Estimating State Totals from the Private School Universe Survey

Easley Hoy, Beverley Causey, Leroy Bailey, Bureau of the Census, Steven Kaufman, NCES
Easley Hoy, Bureau of the Census, Room 3000-4, Washington, DC 20233

Key Words: Coverage Adjustment, Indirect Estimation, Raking

I. Background and History

The Private School Survey (PSS) is designed and conducted by the Census Bureau for the National Center for Education Statistic to collect data for all private secondary schools in 50 states and D.C. Every two years the survey collects data in an attempt to obtain a complete count for all private schools along with counts of students, teachers, and graduates.

The survey collects data from an administrative list of private schools. To improve the coverage of this list frame, additional lists were obtained from private school associations, state records, and other sources. These lists were matched and unduplicated with the list frame. These operations added about 4900 schools for 1991 and about 2300 schools for 1993. Despite these efforts, the private school list frame remains incomplete, with around 8% of private schools missing from the list. The list enumeration estimates are therefore supplemented by a followup area sample that aims to find and represent unlisted private schools. While direct estimation from the followup area sample produces estimates for unlisted schools of adequate precision for the four geographical regions, it fails to do so for individual states. This paper reports the empirical results of an alternative method for providing estimates of such state totals.

II. Current Methodology from the PSS

For this follow up survey, a stratified sample of primary sampling units (PSUs) is drawn with probability proportionate to size (PSS). The PSUs are comprised of counties or groups of counties and the strata cross state boundaries. Eight of the largest counties are included with certainty and about 115 PSUs are selected with noncertainty. In each sample PSU, seven different sources (e.g., yellow pages, local government offices, etc.) were used to identify missing schools which were not on the list frame. For 1991, the search identified a total of 355 missing schools, and for 1993 there were 421 such schools identified.

For direct estimation, each "added school" is first multiplied by its sampling weight (the reciprocal of the PSU's probability of selection). Then the weighted schools are added up to the PSU level, summed over PSUs within a state to obtain state totals, and summed over states to obtain the four region totals of the number of schools, missed by the list frame. Similar estimates can be obtained for students, teachers, and graduates. Such regional totals have adequate precision, but the state totals are dependent upon which of a states' PSUs are selected for the follow up stratified area sample. For the largest states, there should be no problem, as they will surely have at least one sample PSU with which to estimate its private school undercoverage rate. However, for the smaller states, they may not have a sample PSU in the follow up area survey and thus there would be no estimates of uncoverage rate for such states. The direct estimation approach does provide a unbiased estimator of total added schools to an area's total, and the sampling variance is relatively easily estimated.

III. Proposed Methodology

For the indirect estimator, each sample PSU receives the actual number of unweighted added schools from the followup survey. Each nonsample PSU school gets an upward adjustment based on the sample PSUs' estimation of the add rates.

1. Assume for the moment an across the board adjustment. Also suppose that in the area frame sample, the sample PSUs have a total of n schools

For school i let w_i be the weight associated with the school

$y_i = 1$ if school is an add
$= 0$ if school is an original (originally on the list frame)

Let p be the weighted proportion of school adds.

$$p = \left(\sum_{i=1}^{n} w_i y_i \right) / w. \text{ with } w. = \sum_{i=1}^{n} w_i$$

Then q = 1-p is the weighted proportion of original schools. Consequently if 1/q is the ratio of weighted total schools to the number of schools in the original listing. Let r = 1/q be the adjustment factor by which each school outside the sample PSUs is revised upward to reflect undercount. As it turns out, equivalently, we will be estimating for each list school outside the sample PSUs, p/q schools on average to be added. Since

$$r = 1 + \frac{p}{q} = 1 + \frac{p}{1-p} = \frac{1}{q}.$$

2. Knowledge of Postrata

Values of p and r can be quite different for different kinds of schools. NCES has identified 9 groups of such schools:

1. Catholic Parochial
2. Catholic Diocesan
3. Catholic Private
4. Conservative Christian
5. Other Religious, Affiliated
6. Other Religious, Unaffiliated
7. Nonsectarian Regular
8. Nonsectarian Special Emphasis
9. Nonsectarian Special Education

Such groupings helps to distinguish between different "coverage patterns for different schools. Therefore we will develop a value of p within each poststratum j. But also we can use that fact that big schools, (number of students) are easier to find than small schools. So rather than just fit a value p_j for poststratum j we fit a relationship $p_j(x)$ so that a school outside the sample belonging to poststratum j and size x receives an adjustment factor $r_j(x)$

Within the poststratum j we will drop the subscript j and fit a logistical regression.
$p = 1/[1+\exp(A + Bx)]$ subject to the constraints

$$y_i = 0 \qquad \sum w_i R_i = \sum w_i \qquad y=1 \qquad (1)$$

$$y_i = 0 \qquad \sum w_i x_i R_i = \sum w_i x_i \qquad y=1 \qquad (2)$$

Where $R_i = \frac{p_i}{q_i} = \frac{1}{\exp(A + B x_i)}$ and the summation over i on the left hand side of the equation is where $y_i=0$ (original schools) and the summation on the right hand side of the equation is for $y_i=1$ (add schools). We can now solve for coefficients A&B to be used for deriving the adjustment factors. The adjustment for the ith school is $r_i = 1 + R_i$.

Table 1
1993 Coefficients for Linear Adjustment

Poststratum	A	B
1	4.2144	.6137
2	4.5463	.0742
3	1.2842	1.6214
4	2.9289	.2379
5	.9181	1.0255
6	1.6024	.5807
7	1.6024	.5807
8	.9071	.6212
9	1.8757	2.3367

For the estimate of teacher adds, we refit A values with the x variable representing teachers in equation (2). For the estimate of graduate, data were sparse, so we collapsed across poststrata j to one strata.

IV. Results

A. Application Rules of Proposed Methodology for State Totals

The following summarizes the procedure for state estimation in the PSS:

1. A school, its students, teachers and graduates from a sample PSU are included in the adjusted state counts (including adds) but not in original count if it is an add.
2. A shool its students, teachers, and graduates from a non sample PSU are in the original count and with multiplied factor of r in the adjusted count.

For each state, we derived original and adjusted counts of schools, students, teachers, and graduates; also the ratio of students to teachers, and the ratio of adjusted to original counts were computed for each of these categories. In the full paper and the appendix, to be published in a separate document later, the adjustments for all 50 states and D.C. for both 1991 and 1993 will be provided. The range of adjusted values for most states was 8-12% for 1991 and 5-11% for 1993. The 1993 adjustment results for 6 states (California, Indiana, New York, Texas, Vermont, and Wyoming) are provided in Tables A, B, and C at the end of this paper as examples for discussion.

We begin the discussion with Table A. The numbers of Private Schools, Students, Teachers, Graduates, and Student-Teacher Ratio, by States 1993. Note for each state the ratio adjustment for schools is greater than that for students and teachers. For the states of California, New York, and Indiana, the ratio factors are near average. For the states of Texas, Vermont, and Wyoming, the ratio factors are above average.

B. Variance Estimates for Proposed Methodology

For the proposed methodology, there are two components of variance. The first component is the sampling error due to use of sample to estimate parameters A & B of the model. The second component reflects the variability due to the model. It turns out that the second component is many times larger than the first component.

Table B provides the adjusted estimates with standard deviations reflecting both components of variance.

Note that the standard deviations are small relative to the estimates. Also the first component is a smaller proportion to total error.

C. Raking Adjustment to Regional Totals

For precision and consistency sake, we want the

regional totals for the proposed method to equal those for the current method.

Table 2
NCES Totals for Regions

	Northeast	Midwest	South	West
Schools	6183	7146	7558	05207
Students	1275924	1309211	1386268	865039
Teachers	94622	81862	105509	56128
Graduates	77513	60547	67842	36965

For an area sample PSU's actual count, it includes the original list plus the unweighted adds. Subtract from the regional totals the area sample PSU actual count above to obtain a set of reduced regional totals. For each region we rake the nonsample PSU estimates across the board so that their sum equals the reduced total.

Table 2 above gives the 1993 regional totals based on the complete enumeration plus the follow up area sample estimates of the adds. The raking factors associated with the proposed procedure are given in Table 3.

Table 3
Raking Factors for Region

	Northeast	Midwest	South	West
Schools	.9789	.9725	1.0518	.9557
Students	.9933	.9813	1.0156	.9838
Teachers	.9880	.9796	1.0199	.9862
Graduates	.9962	.9973	1.0044	.9862

Note that the factors are all within 5% of 1.0000. It seems in the South our indirect estimator slightly over estimated adds.

Finally, for the raked counts by state for 1993 for the proposed method we have Table C. The table entries which are referred to as scaled counts are the adjusted state counts of Table B after the application of the raking procedure. The indicated errors are root mean square errors. The bias component of the estimates were derived from estimates of bias at the regional level, based on differences between totals using the current estimation procedure (unbiased) and the proposed indirect metnod. These regional bias estimates were proportionately allocated across the respective states.

V. Summary and Recommendations

On the basis of our test results, we recommend the indirect estimation approach to adjust the list frame counts for the PSS along with the raking to regional totals to obtain the adjusted state totals. We would use the poststratification via NCES and size of school vs an across the board adjustment or poststratification based on sample design. We have made preliminary evaluations of the results of the techniques on 1991 and 1993 data and plan to apply them on 1995 data.

VI. References

1. Broughman, S., Gerald, E., Bynum, L. And Stoner, K. (1994), Private School Universe Survey, 1991-92, National Center of Education Statistics, *Statistical Analysis Report NCES 94-350,* Washington, DC; U.S. Department of Education.

2. Hartley, H.O., J.N.K. Rao, and J. Kiefer (1962), "Sampling with Unequal Probabilities and Without Replacement," Annals of Mathematical Statistics, 33, 350-74.

3. Woodruff, R.S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate," Journal of the American Statistical Association, 66, 411-14.

Table A
Numbers of Private Schools, Students, Teachers, Graduates, and Student-Teacher Ratio, by States, for 1993

CALIFORNIA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	3009	562847	34501	16	23746
Adjusted	3224	576047	35718	16	24039
Ratio A/U	1.071609	1.023451	1.035274	0.988580	1.012336
INDIANA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	619	91985	6138	14	4012
Adjusted	686	95447	6450	14	4097
Ratio A/U	1.108312	1.037631	1.050840	0.987429	1.021151
NEW YORK	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	1865	464172	33812	13	25682
Adjusted	1974	472562	34735	13	26048
Ratio A/U	1.058298	1.018076	1.027305	0.991016	1.014259
TEXAS	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	1024	186975	14529	12	7424
Adjusted	1177	199967	15728	12	7787
Ratio A/U	1.149020	1.069484	1.034274	0.987960	1.048821
VERMONT	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	84	9107	945	9	1081
Adjusted	98	9648	1021	9	1116
Ratio A/U	1.165207	1.059432	1.080959	0.980085	1.032137
WYOMING	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Unadjusted	34	1918	167	11	29
Adjusted	40	2112	192	11	35
Ratio A/U	1.185055	1.101031	1.148996	0.958255	1.210001

Table B

Adjusted Figures and Standard Deviations, with Percent of Variance Attributed to 1st Component, for 1993

CALIFORNIA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	3224	576047	35718	16	24039
Stand. Dev.	33	2967	240	0.0373	104
% 1st Comp.	28.982	15.746	20.568		6.478
INDIANA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	686	95447	6450	14	4097
Stand. Dev.	16	1213	96	0.0643	58
% 1st . Comp.	39.289	13.057	16.186		2.955
NEW YORK	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	1974	472562	34735	13	26048
Stand. Dev.	19	2512	249	0.0337	131
% 1st Comp.	22.620	8.874	9.750		4.776
TEXAS	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	1177	199967	15728	12	7787
Stand. Dev.	15	1829	160	0.0359	66
% 1st Comp.	23.996	11.081	14.921		4.893
VERMONT	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	98	9648	1021	9	1116
Stand. Dev.	4	272	35	0.0975	38
% 1st Comp.	13.271	6.618	7.326		2.258
WYOMING	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Adjusted	40	2112	192	11	35
Stand. Dev.	2	93	11	0.2242	5
% 1st Comp.	9.514	3.925	5.630		1.258

Table C
Scaled Counts, and Student-Teacher Ratio, with Associated Errors, by State, for 1993: Proposed Method

CALIFORNIA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	3081	566722	35116	16	23707
Error	140.045	9627.003	636.866	0.563	342.895
INDIANA	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	667	93663	6319	14	4086
Error	24.163	2117.049	159.529	0.853	58.888
NEW YORK	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	1932	469394	34317	13	25950
Error	44.925	4016.141	481.261	0.281	163.220
TEXAS	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	1238	203089	16041	12	7821
Error	66.289	3675.202	359.663	0.367	75.029
VERMONT	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	96	9583	1009	9	1112
Error	4.818	278.540	36.876	0.737	38.791
WYOMING	**Schools**	**Students**	**Teachers**	**S-T Ratio**	**Graduates**
Scaled Value	39	2109	191	11	36
Error	2.814	92.107	11.296	1.828	5.023