Centre for Longitudinal Studies

BCS70 Data Note 1

Enhancing the BCS70 16-year Head Teacher school level dataset

Vania Gerova

Centre for Longitudinal Studies
Bedford Group for Lifecourse and Statistical Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6860
Fax: 020 7612 6880
Email cls@ioe.ac.uk
Web http://www.cls.ioe.ac.uk

Editor’s note:

This Data Note is presented here without its Appendix, which is currently being reviewed.

The Appendix should appear within the next week, containing the SPSS syntax used to derive the variables described in this document.

Enhancing the BCS70 16-year Head Teacher school level dataset

Vania Gerova

Information on the schools attended by the cohort members was previously only available at age 10 for BCS70, since the dataset collected from schools at 16 (Document M: the Headteachers' questionnaire) was incomplete and had been lost. The ESRC funded project 'Single-sex and Co-educational Schooling: Life-course Consequences?' has cleaned the 1986 schools data, and united it with the BCS70 dataset. This technical note provides information on the data cleaning and matching, and on the reliability of the resulting dataset.

Initial steps in adding missing information on sex mix of school and type of school in BCS70 Head Teacher school level dataset

1. Sex mix of school

Using the original BCS70 Head Teacher school level dataset, a syntax was applied defining single sex (boys and girls) and mixed schools.

We generated a variable - *"schsex"* (Table 1) which was based on the proportion of boys and girls on the school register (using htq4b_1 and htq4b_2 variables). The decision was made to code those schools which have less than 5% boys, as girls schools (and those with less than 5% girls, as boys schools; and the rest as mixed schools), as some single-sex schools may have allowed a small number of members of the opposite sex.

Table 1
Schsex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 girls	164	8.2	9.9	9.9
	2 boys	148	7.4	9.0	18.9
	3 mixed	1341	67.1	81.1	100.0
	Total	1653	82.7	100.0	
Missing	System	347	17.4		
Total		2000	100.0		

Source: Deposited BCS70 HT school data

We also looked at the proportion of boys and girls in 5th forms (htq4d_1, htq4d_2), as this seemed a more reliable indicator of single-sex status since it did not include the 6th form. Using the same cut off point as above a variable named *"sex5th"* (Table 2) was generated.

Table 2
sex5th

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 girls	177	8.9	11.8	11.8
	2 boys	162	8.1	10.8	22.5
	3 mixed	1165	58.3	77.5	100.0
	Total	1504	75.2	100.0	
Missing	System	496	24.8		
Total		2000	100.0		

Source: Deposited BCS70 HT school data

Comparing the figures for *"schsex"* and *"sex5th",* 5th forms had higher levels of single sex – 177 (8.9%) girls schools and 162 (8.1%) boys schools (total n 1504). For whole schools, girls schools were 164 (8.2%) and boys 148 (7.4%) of the total (n 1653). This confirmed the suspicion that whole school figures for single-sex would be deflated by mixed 6th forms, so the decision was made to use the 5th form figures. The problem was that there were more missing cases (496) in terms of sex mix of school when using *'sex5th'* variable compared to 347 missing in *'schsex'* variable. We also note that the cross-tabulation of "*schsex"* and "*sex 5th"* revealed 7 anomalous cases; 4 were mixed according to "*sex5th"* but were girls' schools according to "*schsex",* and 3 were girls' schools according to "sex5th" but boys' schools according to "*schsex".* We assume that the discrepancies between the two variables are due to errors, which we have not attempted to correct. In our work we have assumed that the "*sex5th"* variable is more reliable than the "*schsex"* variable. Other users may wish to treat these variables differently.

Another decision was made to use some of the information on sex mix of school from the *'schsex'* variable and to fill the missing information in *'sex5th'* variable. So we mapped girls and boys schools from *'schsex'* to *'sex5th'* and recovered information for 24 schools. We did the same for mixed schools and recovered information for an additional 176 schools. The resulting new variable was *'sex5th_n'* (Table 3). This left us with 296 cases with missing sex mix of school data.

Table 3
sex5th_n

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 girls	189	9.5	11.1	11.1
	2 boys	174	8.7	10.2	21.3
	3 mixed	1341	67.1	78.7	100.0
	Total	1704	85.2	100.0	
Missing	System	296	14.8		
Total		2000	100.0		

Source: Internally modified BCS70 HT school data

2. Type of School

The variable school type *('stype')* had only 4 missing cases, but the category 'other' within it contained 172 schools (Table 4).

Table 4
Stype

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Comprehensive	1097	54.9	55.0	55.0
	2 Independent School	157	7.9	7.9	62.8
	3 Grammar School	64	3.2	3.2	66.0
	4 Special School LEA	396	19.8	19.8	85.9
	5 Secondary Modern School	76	3.8	3.8	89.7
	6 Special School - Independent	31	1.6	1.6	91.2
	7 Technical School	3	.2	.2	91.4
	8 Other School	172	8.6	8.6	100.0
	Total	1996	99.8	100.0	
Missing	System	4	.2		
Total		2000	100.0		

Source: Deposited BCS70 school data

An attempt was made to distribute schools from 'other' category in 'stype' variable to comprehensive, special, modern, etc. 'type of school' categories. To achieve this we used *'othstype'* (other school type) and *'sstype'* (special school type) variables. These are string variables and the text gives us slightly more information on the type of school. From these we recovered information on type of school for 83 cases and added this information to comprehensive, special LEA, grammar and secondary modern categories within *'stype'* variable. The new improved variable was called *'stype_n'* (Table 5).

Table 5
stype_n

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Comprehensive	1150	57.5	57.6	57.6
	2 Independent School	157	7.9	7.9	65.5
	3 Grammar School	81	4.1	4.1	69.5
	4 Special School LEA	406	20.3	20.3	89.9
	5 Secondary Modern	79	4.0	4.0	93.8
	6 Special School - Independent	31	1.6	1.6	95.4
	7 Technical School	3	.2	.2	95.5
	8 Other School	89	4.5	4.5	100.0
	Total	1996	99.8	100.0	
Missing	System	4	.2		
Total		2000	100.0		

Source: Internally modified BCS70 HT school data

We put together grammar and technical schools and created a variable called *'schooltype'* (Table 6).

Table 6
Schooltype

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Comprehensive	1150	57.5	57.6	57.6
	2 Independent	157	7.9	7.9	65.5
	3 Grammar and Technical	84	4.2	4.2	69.7
	4 Special LEA	406	20.3	20.3	90.0
	5 Sec Mod	79	4.0	4.0	94.0
	6 Special Independent	31	1.6	1.6	95.5
	7 Other Type	89	4.5	4.5	100.0
	Total	1996	99.8	100.0	
Missing	System	4	.2		
Total		2000	100.0		

Source: Internally modified BCS70 HT school data

We still had 89 schools in the 'other type' category (see Table 6). We decided to fill this gap from the 1986 School Census data using variable *'e_type'* from the Secondary Schools Census, because 85 out of 89 'other type' schools in the *'schooltype'* variable in TQ school data were secondary schools (for the remaining 4 'other' there was no information).

Matching 1986 School Census data and Register of Educational Establishments (REE) to BCS70 HT school data.

To be able to fill the missing information for the sex mix of school and type of school, we matched the 1986 School Census Data and the Register of Educational Establishments (created between 1996-2000) to our BCS70 HT school level dataset. REE contains information on name, sex and type of schools, while the School Census data has information on sex and type of school. The 1986 School Census Data consists of separate files for:

- Independent schools
- Primary, Middle and Secondary schools
- Special Schools

To be able to match the 1986 School Census, REE and BCS70 HT school data we needed unique identifiers. The identifiers that uniquely define a school are LEA and ESTAB (unique establishment number within LEAs) in the School Census; LEA and DFE (DFE numbers are equivalent to ESTAB) numbers in REE. The establishment number in BCS70 HT school data is ID04 variable (ID04=ESTAB) and new 1998 LEA numbers were additionally added to each school in the HT school data (code/data available within CLS).

We had to overcome several problems with these identifiers. These were:

- REE was created between 1996 and 2000. Since the start of the REE database, many establishments' LEA numbers appear to have changed, as a result of creation, abolition and merger of LEAs in the course of local government reorganisations. Changes to LEA Numbers and DFE Numbers are recorded in the New_LEA_Number, Old_LEA_Number, New_DFE_Number and Old_DFE_Number variables in REE.
- Schools that have closed before the changes in LEAs are kept on REE with their old LEA numbers. Unfortunately for these schools we have no information on gender.
- For the schools that existed at the time of REE creation, we have all the information, including on gender of schools. The problems with these existing schools were the following: a) some of these schools have changed their LEAs from old to new, while the 1986 School Census data is using old LEA numbers, and b) some schools have changed both their LEA and DFE (ESTAB) numbers.

Unfortunately all these changes of identifiers over time did not allow for a straightforward matching of the datasets. [1]

The new file (available within CLS) containing the new 1998 LEA numbers (used also by REE) was matched to BCS70 HT school dataset using ID01 variable. Unfortunately, once matched to BCS70 HT school data, we ended with 838 missing new 1998 LEA numbers. So for all the schools in HT school dataset, for which we did not have the new LEA numbers but had ESTAB number and for which we did not have information on sex mix of school, we tried to impute the missing information on LEA manually. (As mentioned above we needed both LEA and ESTAB numbers to match all datasets.)

We manually added information on LEA and ESTAB variables (and also added the URN variable from REE, which is a unique for each establishment 6 digit reference number, to identify schools) by consulting the 1986 School Census and the REE. We were able to do that as REE had the names of the schools as well as our BCS70 HT school data. In the process we also needed to recode some of the new 1998 LEAs in BCS70 HT school dataset into the old LEAs for the purposes of matching information from the 1986 School Census data.

[1] *Please see http://ndad.ulcc.ac.uk/CRDA/47/detail.html for more information*

Once we filled the gaps with the missing LEAs, we were able to add a few selected variables, including the sex of establishment and type of establishment from the 1986 Census of Secondary, Independent and Special schools files to the BCS70 HT school data. At the end of the matching, a new file was created and all the sex of establishment variables from the different sources were collapsed into one sex mix of school variable. As a result of this we added information on sex mix of school for an additional 240 schools.

We still had 56 schools without sex of establishment information, which also did not have ESTAB or LEA numbers. We proceeded to manually input the missing information by using *Education Year Book 1986*[2]. At the end we were left with 36 schools (mainly in Scotland) for which there was no information on sex mix of school *"sexsch"* (Table 7).

Table 7
sexsch Sex mix of school

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 boys	182	9.1	9.3	9.3
	2 girls	197	9.9	10.0	19.3
	3 mixed	1585	79.3	80.7	100.0
	Total	1964	98.2	100.0	
Missing	System	36	1.8		
Total		2000	100.0		

Source: Externally modified BCS70 HT school data with added information from REE, 1986 School Census and 1986 Education Year Book

We also ended up with a better distribution in Type of School - *"schtype*" variables (Table 8) and School Sector - "*ssector_n"* (derived from type of school) (Table 9), by copying information from the 1986 School Census data (mainly from Secondary Schools as mentioned at the beginning of the paper).

Table 8
schtype School Type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Comprehensive	1215	60.8	60.8	60.8
	2 Independent	157	7.9	7.9	68.7
	3 Grammar and Technical	84	4.2	4.2	72.9
	4 Special LEA	406	20.3	20.3	93.2
	5 Sec Mod	83	4.2	4.2	97.3
	6 Special Independent	31	1.6	1.6	98.9
	7 Other School Type	22	1.1	1.1	100.0
	Total	1998	99.9	100.0	
Missing	System	2	.1		
Total		2000	100.0		

Source: Externally modified BCS70 HT school data with added information from 1986 School Census

[2] London: Longman, for the Association of Education Committees

Table 9
ssector_n School Sector

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 LEA	1679	84.0	84.0	84.0
	2 Independent	157	7.9	7.9	91.9
	3 Special Independent	33	1.7	1.7	93.5
	4 Voluntary	112	5.6	5.6	99.1
	5 Other Sector	17	.9	.9	100.0
	Total	1998	99.9	100.0	
Missing	System	2	.1		
Total		2000	100.0		

Source: Externally modified BCS70 HT school data with added information from 1986 School Census

Finally it is worth mentioning that some of the 2000 schools in the BCS70 HT school data were not attended by identifiable BCS70 cohort members. This is an anomaly, since the target sample for the Headteacher's questionnaire was all schools with a cohort member attending. After adding information on the sex of the schools, we ended up with 1964 (out of a total of 2000) schools for which we had information on the sex mix of school. In 1351 out of these 1964 schools we had cohort members and information on sex mix of school. We had 4514 (out of a total of 4592) cohort members distributed within these 1351 schools, which is on average 3.3 cohort members per school.

Reliability

There was a teachers' strike when the age 16 survey was being carried out in 1986. Therefore, this dataset only covers about one third of the cohort. However, the proportions of students attending schools within each sector is broadly representative, albeit with a slight under-representation of Private and Secondary Modern students, as shown by table 10.

Table 10
Representativeness of Data

	BCS70 Britain		BCS70 England		Statistics of Education England*
	Frequency	%	Frequency	%	%
Comprehensive	3590	87	2355	85	83
Grammar	154	4	139	5	3
Secondary Modern and Technical	134	3	95	3	6
Private	166	4	123	4	7
Special	106	3	56	2	2
Total	4150	101	2768	99	101

**Statistics of Education* figures for 15 year olds in secondary schools in 1984, published 1986. 1984 figures are used because equivalent figures for 1985 were not available. These figures are for England only, Equivalent figures for Scotland and Wales were not available.

Centre for Longitudinal Studies
Bedford Group for Lifecourse and Statistical Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6900
Fax: 020 7612 6880
Email cls@cls.ioe.ac.uk
Web http://www.cls.ioe.ac.uk