Lsay longitudinal surveys of Australian youth

Technical Paper No. 15

WEIGHTING THE 1995 YEAR 9 COHORT SAMPLE FOR DIFFERENTIAL RESPONSE RATES AND SAMPLE ATTRITION

Gary N. Marks
Michael Long

July 2000

ACER

Australian Council for Educational Research

WEIGHTING THE 1995 YEAR 9 COHORT SAMPLE FOR DIFFERENTIAL RESPONSE RATES AND SAMPLE ATTRITION

This technical paper describes the weighting procedure developed for analysis of the first five (1995-1999) waves of the 1995 Year 9 sample (Y95) of the *Longitudinal Surveys of Australian Youth* (LSAY) study.

The weighting procedure addresses three issues:

(a) The original sample used a disproportionate stratified design. Students from States and Territories with smaller numbers of Year 9 students were over-sampled and students from States with larger numbers of Year 9 students were under-sampled. Students from government, Catholic and independent schools were to be proportionately selected within States. In order for the sample to more accurately represent the population of Australian Year 9 students, the sample needs to be weighted so that sample sizes within strata are proportionate to the population sizes of the strata. In the first year of the survey (1995) the weighting compensates for the joint effects of differences in sampling fractions among the strata and differences in response rates among the strata. In subsequent years it also compensates for differences in attrition among the strata.

(b) The initial sample consisted of schools selected with probability proportional to size. In order for each student to have an equal probability of selection, an equal number of students should be selected from each school. Since students were randomly selected in intact classes, different numbers of students were selected from each school. Schools must be weighted so that within strata the students from each school represent the same number of students. In the first year of the survey (1995) the school weighting compensates for sampling unequal numbers of students from each school. In subsequent years it also compensates for differences in attrition among the schools.

(c) Students with different characteristics have a differing probabilities of responding to subsequent waves of the survey. Such differing probabilities reflect both the willingness of the respondent to participate in the survey and the availability of the respondent for interview. A post-stratification weighting schema attempts to compensate for such differential attrition by identifying a variable related to attrition and re-weighting the sample so that in any year the weighted sample reflects the original distribution of that variable. Gender and a combined measure of school achievement have been used in the post-stratification weighting schema for the Y95 LSAY cohort.

The Original Sample Design

The original sample is a stratified cluster sample. In the first stage of the sampling process schools were randomly selected and, in the second stage, classes within schools were randomly selected.

The major stratum in the stratified design was State of schooling. In order to obtain estimates with narrow confidence limits for students in the smaller states, the smaller States and Territories were over-sampled and the larger states under-sampled. Selection of students within States was intended to be proportional to size of Sector. Three sectors were used as strata: Government schools, Catholic schools and Non-government, Non-Catholic (referred to as *independent*) schools. The population data for strata were taken from the *Schools Australia* series (ABS). Within strata, schools were to be selected proportional to their size. Information on the number of Year 9 students in each school came from ACER's Sampling Frame which, in turn, was based on information provided by the relevant State authorities and, in the case of non-government schools, by (the then) DEET. These figures were from the 1994 annual school census. Within schools two classes were to be randomly selected (again, proportional to their size). Schools were asked for a list of the number of students enrolled in each of their Year 9 classes for a subject studied by all Year 9 students in the school (usually English classes). The stratification weights correct for the disproportionate sampling between strata (state and sector) and for the variation between strata due to differential response rates and variable class sizes.

Sample Attrition

In the original sample there were 13,613 respondents. Table 1 shows the number of respondents for the 1996 to 1998 waves. After 4 years in the field the number of active sample members declined to 8,783 a loss of 4,830 respondents, 35 per cent of the original sample. This attrition corresponds to an attrition rate of around 9 per cent per year. The greatest attrition occurred between the first and second waves with smaller attrition levels for subsequent waves. Although the year-to-year attrition rate is reasonably small, further efforts should be made to reduce this rate.

As expected sample attrition does not occur uniformly across social groups. Tables 2 to 6 show the weighted and unweighted percentages and number of cases for the years 1995 to 1999 respectively. Focusing on the unweighted results for 1995 and 1999 shows that attrition was slightly higher among males, Aboriginal and Torres Strait Islanders, students attending Government schools, students from a lower socio-economic status and students who performed less well in the achievement tests. Achievement appears to be a stronger (negative) correlate of attrition than other variables.

Comparison of the weighted frequencies for waves 2 to 6 with the equivalent frequencies from the original sample show that the attrition weights are effective in reducing the impact of differential attrition. Since gender and achievement are the strata used in the post-stratification schema, the 1995 and 1999 weighted distributions for these variables are identical. For other variables their distributions are very similar or close to identical. Most categories are within a single percentage of the weighted 1995 distributions. Hence the weighting compensates for the effect of attrition.

Table 1 Numbers of Percentages of Respondents for the 1995 to 1998 Waves of the Y95 LSAY Cohort

Survey Year	Number of Respondents	Per cent of Original Sample
1995	13,613	100
1996	9,837	72
1997	10,307	76
1998	9,738	72
1999	8,783	65

Table 2 Weighted and Unweighted Frequencies for Selected Variables – 1995 (N=13,613)

Variable	Category	Unweighted		Weighted	
		N	%	N	%
Gender	Male	6717	49.3	6653.5	48.9
	Female	6896	50.7	6959.3	51.1
Aboriginal and Torres Strait Is.	ATSI	385	3.0	373.4	2.9
	Not ATSI	12352	97.0	12362.3	97.1
Non-English Speaking Background	NESB	2580	20.0	2721.4	21.1
	Not-NESB	10325	80.0	10179.2	78.9
State	ACT	599	4.4	266.6	2.0
	NSW	3090	22.7	4556.7	33.5
	VIC	2865	21.0	3310.9	24.3
	QLD	2524	18.5	2498.9	18.4
	SA	1720	12.6	1033.4	7.6
	WA	1837	13.5	1438.9	10.6
	TAS	582	4.3	397.6	2.9
	NT	396	2.9	110.0	0.8
School Sector	Government	9081	66.7	9143.0	67.2
	Catholic	2517	18.5	2743.8	20.2
	Independent	2015	14.8	1726.2	12.7
Region	Metropolitan	7564	55.7	7488.6	55.2
	Regional	3378	24.9	3244.3	23.9
	Rural/Remote	2629	19.4	2830.9	20.1
Father's Occupation	Professional Managerial	5307	44.5	5222.4	44.1
	Sales, Clerks	1807	15.2	1731.4	14.6
	Trades & Skilled Manual	2332	19.6	2373.4	20.1
	Plant Operators Labourers	2485	20.8	2508.8	21.2
Achievement	Bottom Quartile	3268	24.1	3378.4	24.9
	Second Quartile	3406	25.1	3433.3	25.3
	Third Quartile	3345	24.6	3333.6	24.6
	Top Quartile	3559	26.2	3430.8	25.3

Table 3 Weighted and Unweighted Frequencies for Selected Variables – 1996 (N=9,837)

Variable	Category	Unweighted		Weighted	
		N	%	N	%
Gender	Male	4475	54.5	4814.4	48.9
	Female	5362	45.5	5022.6	51.1
Aboriginal and	ATSI	201	2.2	219.7	2.4
Torres Strait Is.	Not ATSI	9095	97.8	9048.7	97.6
Non-English Speaking Background	NESB	1819	19.3	1939.3	20.6
	Not-NESB	7615	80.7	7477.2	79.4
State	ACT	456	4.6	200.6	2.0
	NSW	2102	21.4	3109.8	31.6
	VIC	2097	21.3	2449.0	24.9
	QLD	1814	18.4	1859.9	18.9
	SA	1361	13.8	811.2	8.2
	WA	1321	13.4	1037.5	10.5
	TAS	446	4.5	304.6	3.1
	NT	240	2.4	64.3	0.7
School Sector	Government	6373	64.8	6535.3	66.4
	Catholic	1879	19.1	2012.7	20.5
	Independent	1585	16.1	1289.0	13.1
Region	Metropolitan	5442	55.3.	5318.4	54.1
	Regional	2419	24.6	2371.9	24.1
	Rural/Remote	1976	20.1	2146.7	21.8
Father's Occupation	Professional Managerial	4120	46.4	3932.2	45.0
	Sales, Clerks	1308	14.7	1226.4	14.0
	Trades & Skilled Manual	1691	19.1	1763.6	20.2
	Plant Operators, Labourers	1756	19.8	1814.3	20.8
Achievement	Bottom Quartile	1934	19.7	2428.3	24.8
	Second Quartile	2405	24.5	2473.9	25.2
	Third Quartile	2540	25.9	2415.4	24.6
	Top Quartile	2938	29.9	2492.1	25.4

Table 4 Weighted and Unweighted Frequencies for Selected Variables – 1997 (N=10,307)

Variable	Category	Unweighted		Weighted	
		N	%	N	%
Gender	Male	5001	48.5	5037.6	48.9
	Female	5306	51.5	5269.4	51.1
Aboriginal and Torres Strait Is.	ATSI	228	2.3	241.0	2.5
	Not ATSI	9480	97.7	9437.8	97.5
Non-English Speaking Background	NESB	1900	19.3	2024.9	20.6
	Not-NESB	7959	80.7	7805	79.4
State	ACT	437	4.2	193.5	1.9
	NSW	2264	22.0	3315.8	32.2
	VIC	2231	21.6	2578.7	25.0
	QLD	1865	18.1	1893.2	18.4
	SA	1407	13.7	844.5	8.2
	WA	1409	13.7	1120	10.9
	TAS	436	4.2	293.2	2.8
	NT	258	2.5	67.8	0.7
School Sector	Government	6712	65.1	6807.2	66.0
	Catholic	1990	19.3	2144.6	20.8
	Independent	1605	15.6	1355.4	13.1
Region	Metropolitan	5645	54.8	5587.7	54.2
	Regional	2603	25.3	2511.7	24.4
	Rural/Remote	2057	20.0	2204.6	21.4
Father's Occupation	Professional Managerial	4278	46.2	4119	45.0
	Sales, Clerks	1372	14.8	1814.8	19.8
	Trades & Skilled Manual	1769	19.1	1314.8	14.4
	Plant Operators, Labourers	1850	20.0	1906.7	20.8
Achievement	Bottom Quartile	2154	21.0	2542.8	24.8
	Second Quartile	2503	24.3	2578.9	25.1
	Third Quartile	2657	25.8	2542.4	24.8
	Top Quartile	2970	28.9	2608.2	25.4

Table 5 Weighted and Unweighted Frequencies for Selected Variables – 1998 (N=9,738)

Variable	Category	Unweighted		Weighted	
		N	Per Cent	N	Per Cent
Gender	Male	4679	48.0	4808.5	48.9
	Female	5059	52.0	5029.7	51.1
Aboriginal and Torres Strait Is.	ATSI	204	2.2	223.6	2.4
	Not ATSI	8965	97.8	9007.3	97.6
Non-English Speaking Background	NESB	1784	19.1	1923.1	20.5
	Not-NESB	7532	80.9	7460.4	79.5
State	ACT	415	4.3	185.7	1.9
	NSW	2100	21.6	3106.5	31.6
	VIC	2104	21.6	2467.3	25.1
	QLD	1761	18.1	1821.7	18.5
	SA	1349	13.9	818.2	8.3
	WA	1356	13.9	1093.5	11.1
	TAS	412	4.2	280.7	2.9
	NT	241	2.5	64.6	0.7
School Sector	Government	6294	64.6	6474.2	65.8
	Catholic	1885	19.4	2046.5	20.8
	Independent	1559	16.0	1317.4	13.4
Region	Metropolitan	5344	54.9	5410.7	55.4
	Regional	2448	25.1	2351.2	24.1
	Rural / Remote	1945	20.0	2009.5	20.6
Father's Occupation	Professional Managerial	4085	46.5	3796.1	44.0
	Sales, Clerks	1300	14.8	1249.9	14.5
	Trades & Skilled Manual	1653	18.8	1744.8	20.2
	Plant Operators, Labourers	1739	19.8	1840.0	21.3
Achievement	Bottom Quartile	1938	20.0	2503.7	25.7
	Second Quartile	2357	24.3	2465.3	25.3
	Third Quartile	2549	26.2	2368.7	24.3
	Top Quartile	2872	29.6	2400.4	24.7

Table 6 **Weighted and Unweighted Frequencies for Selected Variables – 1999 (N=8,783)**

Variable	Category	Unweighted		Weighted	
		N	Per Cent	N	Per Cent
Gender	Male	4174	47.5	4292.3	48.9
	Female	4609	52.5	4490.7	51.1
Aboriginal and Torres Strait Is.	ATSI	172	2.1	188.4	2.3
	Not ATSI	8098	97.9	8055.4	97.7
Non-English Speaking Background	NESB	6788	80.8	6610.8	78.9
	Not-NESB	1613	19.2	1763.5	21.1
State	ACT	368	4.2	165.0	1.9
	NSW	1877	21.4	2886.9	32.9
	VIC	1874	21.3	2097.5	23.4
	QLD	1621	18.5	1735.9	19.8
	SA	1250	14.2	654.2	7.8
	WA	1224	13.9	913.7	10.4
	TAS	349	4.0	259.6	3.0
	NT	220	2.5	70.2	0.8
School Sector	Government	5698	64.9	5977.3	68.1
	Catholic	1706	19.4	1747.7	19.9
	Independent	1379	15.7	1058	12.0
Region	Metropolitan	4834	55.0	4848.2	55.2
	Regional	2201	25.1	2118.9	24.1
	Rural / Remote	1747	19.9	1814.6	20.7
Father's Occupation	Professional Managerial	3681	46.4	3412.2	43.9
	Sales, Clerks	1184	14.9	1603	14.5
	Trades & Skilled Manual	1522	19.2	1129.3	20.6
	Plant Operators, Labourers	1545	19.5	1129.3	21.0
Achievement	Bottom Quartile	1667	19.0	2179.8	24.9
	Second Quartile	2114	24.1	2215.3	25.3
	Third Quartile	2323	26.5	2151.0	24.6
	Top Quartile	2659	30.3	2213.6	25.3

Appendix: SAS CODE TO PRODUCE WEIGHTS

```
/*
**  wt95_98a.sas
**
**  Code that produces SAMPLE, ATTRITION and OVERALL weights.
**  Presumes that data for the given year (eg 1998) have
**  been merged onto overall data and that the merged data set
**  has a variable (IN98) that indicates whether the
**  respondent was interviewed in the given year.
*/

options nocentre nofmterr;

libname maindata '[yit2.y95.data]';
libname here   '[]';

data a; set maindata.y95merg(keep=
  v1
  state
  scltyp
  v49
  schlno
  in98
  rename=
  (scltyp=sector
   v49=sex
   schlno=school
   )
  );
if in98;
in=1;

* pseudo schools;
school=floor(school/10);

schwt=1;
if school=3304 then schwt=0.5;
if school=4135 then schwt=0.25;
if school=5114 then schwt=0.5;
if school=5128 then schwt=0.5;
if school=5130 then schwt=0.5;
if school=6126 then schwt=0.666667;
if school=7106 then schwt=0.5;
* construct pseudo-school because of sampling;
if school=7119 then school=7117;
if school=7122 then school=7117;

* fix-up;
if school=2163 then schwt=0.333333;

proc freq; tables in98;

proc summary data=a; class state sector; var in;
  output out=b sum=persons;

proc print data=b;
```

```
data b; set b; retain total;
  if _n_=1 then total=persons;
  sample=persons/total;
  if _type_=3;
  keep state sector sample persons;

data _null_; set b; put state 6. sector 6. sample 8.4;

data c; merge here.abs1 b; by state sector;
wt1=popn/sample;
drop popn sample;

proc sort data=a out=a; by state sector school;
* first check;
data x; merge a c; by state sector;
proc univariate; var wt1;

proc summary data=a nway; class state sector school; id schwt;
  output out=d;
data d; set d; rename _freq_=students;

proc summary data=d nway; class state sector; var schwt students;
  output out=e sum=tschwt tstudent;

proc sort data=b out=b; by state sector;
proc sort data=d out=d; by state sector;
proc sort data=e out=e; by state sector;
data e; merge b c d e; by state sector;
   swt=((tstudent/tschwt)/students)*schwt;
   wt98gen=wt1*swt;

data y; merge a e; by state sector school;

proc univariate; var swt wt1 wt98gen;

proc sort data=y out=y; by v1;

data wt95_98; set y; keep v1 wt98gen;

proc freq; tables wt98gen;

proc summary data=y; id wt98gen;
  class school;
  output out=q;

proc print;

************************;
* Compute post-strat sex * ach weights;
************************;
data q; merge
  maindata.y95merg(keep=v1 achq v64 in98 wt95gen)
  wt95_98(keep=v1 wt98gen);
  ;
  by v1;
in=1;
astratum=0;
if achq ne . and v64 ne . then do;
```

```
   astratum = v64*4 + achq;
   end;
if astratum=0 then delete;
keep v1 in in98 astratum wt95gen wt98gen;

proc summary data=q; class astratum; var in;
  output out=old sum=old;
  weight wt95gen;
data old; set old; retain total;
  if _n_=1 then total=old;
  else old=old/total;
data old; set old; if _n_=1 then delete; keep astratum old;

proc summary data=q; class astratum; var in;
  where in98=1;
  output out=new sum=new;
  weight wt98gen;
data new; set new; retain total;
  if _n_=1 then total=new;
  else new=new/total;
data new; set new;
  if _n_=1 then delete; keep astratum new;

data r; merge old new; by astratum;
  achwt=old/new;
  put astratum 3. old 5.1 new 5.1 achwt 7.2;
  keep astratum achwt;

proc sort data=q out=q; by astratum;

data q; set q; if in98=1;

data s; merge q r; by astratum;
if achwt=. then achwt=1;

proc univariate; var achwt;

proc freq; tables astratum;

proc sort data=s out=s; by v1;

data here.wt95_98a; merge wt95_98 s(keep=v1 achwt); by v1;
  if achwt=. then achwt=1;
  wt98=achwt*wt98gen;

label
  achwt   = 'Weight for 98 data for Attrition Only'
  wt98gen = 'Weight for 98 Sample Only'
  wt98    = 'Weight for 98 data for Attrition+Sample'
  ;

keep v1 wt98gen achwt wt98;
rename achwt=achwt98;

proc univariate; var wt98gen achwt98 wt98;
```