ROSES

The Relevance of Science Education - Second

ROSES Codebook

Data entry, cleaning and reporting

Anders Jidesjö, Magnus Oskarsson & Anna-Karin Westman

Jan 2021

Mittuniversitetet
MID SWEDEN UNIVERSITY

ROSES
The Relevance of Science Education - Second

ROSES Data Entry and Codebook

ROSES Data Entry - Introduction

This text contains information about how to work with data entry to create the file that should be sent back to the ROSES organizers. To assist we provide the ROSES Codebook and ROSES Data entry files. These documents provide information about the character of data and how data should be compiled, processed, and reported. They should be used together with the ROSES master questionnaire to complete the process of data collection.

In a large-scale research project, such as ROSES, it is important that answers of respondents in different countries can be easily interpreted which is the basic purpose of coding. It is also important that data are entered in data files correctly, to be able to compare results. The purpose with coding the ROSES data the way that it is described in the Codebook, is to reduce the large quantity of information into a format that can be analysed and compared in the international setting.

ROSES Codebook

The ROSES Codebook contains technical information about the content and structure of data for each variable. The order of variables follows the structure in the ROSES master quest. In the Codebook they are given a style to be self-explanatory. The variable name is given together with variable label, accepted values and value for missing data. Each variable in the Codebook explain how responses in the questionnaire should be coded. This information also serve as a documentation of how the ROSES data entry file is structured.

In the ROSES questionnaire there are four open questions where data are in a written qualitative data format, e.g. L1, L2, J1 and J2. Themes and patterns within these answers can be interesting to analyse in relation with other ROSES data, but also to compare between countries. Therefore, we will test coding categories empirically to create a structure that hopefully work out well. This means that the coding of these categories should not be sent back now. We will come back with further information. For one of these items, Question L1 “What occupation would you like to have in the future?” the coding can start. We suggest using the international “ISCO 08 codes” for this purpose.

ROSES Data entry files

How data is imported and compiled in the ROSES data file is important. False, uncleaned, or incorrect data has an impact of data quality and subsequently influence the validity of results. As ROSES is an international explorative research project, it is also important to have data entered the same way in being able to carry out international comparisons. Therefore, we provide the ROSES data entry files. Some brief information about data entry and methodological considerations are given below.

Data quality

The validity of ROSES data is dependent on entering the data into the data entry file in the format given in the Codebook. The Codebook gives what is called "data type constraints".

The accuracy of data is about coding data with the correct value. As valid values are given it should be easy to detect invalid values to be omitted. If data have been collected electronically this is a more straightforward procedure. If questionnaires have been distributed manually in a classroom and collected, data entry must be carried out by entering each response with the correct value in the data entry file. In any case, false or incorrect responses must be cleaned.

The completeness is another aspect to consider. Missing responses will occur for various reasons, often unknown to the researcher. In ROSES it is not possible to come back to respondents and ask them to complete. Missing data should be coded. The amount of missing data is another aspect to consider. There is no rule in ROSES where a respondent should be omitted from the data set due to missing data. In broad outlines, if for example a respondent has not given answers to items in one part of the questionnaire but has done it in other parts in a correct manner, we suggest that such responses are included. We rely on researchers in each country to judge about completeness. If respondents should be included or not is depending on the character of responses. This means that completeness in ROSES is to judge about what is called "data consistency" among respondents.

Data entry and Data processing

ROSES data can be entered in two different files. The file 'ROSES_Country.sav' is in SPSS format and 'ROSES_country.xls' in Microsoft Excel format. If possible, we prefer that you use SPSS. If you have experience in SPSS, the empty data entry file is self-explanatory.

In the SPSS data entry file, the variables have names with a maximum length

of eight characters. Most variable names are composed of the question number with capital letters and the item number. Each variable in the data entry file correspond with the questionnaire item by a label. The Excel file was created from the SPSS file and converted to Excel format. The format transformation caused some loss of information, e.g. the labels are not sustained in the Excel file. For this reason, if you have the SPSS software available, we recommend the SPSS version for data entry.

As a rule, the respondent's tick in one of the four response categories in the ROSES questionnaire, which gives the value to be entered in the data entry file. A tick in the first box is coded as '1', a tick in the second box as '2' etc. The coding is given by the Codebook. In addition, in the data entry file, beneath the heading 'Valid value' you find the legal values of the variable.

The Likert scales in use in the ROSES questionnaire have headings only for the outermost values. The two in the middle are untitled. If a respondent has given no response or multiple responses to one item, it should be regarded as a missing value. According to the Codebook, missing values should be coded as '9'.

Processing data can be time consuming and exhausting but is important to explore the data and detect errors. In ROSES almost all data is quantitative and coded in the same way, which allows for "data profiling". This means to summarize data and inspect it. Check columns and rows and search for inconsistencies and patterns. Sort data in different columns and check the data consistency. This procedure can detect if data conforms to the desired standard. You could also do some basic descriptive statistics to identify data outliers which can be reinspected. Such procedures are called "data visualizations".

Data cleaning

Data cleaning can imply several measures. Data can for example be removed or corrected. What to be done with data depends on the character. In ROSES we rely heavily on researchers in each country to judge on these matters.

Irrelevant data is probably not an issue if it have been collected digitally. If questionnaires were distributed manually in a classroom there can be written information in the questionnaires not belonging to the study. As a guiding principle, such information should not be regarded.

Data duplicates can occur for example if the data collection platform allows respondents to click twice when submitting the questionnaires. This often occurs when respondents are unsure if their answers were submitted or not. Duplicates can also occur due to how the questionnaires was distributed to respondents. Duplicates should be detected and probably removed. Type conversion can happen if numbers are not coded with numerical values given in the Codebook. Conversion can be easily detected by inspecting for example columns and check for invalid values. It is often recommended to involve a colleague to inspect data to detect such errors.

In ROSES, data standardization is achieved by giving data correct numerical values according to the Codebook. All values have a certain unit. Missing values should be coded or dropped. If dropped the whole response will be deleted. In ROSES it is not allowed to impute data in the cleaning process. To impute data could be used later in analyses but never in the cleaning and data entry processes.

Data outliers is about divergent responses. A respondent can for example suddenly have responded a part of the ROSES questionnaire with the same value 15 times in a row. It is not easy to judge how outliers should be treated but methodologically they are not the same as missing values. If outliers should be included or omitted must be decided by researchers in each country and is dependent on the character. We recommend discussing such matters with colleagues to come to reasonable decisions. Another way to check for divergent responses and completeness is to check if the platform in use allows to check for start and stop time of respondents. If there is a start time but no stop time, it can be a reason to analyse a respondent's data consistency more critically.

Verifying the data

After data cleaning procedures have been applied it is recommended to verify the data entry file. This is done by going back again and reinspect if any data violates constraints given by the Codebook. If errors are detected, manual correction must be done.

Reporting the data

When having all data processed, cleaned, and entered it is recommended to document what was done. This can be interesting in future discussions, for example to understand why things happened in the first place. We recommend researchers in each country to write a page about how data was collected and under what conditions. Such circumstances have an influence on data quality. This text could also contain some information about how data were cleaned and entered in a data file. This "technical report" should be sent with the national data file. We plan to publish such technical reports for each country on the ROSES web page.

The national data file should be sent to the ROSES organizers together with the descriptive text about data collection and data entry. Please name the data file indicating the name of your country, as an example 'ROSES Sweden.sav'.

In the following pages the concrete Codebook is provided, with names and numbers for each variable in the questionnaire

ROSES **CODEBOOK 2020-11-04**

Number	Variable	Heading
1	V1	**ID**
	Label	*Respondant unique id*
	Type	Numeric
	Measurement	Scale
2	V2	**Sex**
	Label	*Boy/Girl*
	Type	Numeric
	Measurement	Nominal
		1 Girl
		2 Boy
		3 Other
3	V3	**Age**
	Label	*My age (years)*
	Type	Numeric
	Measurement	Ordinal
4	V4	**Country**
	Label	*Which country do you live in?*
	Type	String
	Measurement	Nominal
5	V5	**School**
	Label	*The name of my school is:*
	Type	String
	Measurement	Nominal
6	V6	**Language**
	Label	*Language*
	Type	String
	Measurement	Nominal
7	V7	**Answer Created**
	Label	*Answer Created*
	Type	Date
	Measurement	Ordinal
8	V8	**Answer Completed**
	Label	*Answer Completed*
	Type	Date
	Measurement	Ordinal

		A. What I want to learn about. *How interested are you in*
9	**A1**	A1. Chemicals, their properties and how they react
10	**A2**	A2. The inside of the earth
11	**A3**	A3. How mountains, rivers and oceans develop and change
12	**A4**	A4. Clouds, rain and the weather
13	**A5**	A5. The origin of life on earth
14	**A6**	A6. Heredity, and how genes influence how we develop
15	**A7**	A7. Sex and reproduction
16	**A8**	A8. Birth control
17	**A9**	A9. Atoms and molecules
18	**A10**	A10. The evolution of animals and plants
19	**A11**	A11. Light around us that we cannot see (infrared, ultraviolet)
20	**A12**	A12. How different musical instruments produce different sounds
21	**A13**	A13. Black holes, supernovas and other spectacular objects in outer space
22	**A14**	A14. How meteors, comets or asteroids may cause disasters on earth
23	**A15**	A15. Earthquakes and volcanoes
24	**A16**	A16. How people, animals, plants and the environment depend on each other
25	**A17**	A17. Tornados, hurricanes and cyclones
26	**A18**	A18. Brutal, dangerous and threatening animals
27	**A19**	A19. Poisonous plants in my area
28	**A20**	A20. How the atom bomb functions
29	**A21**	A21. Explosive chemicals
30	**A22**	A22. Biological and chemical weapons and what they do to the human body
31	**A23**	A23. The effect of strong electric shocks and lightning on the human body
32	**A24**	A24. How it feels to be weightless in space
33	**A25**	A25. How the eye can see light and colours
34	**A26**	A26. What to eat to keep healthy and fit
35	**A27**	A27. Eating disorders like anorexia or bulimia
36	**A28**	A28. The ability of lotions and creams to keep the skin young
37	**A29**	A29. How to exercise to keep the body fit and strong
38	**A30**	A30. Plastic surgery and cosmetic surgery
39	**A31**	A31. How radiation from solariums and the sun might affect the skin
40	**A32**	A32. How the ear can hear different sounds
41	**A33**	A33. Rockets, satellites and space travel
42	**A34**	A34. How X-rays, ultrasound, etc. are used in medicine
43	**A35**	A35. How a nuclear power plant functions

Type	Numeric
Measure	Ordinal
1	Not interested
2	2
3	3
4	Very Interested
9	Missing

		B. My future job. How important are the following issues for your potential future occupation or job?		
44	**B1**	B1. Working with people rather than things		
45	**B2**	B2. Helping other people	Type	Numeric
46	**B3**	B3. Working with something easy and simple	Measure	Ordinal
47	**B4**	B4. Working with machines or tools	1	Not important
48	**B5**	B5. Working artistically and creatively in art	2	2
49	**B6**	B6. Using my talents and abilities	3	3
50	**B7**	B7. Making, designing or inventing something	4	Very Important
51	**B8**	B8. Coming up with new ideas	9	Missing
52	**B9**	B9. Having lots of time for my friends		
53	**B10**	B10. Making my own decisions		
54	**B11**	B11. Working independently of other people		
55	**B12**	B12. Working with something I find important and meaningful		
56	**B13**	B13. Working with something that fits my attitudes and values		
57	**B14**	B14. Having lots of time for my family		
58	**B15**	B15. Working with something that involves a lot of travelling		
59	**B16**	B16. Working at a place where something new and exciting happens frequently		
60	**B17**	B17. Earning lots of money		
61	**B18**	B18. Controlling other people		
62	**B19**	B19. Becoming famous		
63	**B20**	B20. Having lots of time for my interests, hobbies and activities		
64	**B21**	B21. Becoming 'the boss' at my job		
65	**B22**	B22. Developing or improving my knowledge and abilities		
66	**B23**	B23. Working as part of a team with many people around me		
		C. What I want to learn about. How interested are you in learning about the following?		
67	**C1**	C1. How a cell phone works		
68	**C2**	C2. The possibility of life outside earth	Type	Numeric
69	**C3**	C3. Life and death and the human soul	Measure	Ordinal
70	**C4**	C4. Why we dream while we are sleeping, and what the dreams may mean	1	Not interested
71	**C5**	C5. Ghosts and witches, and whether they may exist	2	2
			3	3
72	**C6**	C6. Thought transference, mind-reading, sixth sense, intuition, etc.		
73	**C7**	C7. Why the stars twinkle and the sky is blue	4	Very Interested
74	**C8**	C8. Why we can see the rainbow	9	missing
75	**C9**	C9. Properties of gems and crystals and how these are used for beauty		
76	**C10**	C10. How to attain a more sustainable society		
77	**C11**	C11. How emissions of carbon dioxide can affect the climate		

		D. Me and the environmental challenges. To what extent do you agree with the following statements about problems with the environment (pollution of air and water, overuse of resources, global changes of the climate	
78	**D1**	D1. Threats to the environment are not my business	
79	**D2**	D2. Environmental problems make the future of the world look bleak and hopeless	Type Numeric
80	**D3**	D3. Science and technology can solve all environmental problems	Measure Nominal
81	**D4**	D4. I am willing to have environmental problems solved even if this means sacrificing many goods	1 Disagree
82	**D5**	D5. I can personally influence what happens with the environment	2 2
83	**D6**	D6. We can still find solutions to our environmental problems	3 3
84	**D7**	D7. People worry too much about environmental problems	4 Agree
85	**D8**	D8. Environmental problems can be solved without big changes in our way of living	9 Missing
86	**D9**	D9. People should care more about how to protect the environment	
87	**D10**	D10. It is the responsibility of the rich countries to solve the environmental problems of the world	
88	**D11**	D11. Environmental problems should be left to the experts	
89	**D12**	D12. I am optimistic about the future	
90	**D13**	D13. Nearly all human activity is damaging for the environment	

		E. What I want to learn about. How interested are you in learning about the following?
91	**E1**	E1. How the sunset colours the sky
92	**E2**	E2. The greenhouse effect and how it may be changed by humans
93	**E3**	E3. What can be done to ensure clean air and safe drinking water
94	**E4**	E4. How technology helps us to handle waste, garbage and sewage
95	**E5**	E5. How to control epidemics and diseases
96	**E6**	E6. Sexually transmitted diseases and how to be protected against them
97	**E7**	E7. What we know about HIV/AIDS and how to control it
98	**E8**	E8. How alcohol and tobacco might affect the body
99	**E9**	E9. How different narcotics might affect the body
100	**E10**	E10. The possible radiation dangers of mobile phones and computers
101	**E11**	E11. How to protect endangered species
102	**E12**	E12. How to improve the harvest in gardens and farms
103	**E13**	E13. Medicinal use of plants
104	**E14**	E14. Organic and ecological farming without use of pesticides and artificial fertilizers
105	**E15**	E15. How energy can be saved or used in a more effective way E16.
106	**E16**	Renewable sources of energy from the sun and the wind
107	**E17**	E17. How different sorts of food are produced, conserved and stored
108	**E18**	E18. How my body grows and matures
109	**E19**	E19. Animals in my area
110	**E20**	E20. Plants in my area
111	**E21**	E21. Detergents, soaps and how they work
112	**E22**	E22. Human exploration of Mars
113	**E23**	E23. The impact of the Internet on society
114	**E24**	E24. Biological and human aspects of abortion
115	**E25**	E25. Benefits and possible hazards of gene modification (GMO) in farming
116	**E26**	E26. Why religion and science sometimes are in conflict
117	**E27**	E27. Why scientists sometimes disagree
118	**E28**	E28. Famous scientists and their lives
119	**E29**	E29. Big blunders and mistakes in research and inventions
120	**E30**	E30. Inventions and discoveries that have changed the world
121	**E31**	E31. Very recent inventions and discoveries in science and technology
122	**E32**	E32. Phenomena that scientists still cannot explain

Type	Numeric
Measure	Ordinal
1	Not interested
2	2
3	3
4	Very Interested
9	Missing

		F. My science classes. To what extent do you agree with the following statements about the science that you may have had at school?		
123	**F1**	F1. School science is a difficult subject		
124	**F2**	F2. School science is interesting	Type	Numeric
125	**F3**	F3. School science has opened my eyes to new and exciting jobs	Measure	Ordinal
126	**F4**	F4. I like school science better than most other subjects	1	Not important
127	**F5**	F5. The things that I learn in science at school will be helpful in my everyday life	2	2
128	**F6**	F6. School science has made me more critical and sceptical	3	3
129	**F7**	F7. School science has increased my curiosity about things we cannot yet explain	4	Very Important
130	**F8**	F8. School science has shown me the importance of science for our way of living	9	Missing
131	**F9**	F9. School science has taught me how to take better care of my health		
132	**F10**	F10. I would like to become a scientist		
133	**F11**	F11. I would like to get a job in technology		
134	**F12**	F12. School science has helped me to understand sustainability solutions in my everyday life		
		G. My opinions about science and technology. *To what extent do you agree with the following statements?*		
135	**G1**	G1. Science and technology are important for society		
136	**G2**	G2. Science and technology will find cures to diseases such as HIV/AIDS, cancer, etc.	Type	Numeric
137	**G3**	G3. Thanks to science and technology, there will be greater opportunities for future generations	Measure	Ordinal
138	**G4**	G4. Science and technology make our lives healthier, easier and more comfortable	1	Not important
139	**G5**	G5. The benefits of science are greater than the harmful effects it could have	2	2
140	**G6**	G6. Science and technology will help to eradicate poverty and famine in the world	3	3
141	**G7**	G7. Science and technology can solve nearly all problems	4	Very Important
142	**G8**	G8. Science and technology are helping the poor	9	Missing
143	**G9**	G9. Science and technology are the cause of the environmental problems		
144	**G10**	G10. A country needs science and technology to become developed		
145	**G11**	G11. Science and technology benefit mainly the developed countries		
146	**G12**	G12. We should always trust what scientists have to say		
147	**G13**	G13. Learning about sustainability is important		

		H. My experiences of social and digital media. Social and digital media refers to the technologies that facilitate the creation and sharing of information, photos, opinions, chat and other forms of expression via virtual communities and networks.		
			Type	Numeric
			Measure	Ordinal
			1	1-2 hours
			2	2-4 hours
			3	4-6 hours
			4	more than 6 hours
148	**H1**	H1. In school I am using social and digital media in my schoolwork		
149	**H2**	H2. I am using social and digital media at home	9	Missing
		In my Science classes I make use of:		
150	**H3**	H3. Internet search	Type	Numeric
151	**H4**	H4. Online resources (e.g. NASA, Cern… Add national examples)	Measure	Ordinal
152	**H5**	H5. Social media (e.g. Facebook, Snapchat, Instagram… Add national examples)	1	Never
153	**H6**	H6. E-Textbooks	2	2
154	**H7**	H7. Video clips on the web (e.g. YouTube)	3	3
155	**H8**	H8. Computer games	4	Often
156	**H9**	H9. Programming and coding	9	Missing
157	**H10**	H10. Digital fabrication tools (e.g. 3D-printers & Laser cut)		
158	**H11**	H11. Microcontrollers (e.g. Arduino, Mindstorms Add national examples)		
159	**H12**	H12. Graphic and video design		
		The information I find on social and digital media for learning science and technology:		
			Type	Numeric
			Measure	Ordinal
			1	Disagree
			2	2
			3	3
			4	Agree
160	**H13**	H13. Is reliable	9	Missing
161	**H14**	H14. Is better than my science textbook in school		
162	**H15**	H15. Is encouraged by the school		
163	**H16**	H16. Could be better used for learning in school		

		I. My informal science experiences. To what extent do you agree with the following statements? I learn science when I:	
164	**I1**	I1. Visit a zoo, aquarium	Type: Numeric
165	**I2**	I2. Visit a planetarium	Measure: Ordinal
166	**I3**	I3. Visit a science center	1: Disagree
167	**I4**	I4. Visit a museum	2: 2
168	**I5**	I5. Visit a botanical garden, park or nature reserve	3: 3
169	**I6**	I6. Participate in a science event (science fair)	4: Agree
170	**I7**	I7. Participate in an after-school science club	9: Missing
171	**I8**	I8. Visit internet home pages	
172	**I9**	I9. Use social and digital media	
173	**I10**	I10. Read science magazines	
174	**I11**	I11. Play computer games	
175	**I12**	I12. Watch TV	
176	**I13**	I13. Watch video clips on the Internet	
177	**I14**	I14. Use social media sites (e.g. Facebook, Snapchat, Instagram…)	
		J. Myself as a scientist. Assume that you are grown up and work as a scientist. You are free to do research that you find important and interesting. Write some sentences about what you would like to do as a researcher and why.	
	J1	J1. I would like to	
	J2	J2. Because	**Will be coded later**
		K. How many books are there in your home?	
178	**K1**	K1. How many books are there in your home?	1: No Books 2: 1-10 Books 3: 11-50 books 4: 51-100 books 5: 101-250 books 6: 251-500 books 7: More than 500
		L. What occupation would you like to have in the future?	
179	**L1**	L1. What occupation would you like to have in the future?	Use ISCO-08 codes with at least two digits
	L2	L2. Because	Will be coded later