Variable and Value Labelling the School Register of Needs 1996

Alex Montgomery

February 21, 2012

DataFirst

1. Introduction

In the School Register of Needs Survey 1996 (SRNS) dataset, a number of variables that define schools into various categories were incorrectly specified with string variables that vaguely reflected these categories. In many cases, respondents that should probably be grouped together are spuriously defined within different categories because of slight differences in their string values (despite those string values suggesting that they are part of semantically similar groups). For example, some schools were assigned the string value "English" for the variable "medium" (which reflects the language in which students are instructed within the enumerated school), but a small number were defined as "EnglishEnglish". The questionnaire indicates that, of these two string values, only "English" is a valid categorical. Furthermore, it suggests that this string value should actually be coded as an integer value of 1. In this example, respondents with values "English" and "EnglishEnglish" are both recoded as "1", which is then assigned the value label "English".

We are agnostic as to the source of these errors, except that they are assumed to be *actual* errors (and not intentional classifications - we assume that there is no such medium of instruction as "EnglishEnglish", to relate this to the above example). Some assumptions were made to assign as many of these variables into legitimate categories as possible, with some of these assumptions less potentially contentious than others. Given the dearth of metadata on the survey, it is necessary to communicate the recoding process in order to empower researchers to make their own decisions regarding the plausibility of the data editing process. This process separates the latest version of the data from those that were previously available through DataFirst.

This version also includes as much information on the variable and value labels as possible, given the limited information that can be reliably related from the questionnaire provided to the dataset. Some variable names could unambiguously attached to questions observed in the questionnaire, but others were less clear. Occasionally, information on the categorical values was used to deduce the variable label. This process requires documentation for the same reasons as those mentioned previously.

2. Recoding String Variables as Integers

Tables 1-5 summarise information regarding the process of recoding strings as integers, as well as the way in which those integer values were labelled. In some cases, multiple string values were mapped onto a single integer value with a single corresponding value label. The process underlying the mislabelling of certain strings can easily be attributed to typos or spelling errors for most of these variables. However, in the "school type" variable the source of the bizarre string values observed in this dataset is probably related to the process by which those strings were generated. Given that the strings were created based on the integer values in the original data (instead of labelling the integer values!), and that some of the intended "labels" also contain integer values, it is possible that multiple runs of a code that converted integer values to an array

of predefined string values could produce the strange strings observed. To illustrate, imagine the code specified that all values of "2" should be replaced by "Senior Primary (Std 3 - Std 5)". This would produce string values that read, "Junior Primary (Sub A - Std Senior Primary (Std 3 - Std 5))" corresponding to integer values of "1" (which is meant to be "Junior Primary (Sub A - Std 2)". Given that those integers would *also* be redefined with this code, it is simple to see how this process would snowball to produce the string variables related in Table 1. Fortunately, the plausibility of this explanation means that the first part of the variable label is preserved, and can be used to assign the true value label to the proper integer value.

3. Value and Variable Labelling

Variable and value labelling required an identification of variables in the dataset using only the questionnaire. In some cases, the correspondence between certain variable names and their section in the questionnaire was difficult to discern. Those variables remain unlabelled. Those that are more easily identifiable have been labelled in a way that is summarised in Table 6 in the Appendix, which also includes information on the value label name assigned to each variable. Tables 7-12 detail those value labels.

A. Recoded Variables

Table 1: The "type" variable, "Schooltype" value label

Original String Observed in Datafile	Assumed Meaning and Variable Label Assigned	Recoded as Integer Value
Junior Primary (Sub A - Std Senior Primary (Std Primary (Sub A - Std Senior Secondary (Std 9 - Std 10)) - Std Senior Secondary (Std 9 - Std 10)))	Junior Primary (Sub A - Std 2)	1
Junior Primary (Sub A - Std Senior Primary (Std Primary (Sub A - Std Senior Secondary (Std 9 - Std 10)) - Std Senior Secondary (Std 9 - Std 10)))Junior Secondary (Std Secondary (Std 6 - Std 10) - Std Combined (Sub A - Std 10))		
Senior Primary (Std Primary (Sub A - Std Senior Secondary (Std 9 - Std 10)) - Std Senior Secondary (Std 9 - Std 10))	Senior Primary (Std 3 - Std 5)	2
Primary (Sub A - Std Senior Secondary (Std 9 - Std 10))	Primary (Sub A - Std 5)	3
Junior Secondary (Std Secondary (Std 6 - Std 10) - Std Combined (Sub A - Std 10))	Junior secondary (Std 5 - Std 8)	4
Senior Secondary (Std 9 - Std 10)	Senior secondary (Std 9 - Std 10)	5
Secondary (Std 6 - Std 10)	Secondary (Std 6 - Std 10)	6
Intermediate (Std 5 - Std 7)	Intermediate (Std 5 - Std 7)	7
Combined (Sub A - Std 10)	Combined (Sub A - Std 10)	8
Technical colleges 9	Technical colleges	9
Teacher training colleges	Teacher training colleges	10
Technikons	Technikons	11
Universities	Universities	12
Other non-formal training institutions	Other non-formal training institutions	13
Special schools	Special schools	14

Table 2: The "medium" variable, "Language" value label

Original String Observed in Datafile	Assumed Meaning and Variable Label Assigned	Recoded as Integer Value
EnglishEnglish English	English	1
Afrikaans	Afrikaans	2
isiXhosa	isiXhosa	3
isiZulu	isiZulu	4
Sesotho sa Leboa	Sesotho sa Leboa	5
Sesotho	Sesotho	6
Setswana	Setswana	7
Xitsonga	Xitsonga	8
Tshivenda	Tshivenda	9
SiSwati	SiSwati	10
isiNdebele	isiNdebele	11
EnglishAfrikaans	Afrikaans/English	12
African Languages	African Languages	13

Table 3: The "telcom" variable, "Telecom" value label

Original String Observed in Datafile	Assumed Meaning and Variable Label Assigned	Recoded as Integer Value
Telephone	Telephone	1
2-Way Radio	2-Way Radio	2
Cell phone	Cell phone	3
Fax	Fax	4
Modem	Modem	5
None	None	6
Other	Other	7

Table 4: The "owner" variable, "Owner" value label

Original String Observed in Datafile	Assumed Meaning and Variable Label Assigned	Recoded as Integer Value
Government Government0	Government	1
Community	Community	2
Farm	Farm	3
Mine	Mine	4
Factory	Factory	5
Church	Church	6
School/ Governing Body	School/Governing Body	7
Welfare	Welfare	8
Hospital	Hospital	9
Other	Other	10

Table 5: The "power" variable, "Power" value label

Original String Observed in Datafile	Assumed Meaning and Variable Label Assigned	Recoded as Integer Value
Wired & supplied with electricity	Wired and supplied with electricity	1
Wired but not supplied with electricity	Wired but not supplied with electricity	2
Not wired and/ or no electricity available	Not wired and/or no electricity available	3
Generators	Generators	4
Other energy sources	Other	5

B. Variable Labelling

Table 6: Variable labels

Variable	Variable Label	Value Label
contact	Contact person at school	
include_in_sample		
school	Name of school	
post_1	Street address	
post_2	Suburb	
post_3		
postcode	Postal Code	
phys_1		
phys_2		
phys_3		
physcode		
condit	Condition of Buildings	SchoolCondition
rf_type	Roof type	EquipCondition
rf_cond	Roof condition	EquipCondition
wd_cond	Window condition	EquipCondition
dr_cond	Door condition	EquipCondition
wl_type	Wall type	EquipCondition
wl_cond	Wall condition	EquipCondition
fr_cond	Furniture condition	EquipCondition
fl_type	Floor type	EquipCondition
fl_cond	Floor condition	EquipCondition
tl_type	Toilet type	EquipCondition
tl_cond	Toilet condition	EquipCondition
ceiling	Ceiling	Ceiling
o_type	Other type	EquipCondition
o_cond	Other condition	EquipCondition
medium	Medium of instruction	Language
telcom	Telecommunication	Telecom
owner	Owner	Owner
stdprov	Adequacy of student worktop/chair	Adequacy
stdshort	Estimated shortfall of student worktop/chair provision	
eduprov	Adequacy of teach worktop/chair	Adequacy
edushort	Estimated shortfall of teacher worktop/chair provision	

Table 6: Variable labels (continued)

Variable	Variable Label	Value Label
cupprov	Adequacy of cupboard/cabinet in areas of instruction	Adequacy
cupshort	Estimated shortfall of cupboard/cabinet provision in areas of instruction	
dskprov	Adequacy of administrative desk provision	Adequacy
dskshort	Estimated shortfall of administrative desk provision	
typprov	Adequacy of typewriter/computer provision for administrators	Adequacy
typshort	Estimated shortfall of typewriters/computers for administrators	
cpaprov	Adequacy of cupboard/cabinet in administrative areas	Adequacy
cpashort	Estimated shortfall of cupboard/cabinet provision in administrative areas	
dupop	Number of operational duplicators/photocopiers	
dupnop	Number of nonoperational duplicators/photocopiers	
ovhop	Number of operational overhead projectors	
ovhnop	Number of nonoperational overhead projectors	
radop	Number of operational radios	
radnop	Number of nonoperational radios	
filop	Number of operational film projectors	
filnop	Number of operational film projectors	
tvop	Number of operational television sets	

Table 6: Variable labels (continued)

Variable	Variable Label	Value Label
tvnop	Number of nonoperational television sets	
tappop	Number of operational tape recorders	
tapnop	Number of nonoperational tape recorders	
vcrop	Number of operational VCRs	
vcrnop	Number of nonoperational VCRs	
compop	Number of operational computers	
compnop	Number of nonoperational computers	
othop		
othnop		
power	How school is powered	Power
textbook	Adequacy of textbook provision	Adequacy
stnary	Adequacy of stationary provision	Adequacy
mater	Adequacy of materials (used for science projects, experiments, etc.)	Adequacy
equip	Adequacy of general equipment (scissors, science kits, musical instruments, etc.)	Adequacy
mediac	Adequacy of media collection (library books, journals, records, transparencies, etc.)	Adequacy
mediae	Adequacy of equipment in media centre	Adequacy
funding	How the school is funded	Funding
classp		
fence	School fence properties	Fencing
tarroad	Tar road condition	RoadCond
dirtroad	Dirt road condition	RoadCond
noroad	No road access	RoadCond
footpath	Footpath condition	RoadCond
rural__urban	Settlement type in which school is located	

Table 6: Variable labels (continued)

Variable	Variable Label	Value Label
type	Type of school	
province	Province	
telcode	Telephone code	
telnum	Telephone number	
faxcode	Fax code	
faxnum	Fax number	
compn		
compa		
computers_1_	Computers for teaching purposes	
toteduc	Number of educators	
register		
totgirl	Number of girls	
totboy	Number of boys	
total_student	Total number of students	
__student__comp		
__stud_teacher		
totclass		
avrclass		

C. Value Labelling

Table 7: "EquipCondition" value label

Integer Value	Value Label
1	No maintenance needed
2	Need maintenance
3	Need maintenance and structural repair
4	Beyond repair

Table 8: "Ceiling" value label

Integer Value	Value Label
1	Fitted
2	Non-fitted

Table 9: "Adequacy" value label

Integer Value	Value Label
1	Adequate
2	Inadequate
3	None

Table 10: "Funding" value label

Integer Value	Value Label
1	State
2	State-aided
3	Private

Table 11: "Fencing" value label

Integer Value	Value Label
1	High (3m); good condition
2	High (3m); bad condition
3	Medium (1.5m); good condition
4	Medium (1.5m); bad condition
5	No fence
6	Other

Table 12: "RoadCond" value label

Integer Value	Value Label
1	Good
2	Bad