15 Bildung und Wissenschaft   Dezember 2007

PISA 2003

Public Use Sample (PUS)

Methodological Note

PISA 2003 - Public Use Sample (PUS)
Methodological Note

This note is intended to describe and briefly explain the major steps which led to obtaining a PUS from the original Swiss PISA 2003 dataset of 9th-grade students.

Deleted variables

Numerous variables were deleted, more precisely:

- variables regarded as not of interest to the general public (i.e. too much detail, useful only for very specific analyses);

- very technical variables not generally used for common analyses (i.e. WLE –Weighted Likelihood Estimates);

- variables containing similar information to other variables;

- variables, which are index components. Indices summarizing information were kept in the data-set, and the single component variables were eliminated.

About half (236/481) of the variables in the original PISA 2003 9th-grade dataset were kept in the PUS. A comparative table of the variables in the original dataset and the variables in the PUS is given in Appendix 1 to this document.

Note: a few variables were computed by the PISA National Project Management and subsequently added to the original dataset. Some were removed and some were kept in the PUS.

Recoded variables

In order to obtain larger categories and avoid combinations of identifying variables[1] with too few cases – the "unsafe combinations" –, 24 of the PUS variables were recoded.

The threshold value, below which a combination is regarded as unsafe, is set by the user on the basis of practical considerations and of his/her knowledge of the dataset. To create the PISA 2003 PUS, thresholds were set ($n1>n2>n3$):
- $n1$ observations for one-dimensional tables (one variable)
- $n2$ observations for two-dimensional tables (combinations of two variables)
- $n3$ observations for three-dimensional tables (combinations of three variables)

That means that a combination of two identifying variables was regarded as unsafe if there were cells containing less than $n2$ observations; a combination of three variables was considered unsafe if there

---

[1] Identifying variables are those that could make it possible to identify an individual. There are direct identifiers such as name, address and social security number, which are in any case suppressed, and indirect identifiers such as regional variables (i.e. residence), gender, nationality, age, occupation and education. Combinations of identifying variables can lead to re-identification.

were cells containing less than n3 observations and so on. It was not considered necessary to explore higher dimension tables.

Cells containing a number of scores less than a fixed threshold are also called "rare combinations". Rare combinations, i.e. a small number of individuals sharing certain characteristics, might make it easier for an intruder to re-identify a single individual and then to disclose sensitive information about him/her which is also contained in the dataset.

A list of the recoded variables with details about the new and the old categories is given in Appendix 2 to this document. For an overview of all the variables, in which the recoded ones were flagged, see also Appendix 1.

PRAMmed variables

The PRAM (Post Randomization) method was applied to 14 categorical variables, some of which had already been recoded. The PRAM method consists in randomly changing the score on a categorical variable, using a known probability mechanism. Once perturbation was introduced into the dataset it becomes more difficult for an intruder to identify (with certainty) records as corresponding to certain individuals in the population[2]. This raises the safety level in the dataset.

The probability of actually changing the score was kept small in our database (never over 20%) to avoid overly corrupting the variables.

The PRAMmed variables are listed below:

PRAMmed variables

	Variable	Variable label[3]
1	st01q02	Programme
2	st06q01	Father currently doing
3	st15q01	Country of birth self
4	st15q04	Country of birth age
5	st16q01	Language at home
6	st21q01	<ISCED 1> years
7	st22q02	Repeat <ISCED 2>
8	st25q03	Attend other
9	st36q01	Students in maths
10	ic01q01	Available at home ICa
11	age	Age of student
12	msecateg	Mother white collar/blue collar classification
13	hsecateg	Highest parent white collar/blue collar classification
14	natprog_rec	National school programme

Since the transition probabilities (from one category to another) are known, unbiased estimates of contingency tables can be obtained. Other, more elaborate techniques will be needed to compensate

---

[2] μ-ARGUS, User's Manual, CASC-Project, November 2004, page 12.

[3] The variable labels are the original ones and are not always self-explanatory. For more details and to better understand them, see the corresponding questions in the student questionnaire. The variable st06q01 corresponds, for instance, to question Q6 in the questionnaire, variable st22q01 to question Q22, first category (a), variable st22q02 to question Q22, second category (b) and so on. Questions starting with "Q" belong to sections A, B, C, D, E and F and are associated with variables whose names start with "st". Questions starting with "A" belong to section G and are associated with variables whose names start with "ic".

for the PRAM-perturbation in more complex analyses as, for instance, in loglinear models[4]. For an overview of all the variables, in which the PRAMmed ones were flagged, see also Appendix 1.

Local suppression

After recoding and "PRAMming" there were still some unsafe cells (cells containing a small number of cases) in our dataset. The software μ-ARGUS[5], which was used to create a PUS, proceeded to local suppression. Local suppression[6] is the last step in the disclosure control process; it consists of creating missing values by deleting one variable within the remaining unsafe combinations of variables. If there is more than one unsafe combination in a record and the unsafe combinations have a variable in common, μ-ARGUS will suppress the common variable. Otherwise, the software will choose one of the variables minimizing the loss of information. The information loss can be based either on an entropy function (defined in the software) – the variable with the lowest value of the function will be suppressed – or on "suppression weights" assigned by the user. Each weight represents the information loss, and the variable with the lowest loss will be suppressed. To produce the PISA 2003 PUS, we preferred to assign suppression weights rather than use the entropy function. Local suppression resulted in a further 13 missing values in addition to the existing ones. They were necessary for the protection of three-dimensional tables; the one- and two-dimensional tables were already safe.

The table below summarizes all the transformations applied to the variables.

Summary table of modifications to the variables

Variable	Variable label	Recoded	PRAMmed	Local Suppression
st01q02	<Programme> Q1b	X	X	
st05q01	Mother currently doing Q5	X		
st06q01	Father currently doing Q6	X	X	
st15q01	Country of birth Self Q15a_a	X	X	
st15q02	Country of birth Mother Q15a_b	X		
st15q03	Country of birth Father Q15a_c	X		
st15q04	Country of birth Age Q15b	X	X	
st16q01	Language at home Q16	X	X	
st18q01	Possessions cell phones Q18a	X		
st18q02	Possessions TV sets Q18b	X		
st18q03	Possessions computers Q18c	X		
st18q04	Possessions cars Q18d	X		
st18q05	Possessions bathrooms Q18e	X		
st21q01	<ISCED 1>years Q21	X	X	
st22q01	Repeat <ISCED 1> Q22a	X		X
st22q02	Repeat <ISCED 2> Q22b	X	X	
st22q03	Repeat <ISCED 3> Q22c	X		
st25q03	Attend specific programme Q25c		X	
st36q01	Students in maths Q36	X	X	
age	Age of student	X	X	
ic01q01	Available at home IC1a		X	
famstruc	Family structure	X		
fisced	Educational level of father (ISCED)	X		
fsecateg	Father white collar/blue collar classification			X
hisced	Highest educational level of parents	X		

[4] μ-ARGUS, User's Manual, CASC-Project, November 2004, page 13.
[5] The ARGUS software can be downloaded at (http://neon.vb.cbs.nl/cenex/default.htm).
[6] μ-ARGUS, User's Manual, CASC-Project, November 2004, pages 62 and 80.

hsecateg	Highest parent white collar/blue collar classification		X	
misced	Educational level of mother (ISCED)	X		X
msecateg	Mother white collar/blue collar classification		X	
natprog_rec	National school programme	X	X	

For an overview of all the variables, in which the ones containing imputed missing values were flagged, see also Appendix 1.

Other modifications

Student, school and class IDs, obtained by concatenation in the original dataset, were replaced by sequential numbers. The new IDs were randomly sorted.

Additional remarks

Unsafe combinations were dealt with by means of global recoding and PRAM (see above). Since most of the identifying variables were categorical, Statistical Disclosure Control techniques for continuous variables were not taken into consideration[7].

Two important identifying variables, namely the variables "Canton" and "Stratum", had been removed from the dataset from the very beginning, before starting the Statistical Disclosure Control process. This was taken into account in setting the thresholds. If the dataset had contained these variables, the thresholds would have been set at a higher level.

---

[7] For more details see: µ-ARGUS, User's Manual, CASC-Project, November 2004.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

APPENDIX 1

Comparative table of original dataset variables and PUS variables

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
1	ic01q01	ic01q01	Available at home IC1a	P
2	ic01q02	ic01q02	Available at school IC1b	
3	ic01q03	ic01q03	Available at other places IC1c	
4	ic02q01	ic02q01	Used computer IC2	
5	ic03q01	ic03q01	How long using computers Q3	
6	ic04q01	ic04q01	Use often at home IC4a	
7	ic04q02	ic04q02	Use often at school IC4b	
8	ic04q03	ic04q03	Use often at other places IC4c	
9		ic05q01	How often information IC5a	
10		ic05q02	How often games IC5b	
11		ic05q03	How often Word IC5c	
12		ic05q04	How often group IC5d	
13		ic05q05	How often spreadsheets IC5e	
14		ic05q06	How often Internet software? IC5f	
15		ic05q07	How often graphics IC5g	
16		ic05q08	How often educ software IC5h	
17		ic05q09	How often learning IC5i	
18		ic05q10	How often download music IC5j	
19		ic05q11	How often programming IC5k	
20		ic05q12	How often chatrooms IC5l	
21		ic06q01	How well start game IC6a	
22		ic06q02	How well antiviruses IC6b	
23		ic06q03	How well open file IC6c	
24		ic06q04	How well edit IC6d	
25		ic06q05	How well scroll IC6e	
26		ic06q06	How well addresses IC6f	
27		ic06q07	How well copy IC6g	
28		ic06q08	How well save IC6h	
29		ic06q09	How well print IC6i	
30		ic06q10	How well delete IC6j	
31		ic06q11	How well move IC6k	
32		ic06q12	How well Internet IC6l	
33		ic06q13	How well download file IC6m	
34		ic06q14	How well attach IC6n	
35		ic06q15	How well program IC6o	
36		ic06q16	How well spreadsheet plot IC6p	
37		ic06q17	How well PowerPoint IC6q	
38		ic06q18	How well games IC6r	

Espace de l'Europe 10
CH-2010 Neuchâtel
www.statistik.admin.ch

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
39		ic06q19	How well download music IC6s	
40		ic06q20	How well multimedia IC6t	
41		ic06q21	How well draw IC6u	
42		ic06q22	How well emails IC6v	
43		ic06q23	How well web page IC6w	
44		ic07q01	Feel important IC7a	
45		ic07q02	Feel fun IC7b	
46		ic07q03	Feel interested IC7c	
47		ic07q04	Feel forget time IC7d	
48	ic08q01	ic08q01	Learn computer IC8	
49	ic09q01	ic09q01	Learn Internet IC9	
50		st01q01	Grade Q1a	
51	st01q02	st01q02	<Programme> Q1b	R, P
52		st02q02	Birth month Q1month	
53		st02q03	Birth year Q1year	
54	st03q01	st03q01	Sex Q3	
55		st04q01	Lives at home Mother Q4a	
56		st04q02	Lives at home Female guard. Q4b	
57		st04q03	Lives at home Father Q4c	
58		st04q04	Lives at home Male guard. Q4d	
59		st04q05	Lives at home Others Q4e	
60	st05q01	st05q01	Mother currently doing Q5	R
61	st06q01	st06q01	Father currently doing Q6	R, P
62		st07q01	Mother's main job Q7	
63		st09q01	Father's main job Q9	
64		st11q01	Mother <ISCED 3A> Q11a	
65		st11q02	Mother <ISCED3B or 3C> Q11b	
66		st11q03	Mother <ISCED2> Q11c	
67		st11q04	Mother <ISCED1> Q11d	
68		st11q05	Mother none of above Q11e	
69		st12q01	Mother <ISCED5A or 6> Q12a	
70		st12q02	Mother <ISCED5B> Q12b	
71		st12q03	Mother <ISCED4> Q12c	
72		st13q01	Father <ISCED 3A> Q13a	
73		st13q02	Father <ISCED3B or 3C> Q13b	
74		st13q03	Father <ISCED2> Q13c	
75		st13q04	Father <ISCED1> Q13d	
76		st13q05	Father none of above Q13e	
77		st14q01	Father <ISCED 5A or 6> Q14a	
78		st14q02	Father <ISCED 5B> Q14b	
79		st14q03	Father <ISCDED 4> Q14c	
80		st15n01	Country of birth Self -national	
81		st15n02	Country of birth Mother -national	
82		st15n03	Country of birth Father -national	
83	st15q01	st15q01	Country of birth Self Q15a_a	R, P
84	st15q02	st15q02	Country of birth Mother Q15a_b	R
85	st15q03	st15q03	Country of birth Father Q15a_c	R
86	st15q04	st15q04	Country of birth Age Q15b	R, P
87		st16n01	Language - national	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
88	st16q01	st16q01	Language at home Q16	R, P
89		st16q02	Additional language spoken at home, yes/no	
90		st16q03	What is second language? Q16c	
91		st17q01	Possessions desk Q17a	
92		st17q02	Possessions own room Q17b	
93		st17q03	Possessions study place Q17c	
94		st17q04	Possessions computer Q17d	
95		st17q05	Possessions software Q17e	
96		st17q06	Possessions Internet Q17f	
97		st17q07	Possessions calculator Q17g	
98		st17q08	Possessions literature Q17h	
99		st17q09	Possessions poetry Q17i	
100		st17q10	Possessions art Q17j	
101		st17q11	Possessions textbooks Q17k	
102		st17q12	Possessions dictionary Q17l	
103		st17q13	Possessions dishwasher Q17m	
104		st17q14	Possessions music instrument Q17n	
105	st18q01	st18q01	Possessions cell phones Q18a	R
106	st18q02	st18q02	Possessions TV-sets Q18b	R
107	st18q03	st18q03	Possessions computers Q18c	R
108	st18q04	st18q04	Possessions cars Q18d	R
109	st18q05	st18q05	Possessions bathrooms Q18e	R
110	st19q01	st19q01	How many books at home Q19	
111		st20q01	Attend <ISCED 0> Q20	
112	st21q01	st21q01	<ISCED 1>years Q21	R, P
113	st22q01	st22q01	Repeat <ISCED 1> Q22a	R, LS
114	st22q02	st22q02	Repeat <ISCED 2> Q22b	R, P
115	st22q03	st22q03	Repeat <ISCED 3> Q22c	R
116		st23n02	Expect <ISCED 3B> Q23b	
117		st23n03	Expect <ISCED 3C> Q23c	
118		st23q01	Expect <ISCED 2> Q23a	
119		st23q04	Expect <ISCED 3A> Q23d	
120		st23q05	Expect <ISCED 4> Q23e	
121		st23q06	Expect <ISCED 5B> Q23f	
122		st23q07	Expect <ISCED 5A or 6> Q23g	
123		st24q01	School done little Q24a	
124		st24q02	School waste of time Q24b	
125		st24q03	School given confidence Q24c	
126		st24q04	School useful Q24d	
127	st25q01	st25q01	Attend local Q25a	
128	st25q02	st25q02	Attend better Q25b	
129	st25q03	st25q03	Attend specific program Q25c	P
130		st25q04	Attend religious Q25d	
131	st25q05	st25q05	Attend family Q25e	
132	st25q06	st25q06	Attend other Q25f	
133		st26q01	Well with students Q26a	
134		st26q02	Interested in students Q26b	
135		st26q03	Listen to me Q26c	
136		st26q04	Give extra help Q26d	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
137		st26q05	Treat me fairly Q26e	
138		st27q01	Feel an outsider Q27a	
139		st27q02	Make friends Q27b	
140		st27q03	Feel I belong Q27c	
141		st27q04	Feel awkward Q27d	
142		st27q05	Think I'm liked Q27e	
143		st27q06	Feel lonely Q27f	
144		st28q01	Late for school Q28	
145		st29q01	Hours all homework Q29a	
146		st29q02	Hours all remedial Q29b	
147		st29q03	Hours all enrichment Q29c	
148		st29q04	Hours all tutor Q29d	
149		st29q05	Hours all out-of-school Q29e	
150		st29q06	Hours all other study Q29f	
151		st30q01	Attitude enjoy reading Q30a	
152		st30q02	Attitude effort Q30b	
153		st30q03	Attitude look forward Q30c	
154		st30q04	Attitude enjoy maths Q30d	
155		st30q05	Attitude career Q30e	
156		st30q06	Attitude interested Q30f	
157		st30q07	Attitude further study Q30g	
158		st30q08	Attitude job Q30h	
159		st31q01	Confident timetable Q31a	
160		st31q02	Confident discount Q31b	
161		st31q03	Confident area Q31c	
162		st31q04	Confident graphs Q31d	
163		st31q05	Confident linear Q31e	
164		st31q06	Confident distance Q31f	
165		st31q07	Confident quadratics Q31g	
166		st31q08	Confident rate Q31h	
167		st32q01	Feel study_worry Q32a	
168		st32q02	Feel study_not good Q32b	
169		st32q03	Feel study_tense Q32c	
170		st32q04	Feel study_good marks Q32d	
171		st32q05	Feel study_nervous Q32e	
172		st32q06	Feel study_quickly Q32f	
173		st32q07	Feel study_best subject Q32g	
174		st32q08	Feel study_helpless Q32h	
175		st32q09	Feel study_underst. diffc. Q32i	
176		st32q10	Feel study_poor marks Q32j	
177		st33q01	Hours maths homework Q33a	
178		st33q02	Hours maths Remedial Q33b	
179		st33q03	Hours maths Enrichment Q33c	
180		st33q04	Hours maths tutor Q33d	
181		st33q05	Hours maths out-of-school Q33e	
182		st33q06	Hours maths other Q33f	
183		st34q01	Learn_important parts Q34a	
184		st34q02	Learn_new ways Q34b	
185		st34q03	Learn_check myself Q34c	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
186		st34q04	Learn_concepts Q34d	
187		st34q05	Learn_everyday life Q34e	
188		st34q06	Learn_solve when sleep Q34f	
189		st34q07	Learn_by heart Q34g	
190		st34q08	Learn_by relating Q34h	
191		st34q09	Learn_examples Q34i	
192		st34q10	Learn_clarify Q34j	
193		st34q11	Learn_applied Q34k	
194		st34q12	Learn_exactly Q34l	
195		st34q13	Learn_procedure Q34m	
196		st34q14	Learn_relate Q34n	
197		st35q01	Minutes in class period Q35a	
198		st35q02	Maths class periods Q35b	
199		st35q03	All class periods Q35c	
200	st36q01	st36q01	Students in maths Q36	R, P
201		st37q01	Attitudes be the best Q37a	
202		st37q02	Attitudes group work Q37b	
203		st37q03	Attitudes exams Q37c	
204		st37q04	Attitudes project Q37d	
205		st37q05	Attitudes effort Q37e	
206		st37q06	Attitudes work with other Q37f	
207		st37q07	Attitudes do better Q37g	
208		st37q08	Attitudes helping Q37h	
209		st37q09	Attitudes learn most Q37i	
210		st37q10	Attitudes best work Q37j	
211		st38q01	Lesson interested Q38a	
212		st38q02	Lesson don't listen Q38b	
213		st38q03	Lesson extra help Q38c	
214		st38q04	Lesson book work Q38d	
215		st38q05	Lesson help learning Q38e	
216		st38q06	Lesson noise Q38f	
217		st38q07	Lesson understand Q38g	
218		st38q08	Lesson quieten down Q38h	
219		st38q09	Lesson can't work well Q38i	
220		st38q10	Lesson opinions Q38j	
221		st38q11	Lesson late start Q38k	
222		stb1q01	Next year's activity	
223		stb1q02	Next year's activity - apprenticeship: job	
224		stb1q03	Next year's activity - apprenticeship: company	
225		stb1q04	Next year's activity - basic vocational training: job	
226		stb1q05	Next year's activity - Other education: education	
227		stb1q06	Next year's activity - Paid Job: job	
228		stb1q07	Next year's activity - something else: what?	
229		stb2q01	Postcode	
230		stb2q02	Locality	
231		stb2q03	Canton	
232	age	age	Age of student	R, P
233	anxmat	anxmat	Mathematics anxiety (WLE)	
234	atschl	atschl	Attitudes towards school (WLE)	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
235	attcomp	attcomp	ICT: attitudes towards computers (WLE)	
236	belong	belong	Sense of belonging to school (WLE)	
237		bfmj	ISCO code Father	
238		bmmj	ISCO code Mother	
239		bookid	Book ID	
240		cant	Canton	
241		claidgt2	Numeric classid, more than 2 students	
242	classid	classid	Numeric classid	
243		classnam	Name of class	
244		classtyp	Cantonal school type	
245		cnt	Country Alphanumeric ISO Code	
246	comphome	comphome	Computer facilities at home (WLE)	
247	complrn	complrn	Competitive learning (WLE)	
248	cooplrn	cooplrn	Co-operative learning (WLE)	
249		country	Country ID	
250	cstrat	cstrat	Control strategies (WLE)	
251	cultposs	cultposs	Cultural possessions of the family (WLE)	
252		diff_lekt	Q35c-Q35b Lektionen insges minus mathem Q35c-Q35b Total lessons minus maths	
253	disclim	disclim	Disciplinary climate in maths lessons (WLE)	
254	elab	elab	Elaboration strategies (WLE)	
255	famstruc	famstruc	Family structure	R
256	fisced	fisced	Educational level of father (ISCED)	R
257	fsecateg	fsecateg	Father white collar/blue collar classification	LS
258		grade	Grade compared to modal grade in country	
259	hedres	hedres	Home educational resources (WLE)	
260	highconf	highconf	ICT: confidence in high-level tasks (WLE)	
261	hisced	hisced	Highest educational level of parents	R
262		hisei	Highest parental occupational status	
263	hsecateg	hsecateg	Highest parent white collar/blue collar classification	P
264	immig	immig	Country of birth	
265	instmot	instmot	Instrumental motivation in mathematics (WLE)	
266	intconf	intconf	ICT: confidence in internet tasks (WLE)	
267	intmat	intmat	Interest in mathematics (WLE)	
268	intuse	intuse	ICT: Internet/entertainment use (WLE)	
269		iscedd	ISCED designation	
270		iscedl	ISCED level	
271		iscedo	ISCED orientation	
272		iso_f	ISO code country of birth Father	
273		iso_m	ISO code country of birth Mother	
274		iso_s	ISO code country of birth Student	
275		lang	Language at home, national	
276		langch	Language region CHE only	
277	langreg	langreg	Language region	
278	matheff	matheff	Mathematics self-efficacy (WLE)	
279	memor	memor	Memorisation strategies (WLE)	
280	misced	misced	Educational level of mother (ISCED)	R, LS
281	mmins	mmins	Minutes of maths per week	
282	msecateg	msecateg	Mother white collar/blue collar classification	P
283	natprog_rec	natprog	National school programme	R, P

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
284		numclass		
285	pcmath	pcmath	Ratio of maths	
286	prguse	prguse	ICT: programs/software use (WLE)	
287		progcont	Continuous school type	
288		progn	Unique national programme code	
289	pv1math	pv1math	Plausible value in math	
290	pv1math1	pv1math1	Plausible value in math - Space and Shape	
291	pv1math2	pv1math2	Plausible value in math- Change and Relationships	
292	pv1math3	pv1math3	Plausible value in math - Uncertainty	
293	pv1math4	pv1math4	Plausible value in math - Quantity	
294	pv1mlev	pv1mlev	Proficiency scale level PV1math	
295	pv1mlev1	pv1mlev1	Proficiency scale level PV1math1 space and shape	
296	pv1mlev2	pv1mlev2	Proficiency scale level PV1math2 change and relationships	
297	pv1mlev3	pv1mlev3	Proficiency scale level PV1math3 uncertainty	
298	pv1mlev4	pv1mlev4	Proficiency scale level PV1math3 quantity	
299	pv1plev	pv1plev	Proficiency scale level PV1prob	
300	pv1prob	pv1prob	Plausible value in problem solving	
301	pv1read	pv1read	Plausible value in reading	
302	pv1rlev	pv1rlev	Proficiency scale level pv1read	
303	pv1scie	pv1scie	Plausible value in science	
304	pv2math	pv2math	Plausible value in math	
305	pv2math1	pv2math1	Plausible value in math - Space and Shape	
306	pv2math2	pv2math2	Plausible value in math- Change and Relationships	
307	pv2math3	pv2math3	Plausible value in math - Uncertainty	
308	pv2math4	pv2math4	Plausible value in math - Quantity	
309	pv2mlev	pv2mlev	Proficiency scale level PV2math	
310	pv2mlev1	pv2mlev1	Proficiency scale level PV2math1 space and shape	
311	pv2mlev2	pv2mlev2	Proficiency scale level PV2math2 change and relationships	
312	pv2mlev3	pv2mlev3	Proficiency scale level PV2math3 uncertainty	
313	pv2mlev4	pv2mlev4	Proficiency scale level PV2math3 quantity	
314	pv2plev	pv2plev	Proficiency scale level PV2prob	
315	pv2prob	pv2prob	Plausible value in problem solving	
316	pv2read	pv2read	Plausible value in reading	
317	pv2rlev	pv2rlev	Proficiency scale level pv2read	
318	pv2scie	pv2scie	Plausible value in science	
319	pv3math	pv3math	Plausible value in math	
320	pv3math1	pv3math1	Plausible value in math - Space and Shape	
321	pv3math2	pv3math2	Plausible value in math- Change and Relationships	
322	pv3math3	pv3math3	Plausible value in math - Uncertainty	
323	pv3math4	pv3math4	Plausible value in math - Quantity	
324	pv3mlev	pv3mlev	Proficiency scale level PV3math	
325	pv3mlev1	pv3mlev1	Proficiency scale level PV3math1 space and shape	
326	pv3mlev2	pv3mlev2	Proficiency scale level PV3math2 change and relationships	
327	pv3mlev3	pv3mlev3	Proficiency scale level PV3math3 uncertainty	
328	pv3mlev4	pv3mlev4	Proficiency scale level PV3math3 quantity	
329	pv3plev	pv3plev	Proficiency scale level PV3prob	
330	pv3prob	pv3prob	Plausible value in problem solving	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
331	pv3read	pv3read	Plausible value in reading	
332	pv3rlev	pv3rlev	Proficiency scale level pv3read	
333	pv3scie	pv3scie	Plausible value in science	
334	pv4math	pv4math	Plausible value in math	
335	pv4math1	pv4math1	Plausible value in math - Space and Shape	
336	pv4math2	pv4math2	Plausible value in math- Change and Relationships	
337	pv4math3	pv4math3	Plausible value in math - Uncertainty	
338	pv4math4	pv4math4	Plausible value in math - Quantity	
339	pv4mlev	pv4mlev	Proficiency scale level PV4math	
340	pv4mlev1	pv4mlev1	Proficiency scale level PV4math1 space and shape	
341	pv4mlev2	pv4mlev2	Proficiency scale level PV4math2 change and relationships	
342	pv4mlev3	pv4mlev3	Proficiency scale level PV4math3 uncertainty	
343	pv4mlev4	pv4mlev4	Proficiency scale level PV4math3 quantity	
344	pv4plev	pv4plev	Proficiency scale level PV4prob	
345	pv4prob	pv4prob	Plausible value in problem solving	
346	pv4read	pv4read	Plausible value in reading	
347	pv4rlev	pv4rlev	Proficiency scale level pv4read	
348	pv4scie	pv4scie	Plausible value in science	
349	pv5math	pv5math	Plausible value in math	
350	pv5math1	pv5math1	Plausible value in math - Space and Shape	
351	pv5math2	pv5math2	Plausible value in math- Change and Relationships	
352	pv5math3	pv5math3	Plausible value in math - Uncertainty	
353	pv5math4	pv5math4	Plausible value in math - Quantity	
354	pv5mlev	pv5mlev	Proficiency scale level PV5math	
355	pv5mlev1	pv5mlev1	Proficiency scale level PV5math1 space and shape	
356	pv5mlev2	pv5mlev2	Proficiency scale level PV5math2 change and relationships	
357	pv5mlev3	pv5mlev3	Proficiency scale level PV5math3 uncertainty	
358	pv5mlev4	pv5mlev4	Proficiency scale level PV5math3 quantity	
359	pv5plev	pv5plev	Proficiency scale level PV5prob	
360	pv5prob	pv5prob	Plausible value in problem solving	
361	pv5read	pv5read	Plausible value in reading	
362	pv5rlev	pv5rlev	Proficiency scale level pv5read	
363	pv5scie	pv5scie	Plausible value in science	
364	rmhmwk	rmhmwk	Relative time spent on maths homework	
365	routconf	routconf	ICT: confidence in routine tasks (WLE)	
366	schoolid	schoolid	School ID	
367	scmat	scmat	Mathematics self-concept (WLE)	
368	ses	ses	SES aus hisei,hisced,cultposs,hedres,missing imputiert falls 2 gültige Werte Imputed, if there were two valid values	
369		sisced	Expected educational level of student (ISCED)	
370		sprog2	Type TI	
371		sprog51		
372		sprog52		
373		sprog53		
374		sprog54		
375		stidsch	School ID	
376		stidstd	Student ID	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
377		stidstrt	Stratum ID	
378	studid	studid	STUDID	
379	sturel	sturel	Student-teacher relations at school (WLE)	
380		subnat	Sub-nation code	
381	teachsup	teachsup	Teacher support in maths lessons (WLE)	
382	tmins	tmins	Total minutes of instructional time p/w	
383		wlemath	Warm estimate in math	
384		wlemath1	Warm estimate in math - Space and Shape	
385		wlemath2	Warm estimate in math - Change and Relationships	
386		wlemath3	Warm estimate in math - Uncertainty	
387		wlemath4	Warm estimate in math - Quantity	
388		wleprob	Warm estimate in problem solving	
389		wleread	Warm estimate in reading	
390		wlerr_m	Estimation error for wlemath	
391		wlerr_m1	Estimation error for wlemath1	
392		wlerr_m2	Estimation error for wlemath2	
393		wlerr_m3	Estimation error for wlemath3	
394		wlerr_m4	Estimation error for wlemath4	
395		wlerr_p	Estimation error for wleprob	
396		wlerr_r	Estimation error for wleread	
397		wlerr_s	Estimation error for wlescie	
398		wlescie	Warm estimate in science	
399	clcuse3a	clcuse3a	How much effort was invested in the test	
400	clcuse3b	clcuse3b	How much effort would have been invested if marks were counted by school	
401	wp00	wp00	Final student weight	
402	wp01	wp01	BRR-Replicate	
403	wp02	wp02	BRR-Replicate	
404	wp03	wp03	BRR-Replicate	
405	wp04	wp04	BRR-Replicate	
406	wp05	wp05	BRR-Replicate	
407	wp06	wp06	BRR-Replicate	
408	wp07	wp07	BRR-Replicate	
409	wp08	wp08	BRR-Replicate	
410	wp09	wp09	BRR-Replicate	
411	wp10	wp10	BRR-Replicate	
412	wp11	wp11	BRR-Replicate	
413	wp12	wp12	BRR-Replicate	
414	wp13	wp13	BRR-Replicate	
415	wp14	wp14	BRR-Replicate	
416	wp15	wp15	BRR-Replicate	
417	wp16	wp16	BRR-Replicate	
418	wp17	wp17	BRR-Replicate	
419	wp18	wp18	BRR-Replicate	
420	wp19	wp19	BRR-Replicate	
421	wp20	wp20	BRR-Replicate	
422	wp21	wp21	BRR-Replicate	
423	wp22	wp22	BRR-Replicate	
424	wp23	wp23	BRR-Replicate	
425	wp24	wp24	BRR-Replicate	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
426	wp25	wp25	BRR-Replicate	
427	wp26	wp26	BRR-Replicate	
428	wp27	wp27	BRR-Replicate	
429	wp28	wp28	BRR-Replicate	
430	wp29	wp29	BRR-Replicate	
431	wp30	wp30	BRR-Replicate	
432	wp31	wp31	BRR-Replicate	
433	wp32	wp32	BRR-Replicate	
434	wp33	wp33	BRR-Replicate	
435	wp34	wp34	BRR-Replicate	
436	wp35	wp35	BRR-Replicate	
437	wp36	wp36	BRR-Replicate	
438	wp37	wp37	BRR-Replicate	
439	wp38	wp38	BRR-Replicate	
440	wp39	wp39	BRR-Replicate	
441	wp40	wp40	BRR-Replicate	
442	wp41	wp41	BRR-Replicate	
443	wp42	wp42	BRR-Replicate	
444	wp43	wp43	BRR-Replicate	
445	wp44	wp44	BRR-Replicate	
446	wp45	wp45	BRR-Replicate	
447	wp46	wp46	BRR-Replicate	
448	wp47	wp47	BRR-Replicate	
449	wp48	wp48	BRR-Replicate	
450	wp49	wp49	BRR-Replicate	
451	wp50	wp50	BRR-Replicate	
452	wp51	wp51	BRR-Replicate	
453	wp52	wp52	BRR-Replicate	
454	wp53	wp53	BRR-Replicate	
455	wp54	wp54	BRR-Replicate	
456	wp55	wp55	BRR-Replicate	
457	wp56	wp56	BRR-Replicate	
458	wp57	wp57	BRR-Replicate	
459	wp58	wp58	BRR-Replicate	
460	wp59	wp59	BRR-Replicate	
461	wp60	wp60	BRR-Replicate	
462	wp61	wp61	BRR-Replicate	
463	wp62	wp62	BRR-Replicate	
464	wp63	wp63	BRR-Replicate	
465	wp64	wp64	BRR-Replicate	
466	wp65	wp65	BRR-Replicate	
467	wp66	wp66	BRR-Replicate	
468	wp67	wp67	BRR-Replicate	
469	wp68	wp68	BRR-Replicate	
470	wp69	wp69	BRR-Replicate	
471	wp70	wp70	BRR-Replicate	
472	wp71	wp71	BRR-Replicate	
473	wp72	wp72	BRR-Replicate	
474	wp73	wp73	BRR-Replicate	

	PUS	Original	Label	Statistical Disclosure Control R=recoded P: Prammed LS: Local suppression
475	wp74	wp74	BRR-Replicate	
476	wp75	wp75	BRR-Replicate	
477	wp76	wp76	BRR-Replicate	
478	wp77	wp77	BRR-Replicate	
479	wp78	wp78	BRR-Replicate	
480	wp79	wp79	BRR-Replicate	
481	wp80	wp80	BRR-Replicate	

APPENDIX 2

Recoded variables

	Variable name	# categories old	# categories new	Notes
1	st01q02	8	2	Of the eight starting categories (questionnaire), only four contained scores. The others were empty. In the PUS, the four categories were collapsed into two.
2	st05q01	4	3	
3	st06q01	4	3	
4	st15q01	10	2	The Swiss version of this variable contained ten categories (national option), in order to collect more information on the country of birth. In the PUS, the categories were collapsed into two (test country/other), as in the international version.
5	st15q02	10	2	The Swiss version of this variable contained ten categories (national option), in order to collect more information on the country of birth. In the PUS, the categories were collapsed into two (test country/other), as in the international version.
6	st15q03	10	2	The Swiss version of this variable contained ten categories (national option), in order to collect more information on the country of birth. In the PUS, the categories were collapsed into two (test country/other), as in the international version.
7	st15q04	No categories: age between 0 and 16 reported with a half-year detail (i.e. 15, 15.5 etc.)	2	Age collapsed into two categories, regarded as meaningful.
8	st16q01	13	2	The Swiss version of this variable contained 13 categories (national option), in order to collect more information on the language spoken at home. The international version of the same variable contained four categories, which were collapsed into two in the PUS.
	Variable name	**# categories old**	**# categories new**	**Notes**

9	st18q01	4	2	
10	st18q02	4	2	
11	st18q03	4	3	
12	st18q04	4	3	
13	st18q05	4	2	
14	st21q01	No categories	2	The answers (only number possible) were collapsed into three categories, regarded as meaningful
15	st22q01	3	2	
16	st22q02	3	2	
17	st22q03	3	2	
18	st36q01	No categories, answers have generated 37 different entries	3	The answers (only number possible) were collapsed into three categories, regarded as meaningful
19	age	No categories, age was recorded precisely.		The answers (only number possible) were collapsed into three categories, regarded as meaningful
20	famstruc	4	3	index
21	fisced	6	4	index
22	hisced	6	4	index
23	misced	6	4	index
24	natprog_rec	7	4	This variable is a national option

Note: the numbers of categories given does NOT include the categories "Missing", "Invalid" and "N/A". For technical reasons related to the SW ARGUS, the categories, "Invalid" and "N/A" were collapsed into one.