Appendix 2

Weighting procedures for the longitudinal component of the Scottish Young Peoples Survey

by David Raffe

In Spring 1985 the sample for the Scottish Young Peoples Survey included a 10 per cent sample of young people who had started S4 in 1983/4 or left school during or at the end of the 1983/4 session. For this year-group component of the sample the survey achieved a response rate of 81 per cent (based on the target sample) or 85 per cent (based on presumed contacts). Differential non-response was compensated for by weighting the data to match population figures for the sex and S4 O-grade attainment of the year-group. (The other, overlapping sample of 1983/4 leavers was weighted separately but analogously, using sex and total SCE attainment.)[1]

In the following Spring, 1986, the same year-group was surveyed a second time. Those who had responded to the 1985 questionnaire were sent one of three questionnaire types (with two versions of each); a fourth type, version N, was devised for non-respondents to the earlier data-sweep. Version N achieved a 36 per cent response rate, equivalent to nearly seven per cent of the original year-group target sample. Altogether more than 87 per cent of this target had responded to at least one SYPS questionnaire by 1986. By enhancing this proportion version N may prove invaluable to a longer-term strategy of sampling-frame maintenance. However in the shorter term most analyses of the 1986 survey will probably exclude version N respondents. This is partly because the questionnaire was shorter and collected fewer data than the others used in 1986, and partly because most analyses of the 1986 data are likely to use them in conjunction with data from the earlier data sweep.

Over 81 per cent of respondents to the 1985 survey responded again in 1986. Disappointingly, response was skewed, reinforcing the skew already evident in the 1985 response. Response in 1986 ranged from 72 per cent among those with no O-grade awards at the end of S4 to 91 per cent among those with seven or more O-grades at A-C. As a result the weighting coefficients calculated for 1985 respondents were inadequate for the 1986 respondents. This is clearly illustrated by the first two columns of Table 1, which show estimates of the statuses of the year-group in Spring 1985. The first column is based on 1985 respondents, the second column on 1985 respondents who responded again in 1986; the same (1985) weights are used for both columns. The third and fourth columns show the corresponding unweighted figures. The difference between each pair of columns reflects the bias in attrition between the two data-sweeps. For example, the proportion at school in Spring 1985 is some four percentage points higher among the 1986 respondents, reflecting the higher non-response in 1986 among the earlier leavers.

With the discovery of these patterns, the need for a new set of weights for the

1986 respondents - excluding version N - became evident Two main options were considered The first simply repeated the 1985 algorithm - based on sex and a ten-category S4 attainment variable - for the 1986 respondents The second used three variables sex, a slightly collapsed S4 attainment variable, with six categories, and a variable distinguishing early leavers (by Christmas S5) from later leavers

The year-group was to be followed up a third time, in Autumn 1987, with the possibility of further sweeps Later cohorts were to be followed up at unknown and probably varying numbers of times It was therefore felt desirable to find a weighting principle that could be repeated, using the same population figures, for successive data-sweeps of each year-group "New" weights were therefore calculated for the 1985 respondents, as well as "new" and "old" weights for the 1986 respondents, corresponding to the two options described above The "new" weights however made additional demands on the Scottish Education Department which provided the population figures The old weights used population figures derived from the Scottish Examination Board annual returns, but these did not distinguish early from later leavers To provide population figures for the new weights, the SEB returns had to be supplemented by figures from the Qualified School Leavers returns (QSL), these were used to estimate the distribution of young people in each sex and qualification category, whose totals were known from the SEB data, between early and later leavers

This in turn posed two problems The first was caused by the failure, within the QSL system, to match some S5 leavers with the qualifications they gained in S4 This was corrected by a *pro rata* adjustment, but because matching failures are more likely for later leavers, the distribution between early and later leavers may be least reliable for the unqualified, who are not included in the QSL system, and are therefore treated residually This may be reflected in the weighting coefficients, particularly for males, shown in Table 2, but given the small sample numbers for unqualified stayers these weights do not look excessively wild The second problem was that the QSL data became available only about a year after the relevant members of the year-group left school This meant that the 1986 survey could use QSL figures only for fourth- and fifth-year leavers, with figures for sixth-year leavers estimated from previous surveys This in itself may not have been too serious, sixth-year leavers being a relatively stable group More important was the implication that, were the new weighting scheme to be adopted for all subsequent cohort surveys, the first sweep of each survey would be weighted using QSL population figures where only the fourth-year leavers, a minority, were actually represented in current data

The four sets of weights - "new" and "old" weights for 1985 and 1986 respondents respectively - are shown in Table 2.

Table 3 shows estimates of selected 1985 survey variables, obtained by applying these different weights to the respondents for whom they were designed. Two interesting conclusions emerge.

The first is that the new and old weights for 1985 respondents produce virtually identical estimates. This suggests that response to the 1985 survey did not differ between early and later leavers of the same sex and S4 attainment. The second conclusion is that the new and old weights result in slightly different estimates for 1986 respondents; estimates based on the new weights show greater consistency with the 1985 estimates.

A *post hoc* explanation for these differences is as follows. The first sweep questionnaire, which covers many aspects of compulsory schooling and the transition to post-compulsory schooling or non-school activities, is perceived as equally relevant by stayers and leavers alike. Although stayers are more likely to be higher qualified and female, and therefore more likely to respond, staying is not independently associated with response. By the second sweep, however, those who had already left by the first sweep are somewhat less likely to feel that the questionnaire is relevant; there is, in effect, a lagged effect of staying or leaving by time t on response at time t+1, but not on response at time t. Possibly early leavers feel they have got schooling out of their system at the first sweep so that later sweeps are unnecessary; possibly it just reflects the passage of time and the general drift of young people away from the experiences or institutions with which the questionnaire is perceived to be concerned. It may be relevant that the 1986 status category whose estimate is most enhanced by the use of the new weights is that for the employed. Tables 4 and 5 compare the estimates for selected 1986 survey variables based on new and old weights respectively. The direction of difference tends to be the same as in Table 3, but the magnitude is generally smaller.

As a result the new weighting scheme has been adopted for the 1986 respondents, and it is proposed to use the same procedure to generate weights not only for the 1987 and any subsequent data-sweeps of this cohort, but also for the second and subsequent data-sweeps of future cohorts. However, in view of the practical problems associated with the production of population figures, and the delay in the full data becoming available, the old weighting procedure is retained for 1985 respondents and for the first sweep of each later cohort.

1. See Joanne Lamb, "Scottish Young Peoples Survey 1985: Technical Report", CES, 1986.

Table 1

Alternative estimates of April 1985 status (%)

Weighted by:	"old 1985" weights		unweighted	
Respondents in:	1985	1986 (exc N)	1985	1986 (exc N)
School	41.9	46.0	46.5	50.7
YTS	22.6	21.4	20.8	19.3
Job	17.3	16.2	16.3	15.3
Unemployed	11.1	9.8	9.7	8.5
FT FE	3.6	3.5	3.6	3.4
Others	2.1	1.7	1.8	1.5
NK	1.3	1.3	1.3	1.3
Unweighted n	6501	5292	6501	5292

Note: Figures for 1986 respondents exclude version N: in other words they cover only those who responded in 1985 and 1986.

Table 2

Mean values of different weights

	Old 1985 weights		Old 1986 weights	
	Males	Females	Males	Females
No awards	1.18	1.27	1.40	1.37
D-E only	1.31	1.16	1.36	1.27
1 A-C award	1.22	.98	1.26	1.05
2 A-C awards	1.01	1.06	1.07	1.08
3 A-C awards	.93	.87	.95	.84
4 A-C awards	.83	.88	.87	.84
5 A-C awards	.86	.82	.82	.75
6 A-C awards	.84	.88	.82	.76
7 A-C awards	.84	.85	.76	.75
8+ A-C awards	.84	.88	.79	.77

	New 1985 weights				New 1986 weights			
	Early leavers		Later leavers		Early leavers		Later leavers	
	Males	Females	Males	Females	Males	Females	Males	Females
No awards	1.16	1.30	1.51	1.07	1.38	1.42	1.75	1.02
D-E only	1.38	1.23	1.04	.97	1.45	1.41	1.02	.94
1-2 A-C awards	1.11	1.02	1.11	1.01	1.19	1.09	1.11	1.03
3-4 A-C awards	.84	.80	.90	.95	.90	.77	.90	.90
5-6 A-C awards	.89	.80	.83	.87	.90	.78	.79	.76
7+ A-C awards	.92	1.17	.83	.86	1.14	1.29	.75	.74

Table 3

1985 survey variables: alternative estimates (%)

Respondents in:	1985		1986 (exc N)	
Weighted by:	old 1985	new 1985	old 1986	new 1986
April 1985 status				
School	41.9	41.8	43.1	42.0
YTS	22.6	22.8	22.6	23.1
Job	17.3	17.3	16.9	17.4
Unemployed	11.1	11.1	10.7	10.8
FT FE	3.6	3.6	3.6	3.6
Others	2.1	2.1	1.8	1.8
NK	1.3	1.3	1.3	1.3
October 1984 status				
School	56.4	56.4	57.6	56.7
YTS	18.3	18.3	17.6	17.9
Job	11.3	11.3	10.9	11.3
Unemployed	6.6	6.6	6.5	6.5
FT FE	4.4	4.4	4.5	4.6
Others	1.8	1.8	1.7	1.7
NK	1.2	1.2	1.3	1.3
Term left school (respondent's account)				
Before Xmas 83	.3	.3	.3	.3
Xmas 83	3.7	3.7	3.7	3.8
Spring 84	1.0	1.0	.9	.9
Summer 84	33.9	33.9	33.1	33.9
Xmas 84	14.6	14.6	14.4	14.7
Spring 85	1.7	1.7	1.5	1.5
Summer 85	.5	5	.4	.4
After summer 85	.1	.1	.1	.1
Not left/NK	44.4	44.3	45.6	44.5
Ever started YTS (% yes)	32.7	32.3	31.5	32.1

Table 4

1986 survey variables: alternative estimates (%)

	1986 respondents (exc N)	
	old 1986 weights	new 1986 weights
April 1986 status		
School	20.4	20.0
YTS	7.7	7.4
Job	39.8	40.4
Unemployed	18.5	18.6
FT FE/HE	8.4	8.3
Others	4.9	4.9
NK	.4	.4
Ever started YTS (% yes)	42.3	42.5
Term left school		
Before Xmas 83	.3	.3
Xmas 83	3.8	3.8
Spring 84	.9	.9
Summer 84	33.1	33.9
Xmas 84	14.4	14.8
Spring 85	1.8	1.7
Summer 85	17.9	17.3
Xmas 85	3.7	3.5
Spring 86	.8	.8
Summer 86	.2	.2
Not left/ NK	23.2	22.7

Table 5

1986 survey variables' alternative estimates

Old 1986 weights		1986 respondents (exc N) April 1986 status						
April 1985 status		School	YTS	Job	Unemp	FT FE	Others	NK
School	%	46.6	12.1	19.4	4.1	14.4	3.0	.4
YTS	%	0	1.8	56.3	35.6	2.0	4.0	.3
Job	%	0	1.1	83.7	11.3	1.3	2.5	0
Unemp	%	0	9.4	24.7	52.4	1.6	11.8	.1
FT FE	%	0	17.7	31.5	9.8	32.7	6.8	1.5
Others	%	0	3.6	24.8	29.2	2.9	38.3	1.1
NK	%	22.0	15.7	28.7	20.0	9.0	3.5	1.1
New 1986 weights								
School	%	46.9	11.7	19.5	4.0	14.5	3.0	.5
YTS	%	0	1.8	56.3	35.5	2.0	4.1	.3
Job	%	0	1.1	84.0	11.1	1.3	2.5	0
Unemp	%	0	9.3	24.5	52.6	1.6	11.8	.2
FT FE	%	0	17.5	31.4	9.8	32.8	6.9	1.5
Others	%	0	3 0	25.6	28.6	2.8	38.9	1.1
NK	%	21.7	16.2	28.4	19.9	9.1	3.6	1.1