due to illegal or incoherent responses*? (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.

s

This data set corresponds to sections I, II and III of the questionnaire « General School Questionnaire Expenditure and Service Delivery Survey ZAMBIA : Educational Sector The World Bank». This doucment was named generalSchool.doc in the questionnaire file.

3.1 Variables that exhibit errors in more than 5% of observations

S3bq10 is the variable that exhibit errors in more than 5% of observations

3.2 Editing

Four skip patterns have been checked. Among which one of them needed to be corrected :
    (1) s3bq6 has been replaced by missing value if s3bq6 is one and s3bq4 equals to zero.
    (2) s3bq5 has been replaced by missing value if s3bq6 is one and s3bq4 equals to zero.

3.2.1 Other Editing
     (1) '-777' stands for 'NOT APPLICABLE' or '.n' in the user guide. However, the data entry staffs use '10' to stands for 'NOT APPLICABLE' or '.n'. So these errors have been corrected by using '.n' to replace '10'.

3.3 Outliers

There are a few outliers for the following variables : s3q2* s3q3* s3q5* s3q6* s5q2* s5q1* s5q2* s5q3*. However, though technically these variables have outliers, for the research purpose, we prefer to keep these outliers as the existing status and let the researchers to decide the range of outliers that they want to define or use.