Human Sciences Research Council
Lekgotla la Dinyakisišo tša Semahlale tša Setho
Raad vir Geesteswetenskaplike Navorsing
Umkhandlu Wezokucwaninga Ngesayensi Yesintu
Ibhunga Lophando Ngenzulu-Lwazi Kantu

USER GUIDE

HIV/AIDS Educators study

ELRC
2004-2005

Introduction to the HIV/AIDS educators study

The main objective of this research was to examine the impact of HIV/AIDS on South African educators. This is a national study conducted in all provinces, using a variety of research approaches. These are qualitative methods involving focus groups, key informant surveys, archival research involving review of records and quantitative methods involving questionnaire administrations and HIV testing of nearly 26 000 educators. The target population included everyone employed in schools. This included school-based educators from grades R to 12, Further Education and Training bands, students preparing for tutor training in tertiary institutions and principals. The study is limited to the public education sector.

Please refer to the Study Information document for the metadata of the study.

The following three questionnaires were used in the study:

- The educator questionnaire
- The student educators questionnaire
- The institution questionnaire

Available Data sets

The following data sets that are provided for further analysis, contain micro data.

SPSS data sets:

ELRC2004_2005_EDUCATORS_DATA.SAV (n=21,358); 609 variables.
ELRC2004_2005_EDUCATORS_DATA.SPS (syntax file)

ELRC2004_2005_STUDENT_DATA.SAV (n=919); 326 variables
ELRC2004_2005_STUDENT_DATA.SPS (syntax file)

ELRC2004_2005_INSTITUTION_DATA.SAV (n=1508); 119 variables
ELRC2004_2005_INSTITUTION_DATA.SPS (syntax file)

Stata data sets:

ELRC2004_2005_EDUCATORS_DATA.DTA (n=21,358); 609 variables.
ELRC2004_2005_EDUCATORS_DATA.DO. (syntax file)
ELRC2004_2005_EDUCATORS_DATA.DCT (dictionary file)

ELRC2004_2005_STUDENT_DATA.DTA (n=919); 326 variables
ELRC2004_2005_STUDENT_DATA.DO (syntax file)
ELRC2004_2005_STUDENT_DATA.DCT (dictionary file)

ELRC2004_2005_INSTITUTION_DATA.DTA (n=1508); 119 variables
ELRC2004_2005_INSTITUTION_DATA.DO (syntax file)
ELRC2004_2005_INSTITUTION_DATA.DCT (dictionary file)

SAS data sets:

ELRC2004_2005_EDUCATORS_DATA.SAS7BDAT . (n=21,358); 609 variables.
ELRC2004_2005_EDUCATORS_DATA.SAS (syntax file).
FORMATS.SAS7BCAT (formats)

ELRC2004_2005_STUDENT_DATA.SAS7BDAT (n=919); 326 variables
ELRC2004-2005_STUDENT_DATA.SAS (syntax file).
STUDENTFORMATS.SAS7BCAT (formats)

ELRC2004_2005_INSTITUTION_DATA.SAS7BDAT (n=1508); 119 variables
ELRC2004_2005_INSTITUTION_DATA.SAS (syntax file)
INSTITUTIONFORMATS.SAS7BCAT (formats)

ASCII data sets: (fixed format)

ELRC2004_2005_EDUCATORS_DATA.DAT (use with SAS, SPSS and Stata syntax files)
ELRC2004-2005_STUDENT_DATA.DAT (use with SAS, SPSS and Stata syntax files)
ELRC2004_2005_INSTITUTION_DATA.DAT (use with SAS, SPSS and Stata syntax files)

Conversion of Data

Ingest format(s) of the data: SPSS (.sav)
Additional comment about each data format is given below, please read the wording corresponding to the data format you have downloaded:

SPSS:
If SPSS has not been the ingest format this format has been created according to StatTransfer standards. Three files are created:
ASCII data file (.dat),. SPSS syntax file (.sps) and a SPSS data set (.sav).
Use the SPSS syntax file to convert data to a SPSS data set (if not provided). Be sure to change the path in the DATA LIST FILE= statement to point to where you stored the ASCII data file.

SAS:
If SAS is not the ingest format, SAS files are created according to StatTransfer standards. Four files are created: ASCII data file (.dat), SAS program file (.sas) and SAS data set (.sas7bdat) and SAS formats (sas7bcat). The formats.sas7bcat file contains the formats of the variables and must be used with the data set as illustrated under the heading: 'Additional notes on SAS'.

In the SAS program file, change the infile statement to point to where you stored the ASCII data file. To create a SAS data set (if not provided) open the SAS program in the SAS program editor and submit. If you do not want the value labels permanently associated with the variable values in the data set, comment out the format statement in the data step and rather use it in the PROCS.

STATA:
If STATA is not the ingest format, STATA files are created according to StatTransfer standards. Four files are created: A Stata data file (.dta), :ASCII data file (.dat) and Stata program files (.do and .dct).
To create a Stata data set (if not provided) open the .do program in STATA and change the path names to point to where you stored the ASCII data and the .dct (dictionary) files. Also indicate the path of the .dta output file. When the .do file is opened in Stata, run the .do (syntax) file to create a Stata data set.

The data sets contain a mix of the following measurement levels: ordinal, nominal and scale (continuous) variables. (See Appendix A for an explanation.)

Additional notes on SPSS

When using SPSS please note the following: if statistical analysis, other than descriptive statistics is done, the measurement level for variables should be set to the appropriate level as per the example below.

Example: Measurement Level syntax (SPSS): Educator data set

```
VARIABLE LEVEL
q4_7_03 q4_7_02
 (SCALE)
/ q1_2 q1_3 q1_4
(NOMINAL)
/ q13_10a q13_10b
 (ORDINAL).
```

Additional notes on SAS

In the downloadable SAS programs the formats were stored in a SAS catalog **formats.sas7bcat** for the **educators, STUDENTFORMATS.SAS7BCAT** for the **students.** This was the result of using the **library** option in proc format and

putting the format statement to associate the value labels with the variables, in the data step. Whenever the SAS data set is used, a libname statement must indicate where the permanent SAS data set and the format catalog are located. Store them in the same folder.

Example: Referencing the permanent SAS data set and the format catalog

Example1: Educators

Libname library '*full path where the data set and catalog are stored e.g. c:*\';

PROC FREQ DATA=library.ELRC2004_2005_Educators_data;
Run;

Please note that there is no need to specify the formats catalog if it is ***formats.sas7bcat*** *(default) as long as it is in the same folder as the data set.*

If however the formats has a name such as ***STUDENTFORMATS.SAS7BCAT*** *you need to put in additional statements.*

Example2: Students

libname library 'c:/myfolder' ;

OPTIONS FMTSEARCH=(library.Studentformats);

PROC FREQ DATA=library.ELRC2004_2005_Student_data

Note:
When searching for a format or informat, SAS always searches in WORK.FORMATS first, and then LIBRARY.FORMATS, unless one of them appears in the FMTSEARCH= list. SAS searches the catalogs in the FMTSEARCH= list in the order that they are listed until the format or informat is found.

If you want the values and not the formats in the data set, then use the following statements to disassociate the formats from the variables.

Example: Getting rid of the formats in the data set

The following code will remove all formats associated with the variables in the SAS data set.

Libname library '*full path where the data set and catalog are stored e.g. c:\myfolder*';

DATA *library.newdsn*;
SET *library.dsn*;
FORMAT _all_;
RUN;

- *library* refers to name you used in the libname statement.
- *dsn* is the name of the SAS data set, whose formats needs to be removed.
- *newdsn* is the name of the new SAS data set with the formats removed.
- format _all_ will delete all the formats in the SAS data set.

The following code will remove only the formats of specific variables in the SAS data set. Specify the variables without their formats in the FORMAT statement. This is also applicable to PROCS.

DATA *library.newdsn*;
SET *library.dsn*;
FORMAT q14_5a q14_5b ;
RUN;

Other formats:

Data are only made available in other formats on the rare occasion when there is no reliable method of extracting the data into a more accessible format. Other formats will be created upon request.

Missing values within the data set

Please refer to the project code books to find useful information about the distribution of responses and the proportion of missing values. In the data sets provided the missing values were all blanks. Different codes were used for not applicable (see codebook).

Variables and labels within the data set(s)

Please refer to the following code books for the variables, variable labels and frequencies within the respective data sets:

ELRC2004_2005_Educators_codebook.pdf
ELRC2004_2005_Student_codebook.pdf
ELRC2004_2005_Institutions_codebook.pdf

Derived / Recoded (categorized) variables (Educators data)

Recoded variables are described in the data labels. Computed scales were: Educator support index, Violence at school index, alcohol use scale (AUDIT), self efficacy scale for HIV risk behaviour, HIV/AIDS knowledge index, HIV and sexuality communication comfort index, HIV risk perception scale, TB social distance scale, HIV/AIDS stigma scale.

APPENDIX A

Nominal

A variable can be treated as nominal when its values represent categories with no intrinsic ranking; for example, the department of the company in which an employee works. Examples of nominal variables include region, zip code, or religious affiliation.

Ordinal

A variable can be treated as ordinal when its values represent categories with some intrinsic ranking; for example, levels of service satisfaction from highly dissatisfied to highly satisfied. Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores. For ordinal string variables, the alphabetic order of string values is assumed to reflect the true order of the categories. For example, for a string variable with the values of low, medium, high, the order of the categories is interpreted as high, low, medium which is not the correct order. In general, it is more reliable to use numeric codes to represent ordinal data.

Scale

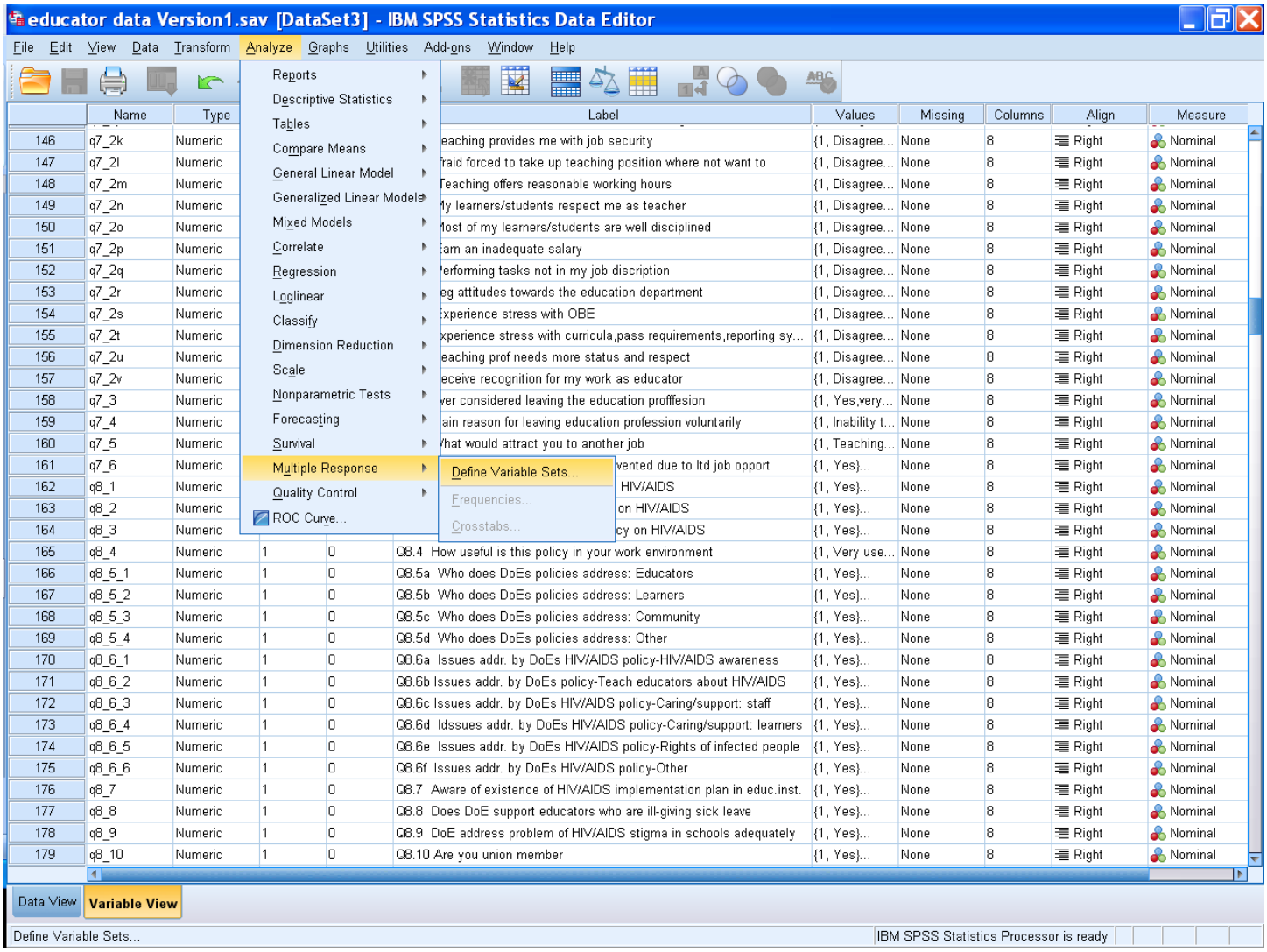
A variable can be treated as scale when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate. Examples of scale variables include age in years and income in thousands of dollars.

APPENDIX B

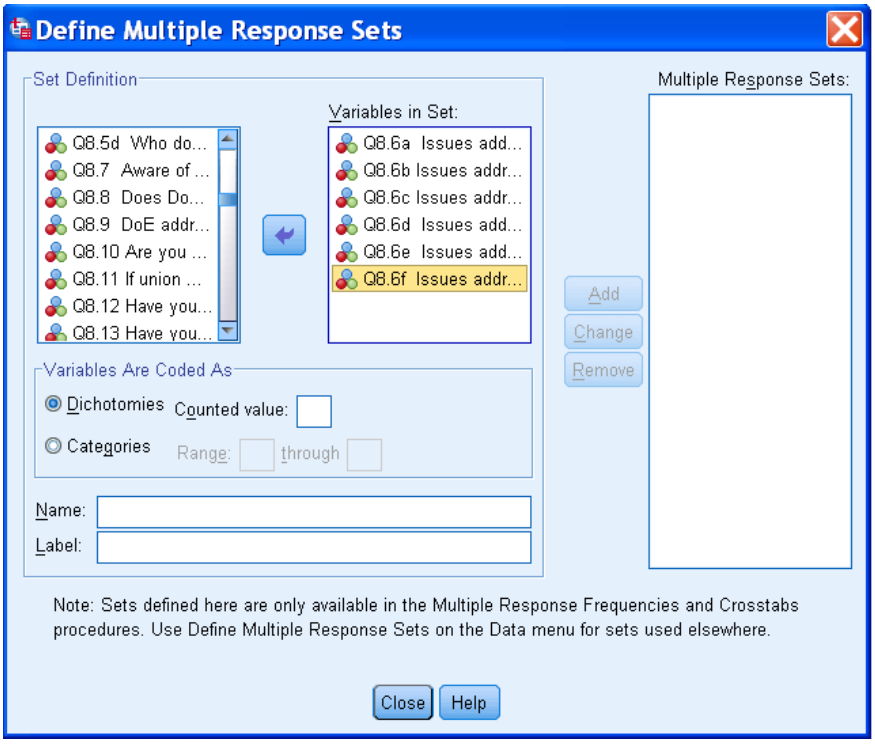
Analysing multiple response questions in SPSS:

Multiple response questions:
ELRC Educators questionnaire. Q8_6_1 – Q8_6_6.

If you want to analyse multiple response questions to get an overview of the responses, you can use SPSS multiple response set functionality under *Analyze, Multiple Response, Define Variable Sets:*



Select the variables you will need for the multiple response set: and move to the *Variables in Set:* box by clicking the ► arrow.

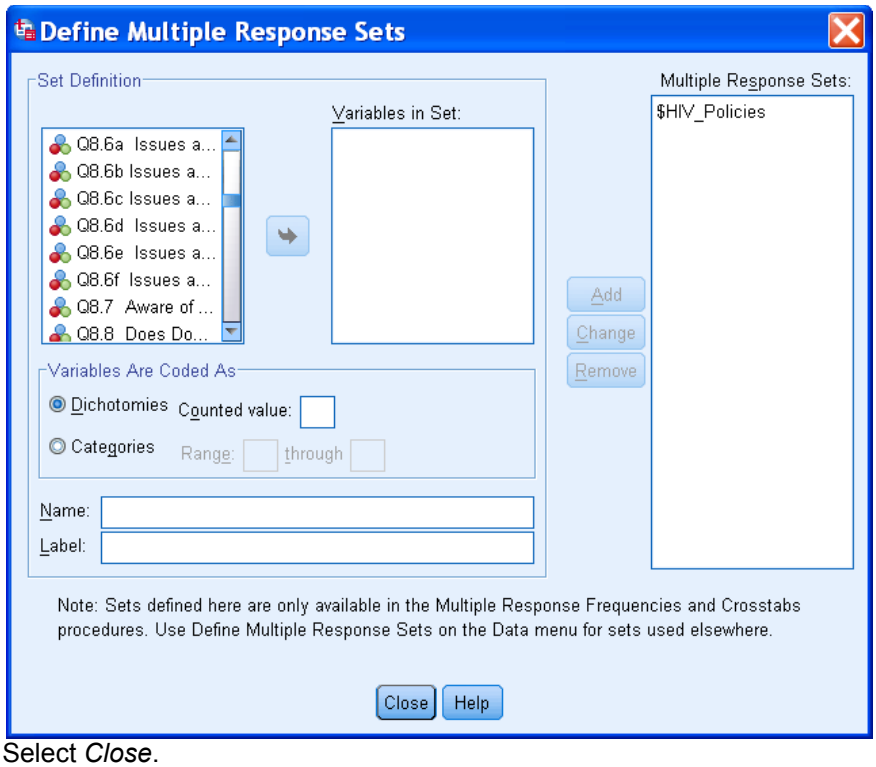


In this example *Dichotomies* will be selected, when answer options are all coded as 1 or left blank, *Counted value* will be assigned as 1.

Categories will be selected, when answer options are categorised, e.g. coded from 1 to say 20 and the Range will be specified.

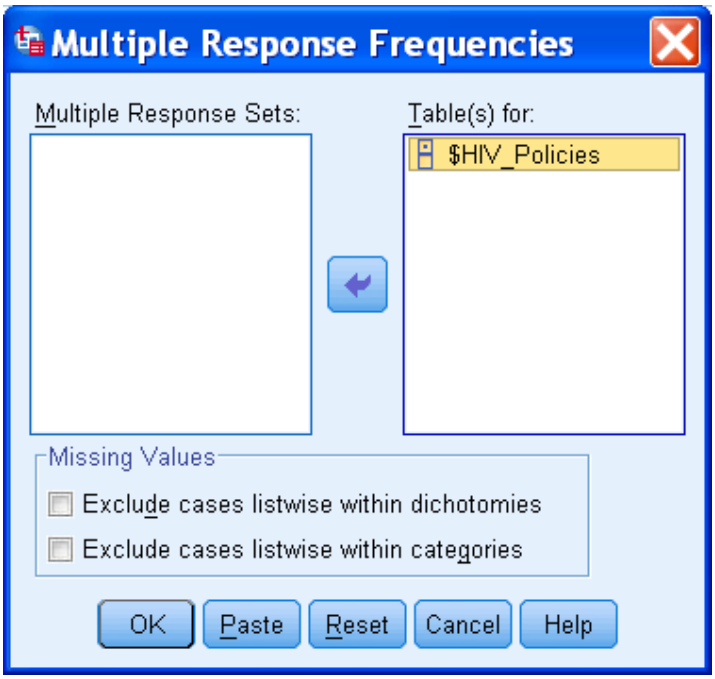
Assign a variable name in the *Name* box (e.g. HIV_Policies) and a description in the *Label* box.

Click **Add** to add the variable HIV_Policies to the Multiple Response Sets box



Select *Close*.

To run a frequency or cross tabulation:
Analyze, Multiple response, select *Frequencies….*or *Crosstabs…:*
'The created variable HIV_Policies can be used in a frequency table or crosstab.



Example of frequency output for multiple response questions:

$HIV_Policies Frequencies

		Responses		Percent of Cases
		N	Percent	
issues addressed by the Department of Ed[a]	Q8.6a Issues addr. by DoEs HIV/AIDS policy-HIV/AIDS awareness	8763	33.0%	82.9%
	Q8.6b Issues addr. by DoEs policy-Teach educators about HIV/AIDS	5496	20.7%	52.0%
	Q8.6c Issues addr. by DoEs HIV/AIDS policy-Caring/support: staff	4243	16.0%	40.2%
	Q8.6d Idssues addr. by DoEs HIV/AIDS policy-Caring/support: learners	4134	15.5%	39.1%
	Q8.6e Issues addr. by DoEs HIV/AIDS policy-Rights of infected people	3338	12.6%	31.6%
	Q8.6f Issues addr. by DoEs HIV/AIDS policy-Other	618	2.3%	5.8%
Total		26592	100.0%	251.7%

a. Dichotomy group tabulated at value 1.

Note that percentages do not add up to 100.0% in the 'Percent of Cases' column as one person could select more than one answer option. (The above is an example and calculations were not performed on weighted data.)