DOCUMENT RESUME

ED 336 399 TM 017 042

AUTHOR Kaufman, Steven
TITLE 1988 Schools and Staffing Survey Sample Design and Estimation. Schools and Staffing Survey. Technical Report.
INSTITUTION National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-91-127
PUB DATE May 91
NOTE 85p.; Data Series: SP-SAS-87/88-7.4.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Administrators; Databases; Elementary School Teachers; Elementary Secondary Education; *Estimation (Mathematics); Mail Surveys; Mathematical Models; *National Surveys; Private Education; Public Education; *Research Design; *Sampling; School Districts; *School Surveys; Secondary School Teachers; Tables (Data); Teacher Supply and Demand
IDENTIFIERS *Schools and Staffing Survey (NCES); *Staffing Patterns

ABSTRACT

The Schools and Staffing Survey (SASS) represents the first time the National Center for Education Statistics has integrated three of the Elementary and Secondary Education Surveys: the Teacher Demand and Shortage Surveys, Public and Private School Surveys, and Teacher Surveys. The SASS was designed to measure the critical aspects of teacher supply and demand, the composition of teacher and administrator work force, and the general status of teaching and schooling. The SASS was conducted by the Bureau of the Census in the 1987-88 school year. The SASS sample included 12,823 public and private schools and administrators, 65,124 teachers, and 5,592 Local Education Agencies (LEAs). The merger of the three surveys produced one database to provide comparable linkage data among LEAs, schools, and teachers. Response rates ranged from 66% for private school teachers to 94.4% for public school administrators. Sample selection is discussed for public and private sectors. The discussion of estimation considers: weighting; item response rates and imputation; variance estimation; and frame evaluation. Data are reported in 18 tables. Three appendices discuss: minimizing school overlap with other national and longitudinal surveys; allocating sample groups to minimize loss of prediction for specific estimators; and the effect of a finite population correction on SASS variance estimates. (SLD)

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **May 1991**

Schools and Staffing Survey

1988 Schools and Staffing Survey Sample Design and Estimation

Data Series:
SP–SAS–87/88–7.4

U.S. Department of Education
Office of Educational Research and Improvement NCES 91-127

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report May 1991

Schools and Staffing Survey

1988 Schools and Staffing Survey Sample Design and Estimation

Steven Kaufman
Elementary and Secondary Education Statistics Division
National Center for Education Statistics

Data Series:
SP–SAS–87/88–7.4

U.S. Department of Education
Office of Educational Research and Improvement NCES 91-127

U.S. Department of Education
Lamar Alexander
Secretary

Office of Educational Research and Improvement
Bruno V. Manno
Acting Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Acting Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

May 1991

Contact:
Steven Kaufman
(202) 219–1337

Table of Contents

Table of Tables

Introduction

General Goals

The Schools and Staffing Survey (SASS) is the first time the National Center of Education Statistics (NCES) has integrated three of the Elementary and Secondary Education Surveys: the "Teacher Demand and Shortage Surveys," the "Public and Private School Surveys," and the "Teacher Surveys." SASS components are: "Teacher Demand and Shortage Survey", the "School and School Administrator Surveys", and the "Teacher Survey". The survey was designed to measure more accurately the critical aspects of teaching supply and demand, the composition of the administrator and teacher work force, and the status of teaching and schooling generally. The SASS was conducted by the Bureau of the Census during the 87-88 school year. Approximately 13,000 schools and administrators, 65,000 teachers, and 5,600 Local Education Agencies comprise the SASS sample. The idea behind the merger is to produce one database that would provide comparable linkage data between the LEAs (local education agencies), schools, and teachers. In addition, SASS is the first time these three surveys have the same reference period. To accomplish this:

1) Schools were selected first. Each selected school received a school and administrator questionnaire.

2) A sample of teachers was selected, within each selected school. Each selected teacher received a teacher questionnaire.

and 3) For public schools, the LEAs associated with the selected schools received a Teacher Demand and Shortage (TDS) questionnaire. Selected private schools received a TDS questionnaire, but were asked to respond only for the school.

An additional requirement of the survey design is that there be minimal school overlap between the National Assessment of Educational Progress (NAEP) and the National Education Longitudinal Study (NELS) surveys which also went to field in '88. The purpose of this requirement is to reduce the collection burden placed on the selected schools by NCES. This requirement was satisfied by adjusting the selection probabilities given the NAEP and NELS samples so that the overlap between the surveys was minimal. This was done in a fashion which maintained the original selection probabilities.

The SASS is designed to provide the following analytical comparisons for each of the SASS surveys, except where otherwise noted:

1) Public and private sector comparisons,
2) State public sector comparisons,
3) Elementary, secondary and combined school comparisons for the public and private sectors,
4) Detailed private sector affiliation comparisons,
and 5) Fields of teaching assignment comparisons for public and private sector teachers.

These requirements were satisfied by allocating the sample using an optimization algorithm which minimized the relative loss in precision with respect to sector, national, state and level estimates.

The organization of this paper is to describe the procedures used for the: 1) school and teacher sample stratum allocation, 2) minimization of overlap with NAEP and NELS, 3) public school sample design, 4) LEA sample design, 5) private school sample design, 6) teacher sample design (including within school teacher allocation), 7) weighting, 8) imputation, 9) variance estimation techniques and 10) frame evaluation.

Response Rates

Below are the weighted questionnaire response rates for the SASS components. These response rates are defined as the weighted number of in scope responding questionnaires divided by the weighted number of in scope sample cases. The overall response rate for a particular item (questionnaire response rate times item response rate) may be lower than the respective response rate below, because the item nonresponse rates are not included in the figures below.

Table 1. -- Questionnaire response rates

Teacher Demand and Shortage Survey	
Public LEAs	90.8%
Private schools	66.0%
School Administrator Survey	
Public	94.4%
Private	79.3%
School Survey	
Public	91.9%
Private	78.6%
Teacher Survey	
Public	86.4%
Private	79.1%

A future paper will examine survey response rates and possible bias in more detail.

Scope of SASS

The following survey terms define the scope of the components of the SASS.

Teacher Demand and Shortage Survey

1) Private School

A private school was defined as a school not in the public system that provides instruction for any of grades 1-12 where the instruction was not given exclusively in a private home.

2) Local Education Agency (LEA)

An LEA, or public school district, was defined as a government agency administratively responsible for providing public elementary and/or secondary instruction and educational support services. The agency or administrative unit must operate under a public board of education. Districts which operate only one school, as well as districts which do not operate schools but do hire teachers are included.

3) Out-of-Scope

An LEA or private school was considered out-of-scope for the Teacher Demand and Shortage Survey if it did not employ elementary or secondary teachers. In addition, if it was undetermined whether a private school operates in a private home and its enrollment was less than 10 students and it had only one teacher, the school was considered out-of-scope. If a private school was classified as out-of-scope for the Private School Survey, it was automatically classified as out-of-scope for the Teacher Demand and Shortage Survey.

School Survey

1) Public School

A public school was defined as an institution which provides educational services, has one or more teachers to give instruction, is located in one or more buildings, receives public funds as primary support, and is operated by an education agency. Prison schools, schools operated by the Department of Defense and the Bureau of Indian Affairs were included.

2) Private School

See above for definition.

3) Out-of-Scope (public schools)

A public school was considered out-of-scope if it did not have any students in any of grades 1-12. Schools offering only kindergarten and pre-kindergarten were deleted from the sampling frame before the sample was selected. If a school was determined to be out-of-scope after editing its questionnaire, it was deleted from the data tape.

4) Out-of-Scope (private schools)

A private school was considered out-of-scope if it did not have any students in any of grades 1-12, if it operated in a private home that was used as a family residence, or if it was undetermined whether it operated in a private home and its enrollment was less than 10 students or it had only one teacher. Out-of-scope schools were deleted from the sampling frame before the sample was selected. If a school was determined to be out-of-scope after editing its questionnaire, it was deleted from the data tape.

School Administrator Survey

1) Out-of-Scope

A school administrator sample case was considered out-of-scope if the school did not have an administrator. Also, if a sample administrator's school was considered out-of-scope, the administrator was automatically classified as out-of-scope.

Teacher Survey

1) Teacher

A teacher is defined as any full-time or part-time teacher whose primary assignment was teaching in any of grades K-12. Itinerant teachers were included, as well as long-term substitutes who were filling the role a regular teacher on an indefinite basis. An itinerant teacher is defined as a teacher who teaches at more than one school.

2) Out-of-Scope

A sample teacher was considered out-of-scope if he/she was a short-term substitute, a student teacher, a nonteaching specialist (e.g., guidance counselor, librarian, etc.), an administrator (e.g., principal, assistant principal, etc.), a teacher's aide, or in some other professional or support staff position. If a sample school was considered out-of-scope, all teachers from that school were also considered out-of-scope.

If an LEA was classified as out-of-scope, its teachers, administrators and schools were also classified as out-of-scope. Likewise if a school was classified as out-of-scope, its teachers and administrators were also considered out-of-scope. In addition, if a private school was out-of-scope for the Private School Survey, it was also considered as out-of-scope for the Teacher Demand and Shortage Survey.

School and Teacher Allocation

School and Teacher Allocation

In order to allocate the SASS school and teacher samples, it was necessary to determine the analytical goals for both school and teacher estimates. The major analytical school and teacher goals in terms of desired estimates are listed below:

1) Overall national estimates

2) State comparisons for public sector

3) School level (elementary, secondary, and combined) comparisons for both public and private sectors

and 4) Public vs. private comparisons

From these design goals, it was decided to stratify the school selection process by sector/state/level. A description of how the schools and teachers were allocated to these strata will be provided first. Then, additional design goals will be stated. Finally, how the sector/state/level allocation was modified to satisfy the additional goals will be described.

Generally speaking, the following can be said: The optimum allocation for each of the above comparisons is quite different. For national estimates, schools and teachers should be allocated proportional to their numbers within each stratum. State comparisons are best when each state gets an equal sample allocation. Using the best state allocation would provide less efficient national estimates (e.g., California would get the same sample size as Delaware, even though California contributes a lot more to the national variance than Delaware). Likewise, the optimum allocation for level and sector are both different and what's good for one comparison is less efficient for the other. Therefore, a compromise allocation was determined. The compromise allocation weighted each of the goals stated above equally.

The idea behind the compromise allocation was to look at the relative loss of efficiency (i.e., the ratio of the variance for the proposed allocation to the variance for the optimum allocation), for each of the four individual optimum allocations stated above (i.e., national, state, level and sector). The optimum allocations assumed a linear cost function with fixed total costs. These relative efficiencies were computed for both school and teacher estimates. The compromise SASS school and teacher allocation was the allocation that minimized the sum of these eight relative efficiencies (four school and four teacher).

An additional design goal was:

5) for teacher estimates, to provide comparisons by teacher's primary teaching assignment.

To increase the number of selected teachers per school so that primary teacher assignment comparisons can be made (goal 5), the within school variance component of the total teacher variance, in the compromise allocation, was multiplied by the number of desired teacher assignments (the teacher assignments are provided in the teacher sample selection section under teacher sorting). Appendix 2 provides more details about the allocation.

Tables 2 and 3 provide the results from the compromise allocation. Table 2 provides the average number of selected teachers per school and table 3 provides the school allocation for the state by level stratum allocations. Table 4 provides the number of public LEAs selected for SASS.

The final design goal was:

6) for private schools, to provide comparisons by detailed affiliation.

Table 6 provides the required sample sizes for the thirteen private school affiliations to satisfy goal 6. This table was produced by allocating at least 100 schools to each affiliation. If an affiliation had less than 100 schools on the frame then all schools in the affiliation were selected. The remainder of the private school sample was allocated proportional to number of schools in the affiliation.

Within a state/level stratum, a raking procedure was used to distribute the private school allocation (table 3) to the thirteen affiliations. This process maintained the affiliation allocation in table 6, as well as the allocation in table 3.

The private school stratification was then state by level by affiliation. Table 5 provides the number of schools in the private school list frame by affiliation/level strata.

Table 2. -- Average number of selected teachers per school

Public schools

	teachers/school
Elementary schools	4
Secondary schools	8
Combined schools	6

Private schools

	teachers/school
Elementary schools	4
Secondary schools	5
Combined schools	3

Table 3. -- SASS school allocation (state by school level)

State	Public schools				Private schools				Total
	Elem.	Sec.	Comb.	Total	Elem.	Sec.	Comb.	Total	
Total	4994	2907	1430	9331	2017	409	1065	3491	12823
Alabama	73	45	56	174	20	2	25	47	220
Alaska	43	15	43	100	4	1	5	10	110
Arizona	77	35	10	122	20	4	15	38	160
Arkansas	65	51	21	137	10	1	9	19	157
California	329	187	139	655	199	41	115	355	1010
Colorado	79	45	16	139	19	4	14	37	176
Connecticut	78	39	11	128	35	13	17	65	192
Delaware	44	19	8	72	19	4	14	37	109
D. of Columbia	54	11	7	72	20	8	10	38	111
Florida	135	87	53	275	63	11	71	144	419
Georgia	110	74	19	202	19	4	44	67	269
Hawaii	55	14	5	74	25	5	9	39	113
Idaho	68	30	4	103	6	1	3	10	113
Illinois	194	117	45	357	131	23	43	197	553
Indiana	107	66	17	190	47	6	17	71	261
Iowa	76	55	32	163	32	5	5	42	205
Kansas	82	51	13	146	19	4	6	29	174
Kentucky	83	40	31	154	29	6	13	47	201
Louisiana	75	57	54	187	43	11	32	86	273
Maine	72	27	6	105	7	5	12	24	130
Maryland	82	42	28	152	40	11	19	70	221
Massachusetts	112	66	18	196	48	22	25	95	291
Michigan	153	109	42	304	81	14	34	129	433
Minnesota	71	56	50	177	49	6	15	70	247
Mississippi	59	40	37	136	12	4	22	38	174
Missouri	95	69	34	198	58	10	17	84	283
Montana	72	23	11	106	10	1	3	14	120
Nebraska	70	25	29	124	26	6	4	35	160
Nevada	67	20	8	95	9	1	3	13	108
New Hampshire	67	21	4	91	11	6	9	26	117
New Jersey	135	78	30	243	83	19	35	137	381
New Mexico	67	29	7	103	13	2	7	22	125
New York	214	162	104	480	198	41	83	321	801
North Carolina	113	78	12	203	17	3	34	53	256
North Dakota	47	17	36	100	12	2	1	15	115
Ohio	174	134	51	360	100	19	19	138	498
Oklahoma	87	68	25	179	8	2	4	14	193
Oregon	83	40	8	132	16	3	10	29	160
Pennsylvania	175	121	42	338	156	27	61	244	582
Rhode Island	60	17	3	80	20	5	9	35	115
South Carolina	75	51	20	146	13	2	28	43	188
South Dakota	62	18	20	100	13	2	6	21	121

Table 3. -- SASS school allocation (state by school level) -- Continued

State	Public schools				Private schools				Total
	Elem.	Sec.	Comb.	Total	Elem.	Sec.	Comb.	Total	
Tennessee	94	50	26	170	22	6	27	55	226
Texas	262	224	93	579	67	10	25	102	681
Utah	75	28	12	115	3	1	1	6	121
Vermont	74	15	4	93	7	5	6	18	110
Virginia	107	69	26	203	29	6	29	64	267
Washington	88	55	19	163	26	4	22	51	21
West Virginia	85	33	12	131	7	2	5	14	145
Wisconsin	98	62	22	182	95	9	19	123	305
Wyoming	70	21	7	99	7	0	2	9	107
Total teachers	19975	23257	8582	51814	8070	2045	3195	13310	65124

Table 4. -- Number of sampled public LEAs

State	LEAs
Total	5592
1 Alabama	92
2 Alaska	40
3 Arizona	76
4 Arkansas	104
5 California	354
6 Colorado	64
7 Connecticut	87
8 Delaware	19
9 D. of Columbia	1
10 Florida	53
11 Georgia	110
12 Hawaii	1
13 Idaho	61
14 Illinois	266
15 Indiana	137
16 Iowa	132
17 Kansas	108
18 Kentucky	99
19 Louisiana	66
20 Maine	83
21 Maryland	24
22 Massachusetts	148
23 Michigan	208
24 Minnesota	139
25 Mississippi	95
26 Missouri	147
27 Montana	85
28 Nebraska	100
29 Nevada	17
30 New Hampshire	66
31 New Jersey	180
32 New Mexico	53
33 New York	294
34 North Carolina	105
35 North Dakota	78
36 Ohio	275
37 Oklahoma	142
38 Oregon	90
39 Pennsylvania	244
40 Rhode Island	31
41 South Carolina	69

Table 4. -- Number of sampled public LEAs -- Continued

42 South Dakota	80
43 Tennessee	84
44 Texas	350
45 Utah	30
46 Vermont	78
47 Virginia	92
48 Washington	95
49 West Virginia	57
50 Wisconsin	141
51 Wyoming	44

Table 5. -- List frame sizes for private school affiliations

Affiliation	Frame sizes Elementary	Secondary	Combined	Total
Total	14653	2465	7064	24182
Catholic	7618	1377	184	9179
Friends	26	8	30	63
Episcopal	177	41	69	287
Jewish	276	110	186	572
Lutheran	1502	81	76	1659
Seventh-Day Adventist	779	62	451	1292
Christian Schools Intl	146	31	80	257
Amer Assoc Christ. Schools	139	6	901	1046
Exceptional Children	19	10	124	153
Military Schools	0	17	9	26
Montessori	228	1	64	293
Independent Schools	240	257	487	984
Other	3505	466	4400	8371

Table 6. -- Sample sizes for private school affiliations

Affiliation	List frame Elementary	Secondary	Combined	Total
Total	1705	353	910	2968
Catholic	637	131	12	781
Friends	26	8	30	63
Episcopal	72	20	28	120
Jewish	76	29	37	142
Lutheran	202	12	8	222
Seventh-Day Adventist	125	10	60	195
Christian Schools Int'	74	15	30	118
Amer Assoc Christ. Schools	25	1	150	177
Exceptional Children	17	7	86	110
Military Schools	0	17	9	26
Montessori	97	1	23	121
Independent Schools	45	54	73	172
Other	309	49	363	721

An additional 551 schools were allocated to the area frame sample (see the area frame sample section for more information).

Sample Selection

Minimizing the School Overlap between the NAEP and NELS Samples

The collection period for the NAEP (National Assessment of Educational Progress) and NELS (National Education Longitudinal Study) surveys was the same as the SASS survey. To minimize respondent burden from NCES surveys, it was decided to minimize the school overlap between the three surveys. Each survey used the Qualit. Education Data (QED) file as a school frame. Hence, it was easy to match schools between the surveys. Each survey, however, did independent frame refinements, so the frames did not match exactly. Given the time constraints to do this process, the minimization process was done only on schools that matched on either the NAEP or NELS frames via the QED identification number. Since SASS was by far the largest survey, SASS was selected last. This reduced the impact of the potentially large SASS ' rlap on NAEP and NELS. NAEP selected first; NELS selected next .nd unduplicated the NAEP sample from the NELS frame before selecting their sample. SASS selected last and minimized the overlap between both the NAEP and NELS samples.

The approach taken by SASS was to minimize the school overlap in a fashion such that the SASS selection probability for a school averaged over all possible NAEP, NELS and SASS samples equaled the original probability of selection, if no unduplication was required. Taking this approach meant giving some NAEP and NELS sampled schools a small probability of selection. However, by maintaining the original probability of selection, the impact the unduplication process had on SASS estimates was minimized. Only four SASS schools overlapped with the NAEP sample and there was no overlap with the NELS sample. For more details on the minimization process see appendix 1.

Public School and LEA Sample Selection

Public School Sample

This section describes the frame, stratification, sorting and sample selection. The school allocation was described earlier in the School and Teacher Allocation section. In total, 9317 public schools were selected.

The SASS school sample was selected to minimize the overlap between the NAEP and NELS school samples. See section, Minimizing the Overlap between the NAEP and NELS School Samples, for a description of that process.

Frame

The public school frame was the 1986 Quality Education Data (QED) file. The frame included regular public schools, Bureau of Indian Affairs schools and Department of Defense schools. Nonregular schools such as special education, vocational or technical schools were also included in the sample frame. Before sampling, duplicate schools and Department of Defense Schools outside of the United States were removed from the frame. Schools that only teach prekindergarten, kindergarten or adult education were also removed. There were a total of 80,384 schools on the public school frame. Some public schools were missing from the frame (see frame evaluation section). However, the QED file was believed to be the best frame available at the time.

Number of schools on public frame

School type	Number of schools
Elementary	54,463
Secondary	18,737
Combined	7,184

Stratification

Public schools had a total of 153 strata.

The first level of stratification was State (51) -- each of the 50 States and the District of Columbia.

Within each state there were 3 grade level strata (elementary, secondary, and combined school), defined as follows:

Regular Schools

Elementary	Lowest grade ≤ 6 and Highest grade ≤ 8
Secondary	Lowest grade ≥ 7 and Highest grade ≤ 12
Combined	Lowest grade ≤ 6 and Highest grade > 8

Nonregular School

Nonregular schools, which include special education, vocational, technical, adult education (if part of in-scope school) or alternative/continuation grades were classified as combined schools.

School Sorting

Before the sample was selected, the schools within each stratum were sorted. To facilitate the calculation of LEA weights, it was important to keep all schools within a stratum/LEA together. To accomplish this, some of the sort variables' values were changed to keep them the same for every school within a stratum/LEA. They were changed in the following manner:

1) All schools within a stratum/LEA had the first three digits of the ZIP code set equal to the ZIP code of the first school in the stratum/LEA.

2) All schools within a stratum/LEA had the urbanicity code changed to the urbanicity code most prevalent among the first 11 schools within the stratum/LEA. If there were less than 11 school in the LEA, the most prevalent value of urbanicity was used. If there was a tie the lower value was used.

After these fields were changed the schools within a stratum were sorted by the following variables:

1) LEA urbanicity -- 0 - unclassified, 1 - urban, 2 - suburban and 3 - rural;

2) LEA percent minority -- 1 - (0-5%), 2 - (6-20%), 3 - (21-50%), and 4 - (51% or more);

3) LEA Zip code -- The first three digits were used;

4) LEA ID number;

5) Highest grade in school;

6) School enrollment; and

7) PIN number -- The PIN number is a unique number assigned by QED which identifies the school.

Sample Selection

Within each stratum, schools were systematically selected using a probability proportionate to size algorithm. The measure of size used was the square root of the number of teachers in the school as reported on the QED file. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in the sample with certainty.

LEA Sample for Public Schools

During the initial design development of the SASS, selecting the LEAs first and then selecting schools within LEAs was considered. It was hypothesized that doing this would reduce the reliability of both school and teacher estimates. Simulations done on the reliability of LEA estimates when the LEAs were selected first confirmed this. The simulations also showed that selecting schools "first" would produce reasonably accurate LEA estimates. For these reasons the SASS sample selected schools first.

Hence, the LEA sample consists of the set of LEAs that were associated with the SASS public school sample. This provides the linkage between the LEA and the school. Each Bureau of Indian Affairs school and Department of Defense school was defined to be an LEA. This portion of the LEA sample represented the set of LEAs associated with schools.

Some LEAs were not associated with schools. Such LEAs may hire teachers that teach in schools from other LEAs. For SASS to represent such LEAs, a sample of these LEAs was also selected. The frame for this sample consisted of all LEAs on the '86 QED file that were not associated with schools. 1,077 LEAs were on this frame. 70 LEAs were selected and only 8 of the 70 sampled LEAs were actually in-scope (i.e., reported hiring teachers in SASS).

The frame was stratified into two strata: the first stratum consisted of LEAs not associated with schools that reported at least one teacher on the QED file; the second stratum consisted of LEAs not associated with schools that reported zero teachers on the QED file.

For the first stratum, a 1 in 10 sample was taken. The sample was selected using a systematic probability proportionate to size algorithm. The measure of size was the square root of the QED number of teachers. The sort variables were:

1) State (51) -- one for each state and the District of Columbia;

2) Urbanicity (3) -- 1 - urban, 2 - suburban, and 3 - rural;

3) Percent minority (4) -- 1 - (0-5%), 2 - (6-20%), 3 - (21-50%), and 4 - (51% or more);

4) First three digits of Zip code;

and 5) LEA ID.

For the second stratum, a 1 in 20 sample was taken. The sample was selected using a systematic equal probability algorithm. The sort variables were:

1) State (51) -- one for each state and the District of Columbia;

2) Urbanicity (3) -- 1 - urban, 2 - suburban, and 3 - rural;

3) Percent minority (4) -- 1 - (0-5%), 2 - (6-20%), 3 - (21-50%), and 4 - (51% or more);

4) First three digits of Zip code;

and 5) LEA ID.

Delaware, Nevada and West Virginia LEAs

For each state, a simulation study was done to assess the reliability of SASS LEA estimates. It showed that standard errors from Delaware, Nevada and West Virginia were very high relative to the LEA sampling rate (i.e., CVs larger than 20 percent with 90 percent of LEAs in sample). Since almost all LEAs were selected in these states, it was decided to include the nonselected LEAs in the LEA sample. Hence, all LEAs in these states were selected with probability one.

Private School Sample

This section describes the frame, stratification, sorting and private school selection. The school allocation is described in the School and Teacher Allocation section.

Frames

3,513 private schools were selected using a dual frame approach. A list frame was the primary private school frame and an area frame was used to find schools missing from the list frame and thereby compensate for the coverage problems of the list frame.

List Frame

The list frame used for private schools was a supplemented 1986 QED file. Various private school associations were asked to supply lists of their schools. Seventeen such lists were received. These lists were matched with the QED list and any association list school not found on the QED file was added to the frame. Before sampling, duplicate schools were excluded from the frame. Schools that only teach prekindergarten, kindergarten or adult education were also removed. The list frame consisted of 22,600 schools from the QED file and 1,586 schools added on from the association lists.

Area Frame

The area frame sample PSUs were originally selected by Westat for the 1983 Private School Survey. They used a dual frame design similar to the SASS design. The area frame consisted of 2,497 primary sampling units (PSUs). Each PSU consisted of a single county or independent city or cluster of such geographically contiguous areas defined so that each PSU had a minimum population of 10,000 according to the 1980 Census of Population. To avoid having PSUs covering too large a geographic area some PSUs had less than 10,000 in population. Counties of Alaska were excluded.

Eight PSUs with populations greater than 1,700,000 were included in the sample with certainty.

The strata were: a) census region (4 levels), b) metro/nonmetro status (2 levels) and c) whether the PSU's private school enrollment exceeded the median enrollment of the other PSUs in the census region/metro status strata (2 levels).

A minimum of four PSUs were allocated to each of the 16 strata (64 PSUs). 28 additional PSUs were allocated to the 16 strata to

more nearly approximate a uniform fraction of PSUs in the sample from each stratum.

The PSUs were selected as a systematic sample with probability proportionate to the square root of the 1980 PSU population. A total of 100 PSUs were in sample. Due to budget constraints during the '83 Private School Survey, the 100 PSUs were subsampled to 75 PSUs with a minimum of 3 selected PSUs per stratum. These 75 PSUs were used in the '88 SASS survey.

Area Sample School Frame Building

Within each of the 75 Westat PSUs, the Census Bureau attempted to find all eligible private schools (i.e., nonpublic schools providing the following: instruction for any grades 1-12, instruction not provided exclusively in the home, a school year at least 160 days long and a normal school day at least 4 hours long). An area canvas was not attempted. However, regional staff were used to create the frame using such sources as: yellow pages, non-roman catholic religious institutions, local education agencies, chamber of commerce, local government offices, commercial milk companies and commercial real estate offices. Roman Catholic religious institutions were not contacted because QED calls each Catholic diocese during its annual list update. Once these lists were constructed, they were matched with the updated QED file. Schools that did not match the QED list were contacted to make sure they were eligible schools. 746 eligible and 250 unable-to-contact schools comprised the area school frame. The 746 eligible and 250 unable-to-contact school weight up to respectively 6,909 and 753 schools nationally.

Frame Summary

The total number of schools from the list and area school frames are provided in table 7 below.

Table 7. -- Private school frames

	Unweighted count	Weighted count
List frame		
Total	24,186	24,186
QED list	22,600	22,600
Association add-ons	1,586	1,586
Area frame		
Total	996	7,662
Eligibles	746	6,909
No contacts	250	753

Private School List Sample

Stratification

For private schools, the frame was partitioned into an initial set of 1,989 cells. Only cells with schools in them comprised the sampling strata. Some of the initial cells didn't have any schools.

The first level of stratification was state (51) -- 1 for each state and the District of Columbia.

Within each state, school affiliation (13) was the next level of stratification:

1) Catholic schools

2) Friends schools

3) Episcopal schools

4) Jewish schools

5) Lutheran schools

6) Seventh-Day Adventist schools

7) Christian Schools International

8) Association of Christian Schools

9) Association of Private Schools for Exceptional Children

10) Association of Military Colleges and Schools of the U.S.

11) American Montessori Society Schools

12) National Association of Independent Schools

13) All Else

Within each state/affiliation, schools were stratified by school level (elementary, secondary, and combined schools). The definitions are provided below:

Regular Schools

Elementary	Lowest grade ≤ 6 and Highest grade ≤ 8
Secondary	Lowest grade ≥ 7 and Highest grade ≤ 12
Combined	Lowest grade ≤ 6 and Highest grade > 8

Nonregular School

Nonregular schools, which include special education, vocational, technical, adult education (if part of in-scope school) or alternative/continuation grades were classified as combined schools.

Due to the raking procedure, described in the school allocation section, some strata did not have any schools designated in the sample; in such cases, strata were collapsed with other strata until all strata had a least one designated sample school.

School Sorting

Within each stratum, sorting took place on the following variables:

1) Urbanicity -- 0 unclassified, 1 urban, 2 suburban, 3 rural and 4 affiliation add-on with no classification;

2) Zip code -- The first two digits were used;

3) Highest grade in the school;

4) Enrollment;

and 5) PIN number -- The PIN number is a unique number assigned by QED which identifies the school.

Sample Selection

Within each stratum, schools were systematically selected using a probability proportionate to size algorithm. The measure of size used was the square root of the QED number of teachers in the school. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in sample with certainty.

Area Frame Sample

Allocation

551 private schools were selected from the area sample school frame. In most states, all area frame schools were in sample. However, in six states: California, Florida, Illinois, Montana, Nevada and Texas, the number of area frame schools exceeded 30 percent of the state's total allocation. To maintain a reasonable balance between the numbers of list and area frame sample cases, in these six states the area frame schools were subsampled to 30 percent of the state's total allocated sample. After a state sample size was determined for the area frame, sample was allocated to PSUs so as to maximize the "self-weightedness" of the school sample (i.e., minimize the range of the weights). The sample sizes for the area sample school frame are provided in table 8.

Table 8. -- Sample sizes for the area sample school frame

	Sampling rates			
	Elementary	Secondary	Combined	Unable to contact
California	46 of 123	9 of 13	26 of 45	37 of 198
Florida	19 of 56	3 of 7	22 of 26	4 of 8
Illinois	40 of 74	7 of 8	13 of 45	6 of 13
Montana	1 of 1	1 of 1	2 of 5	1 of 2
Nevada	4 of 5	--	--	--
Texas	22 of 49	3 of 7	8 of 23	1 of 6

All area frame schools in the other states were selected with certainty.

-- means there were no cases in the frame to sample.

For the six states, the following sampling scheme was implemented.

Stratification

Schools were stratified by the following variables:

1) State;

2) PSU;

3) Two strata
   a. Eligible schools - Schools that were contacted and verified in-scope;

   and

   b. Unable-to-contact schools - Schools who could not be contacted and therefore it was not possible to verify in-scope status. Sort variables (grade level, affiliation and enrollment) could not be collected for these schools; and

4) School level - Eligible schools were stratified by grade level (elementary, secondary, and combined). In some cases, when the grade level was unknown, it was imputed.

   Since the school level was unknown for the unable-to-contacts, they were not stratified by this variable.

Sort Variables

Eligible schools were sorted using the following variables:

1) Affiliation - Catholic, other religious, and nonsectarian. In some cases, when the affiliation was unknown, it was imputed based on the name;

2) Enrollment;

and 3) Alphabetical order of name.

Unable-to-contact schools were sorted by alphabetical order of name.

Measure of Size

For eligible schools, the measure of size was the square root of the number of reported teachers. For unable-to-contact schools, this measure of size was unknown so they were assigned a measure of size equal to one.

Sample Selection

Within each stratum, eligible schools were systematically selected using a probability proportionate to size algorithm. The measure of size used was described above. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in sample with certainty.

Public and Private Teacher Sample Selection

This section describes the frame, stratification, within school allocation, sorting, and the selection for the public and private teacher sample.

Selecting the teacher sample involved the following steps. First, the selected schools were asked to provide teacher lists for their schools. From the lists, 56,242 public and 11,529 private teachers were selected. Teachers could be selected in one of two ways:

1) Most teachers were selected as part of the basic teacher sample. These teachers were allocated to approximately equalize the teacher weights within a school stratum;

or

2) 2447 bilingual/ESL teachers were selected independently from the basic sample. This sample was designed to produced reliable bilingual/ESL teacher estimates in California, Texas, Florida, Illinois, New York and all other states.

The public and private teacher samples will be described together because they were selected using identical methodology. The only differences were in the rate new teachers were oversampled and the average number of teachers selected within a school.

The details of the teacher selection are provided below.

Teacher Frame

Each selected school was asked to provide a list of their teachers. For each teacher on the list, the following was to be specified:

1) New/experienced - Teachers in their first, second or third year of teaching should be classified as a new teacher.

2) Bilingual/ESL - Teachers who use NATIVE LANGUAGE to instruct students with limited English proficiency (bilingual)
or
teachers providing students with limited English proficiency with intensive instruction in English (English as a Second Language).

Field of Teaching - Elementary teachers should be classified as: general elementary, special education or other.

Secondary teachers depending on primary teaching subject taught should be classified as: math, science, English, social studies, vocational education or other.

The above information for each teacher in a selected SASS school comprises the school teacher frame.

Twelve percent of the private schools and four percent of the public schools did not provide teacher lists. For these schools no teachers were selected. A factor in the teacher weighting is used to adjust the weights to reflect the fact that some schools did not provide teacher lists.

Basic Teacher Sample

Teacher Stratification

Within each selected school, teachers were stratified by their new/experienced classification. This was done to allow for the possibility of oversampling these groups in order to ensure sufficient sample for analytical purposes.

Within School Teacher Allocation

For public schools, it was decided that new/experienced teachers did not require oversampling. Therefore, teachers were allocated to the new and experienced categories proportional to their numbers in the school. However, for private teachers, to ensure that there would be enough new teachers in the Teacher Followup Survey (TFS), it was decided to oversample new teachers.

Before teachers were allocated to the new/experienced strata, the school was first allocated an overall number of teachers to be selected. This overall sample size was chosen so as to equalize the teacher weights within the school stratification (state/level and state/affiliation/level for public and private sectors, respectively), assuming no further teacher stratification.

Table 9 provides the average number of teachers to be selected within a school.

Table 9. -- Average selected teachers per school

	School Level		
	Elementary	Secondary	Combined
Public Schools	4	8	6
Private Schools	4	5	3

Given the numbers in table 9, the overall teacher sample size was chosen to equalize the teacher weights within a school stratum. Since the school sample was selected proportional to the square root of the number of teachers in the school, an equally weighted teacher sample within a school stratum was obtained by selecting t_i teachers in school i.

$$t_i = W_i * T_i (C/Y)$$

where: W_i is the school weight for school i (the inverse of the school selection probability).

T_i is the number of teachers in school i, as reported on the teacher list form.

C is the average number of teachers selected per school (See table 9).

Y is the sample average of the school's weighted measure of size over all schools in the school stratum. The weighted measure of size is defined below.

For noncertainty schools, the weighted measure of size equals school sampling interval times the square root of the QED number of teachers in school. The measure of size for certainty schools is the QED number of teachers in the school.

To make sure a school wasn't overburdened, the maximum number of teachers per school was set at twice the average number of teachers per school from table 9. Each school was asked for at least one teacher.

65,124 teachers were designated for selection, while 65,327 were actually selected. This slight difference was due to the fact that in allocating the sample, Y was based on QED teacher counts instead of reported teacher counts from the school. This caused the overall average number of teachers per school to be slightly different than the target numbers in table 9.

Given the overall school allocation of teachers, t_i, teachers were allocated to the new/experienced strata, t_{ni} and t_{ei} respectively, in the following manner.

$$t_{ni}=(A*T_{ni}*t_i)/(T_{ei}+A*T_{ni})$$

and

$$t_{ei}=(T_{ei}*t_i)/(T_{ei}+A*T_{ni})$$

where: A is the oversampling factor (see table 10).

T_{ni} is the number of new teachers in school i.

T_{ei} is the number of experienced teachers in school i.

Table 10. -- Oversampling factor for new teachers (A)

	A (oversampling factor)
Public teachers	1.0
Private teachers	1.6

The oversampling in private schools was done to ensure the ability to analyze new private school teachers for both the SASS and the TFS surveys. The expected public new teacher sample size was large enough to meet the NCES's analysis plans without oversampling.

Teacher Sorting

Within each new and experienced teacher stratum, teachers were sorted by primary field of teaching. Elementary teachers were sorted by general elementary, special education or other teaching assignment. Secondary teachers were sorted by math, science, English, social studies, vocational education or other teaching assignment. When combined schools had both elementary and secondary teachers, the teachers were sorted by grade level/primary field of teaching. This was done to assure a good distribution of teachers by field of teaching.

Teacher Selection

Within each school and teacher stratum, teachers were selected systematically with equal probability.

Bilingual/ESL Teacher Supplemental Sample

Independent of the teacher selection process above (the basic sample), a sample of bilingual/ESL teachers was selected from the teacher lists from both public and private SASS school samples, using the bilingual/ESL code. This supplemental sample was designed to provide estimates for California, Texas, Florida, Illinois, New York, and the rest of the United States.

Bilingual Supplement Stratification

There were six strata:

1) California
2) Texas
3) Florida
4) Illinois
5) New York
6) All other states

Bilingual Sample Allocation

The goal of the bilingual supplement was to be able to analyze bilingual teachers in the strata stated above. To achieve this goal, it was decided to select approximately 500 teachers in each strata. Since accurate counts of bilingual/ESL teachers were not available, determining the expected number of bilingual/ESL teachers in the selected SASS schools was very difficult. Since teachers were selected on a flow basis as the lists came in from the field, an estimate could not be made using SASS. To minimize this problem, conservative sampling rates were used so that more than the required number of teachers would be selected. After a specified time, estimates of final sample sizes were made and a sampling reduction rate was determined and applied to provide the desired sample sizes.

The sample size within each school was chosen to be proportional to the weighted number of bilingual teachers in the school (i.e., proportional to the inverse of the school selection probability times the number of bilingual teachers in the school).

No school was asked to provide a sample of more than 20 teachers. If the sum of the basic and bilingual teacher samples was greater than 20 in any school, the bilingual sample was systematically reduced to make the total teacher sample equal to 20. Table 11 provides the actual number of selected bilingual teachers in the final sample.

Table 11. -- Number of bilingual teachers in the supplement

State	Number of teachers	
Total	2,447	
Illinois	195	All SASS bilingual teachers selected
Florida	190	All SASS bilingual teachers selected
California	513	
Texas	563	
New York	468	
Rest of U.S.	518	

Some state allocations were greater than 500 because the sample reduction rates were computed due to timing constraints before all teachers were selected.

Sort Variables

Within each school, teachers were sorted by the new/experienced category by primary field of teaching. Elementary teachers were sorted by general elementary, special education or other teaching assignment. Secondary teachers were sorted by math, science, English, social studies, vocational education or other teaching assignment. This was done to assure a good distribution of teachers by field of teaching.

Teacher Selection

Teachers were selected systematically with equal probability among those schools having bilingual/ESL teachers.

Teacher Sampling for Telephone Followup

Given the size of the teacher sample (67,771 basic and 2,447 bilingual/ESL teachers) and the fact that this was the first time SASS was fielded, the Census Bureau was not sure there was enough time to do telephone followups for all teachers not reporting by mail. To handle this situation, the teachers selected above were systematically placed into 100 approximately equal sized groups. Once the mail collection was closed out, this provided the capability of producing different sized samples of non-mail return teachers for telephone followup collection. Hopefully, all non-mail return teachers would be followed up. However, if time or money constraints became critical, this provided the capability to reduce the number of followup cases. In fact, 57 percent of the non-mail return teachers (i.e., 23,365 teachers), were selected for telephone followup. Therefore, 57 of the 100 groups were randomly chosen for telephone followup.

Estimation

Weighting

This section describes the weighting processes for the SASS samples. For each questionnaire, the formula for the weight will be presented, along with a brief description of each component. When computations are done within cells, such as nonresponse adjustments, the cells will be described. Sometimes a cell did not have enough data to produce a reliable estimate; in such cases, cells were collapsed. The least important variables were always collapsed first. The collapse criteria are also described.

First the school weight will be described. Since the public and private school weights have the same structure, they will be presented together. They differ only in the definition of the cells used to compute the nonresponse adjustment and the first-stage ratio adjustment factors. These cells will be described separately within the school weight section.

Since the public and private administrator weights, as well as the private teacher demand and shortage weights are similar to the school weights, they will be described next.

In the third section, the public teacher demand and shortage weights will be described.

In the last weighting section, the teacher weights will be described. Since the public and private weights have the same structure, they will be presented together. They differ only in the definition of the cells used to compute the various weighting factors. These cells will be described separately within the school weight section.

School Weight

The final weight for the public and private school data is:

(Basic Weight) X (Sampling Adjustment Factor) X (School Noninterview Factor) X (Frame Ratio Adjustment Factor)

Where:

Basic Weight	is the inverse of the probability of selection of the school.
Sampling Adjustment Factor	is an adjustment that accounts for unusual circumstances that affect the school's probability of selection, such as merger, split or duplication.
School Noninterview Adjustment	is an adjustment that accounts for total school nonresponse. It's the weighted (basic weight*sampling adjustment factor) ratio of total eligible in-scope schools to the total responding in-scope schools within cells.
Frame Ratio Adjustment Factor	This factor adjusts the sample estimates to known frame totals. For public schools, it is equal to the ratio of the total number of SASS frame noncertainty schools to the weighted sample estimate of total number of noncertainty schools in the frame. For private schools, the adjustment is the same, except for the area frame. For the area frame, instead of SASS frame totals, weighted (PSU weight) area frame totals are used in the numerator. This ratio is calculated within cells.

Cells

The school noninterview and frame ratio adjustments are computed within cells.

Public

For public schools, the noninterview adjustment cells were: state by school level by enrollment size class by urbanicity. If the factor was less than 1.5 and there were at least 15 schools in the cell, no collapsing was done. Otherwise, cells were collapsed (enrollment size first, urbanicity second and level third). The frame ratio adjustment cells were: state by grade level by urbanicity. The results of the QED factors were reviewed by Census Bureau staff. If collapsing was necessary (e.g., where a cell contains QED schools, but no SASS schools), urbanicity was collapsed first and grade level was collapsed last.

Private

For private list frame schools, the noninterview adjustment cells were: census region by affiliation (Catholic, other religious and nonsectarian) by grade level. If the factor was less than 1.5 and there were at least 15 schools in the cell, no collapsing was done. If collapsing was done, urbanicity was collapsed first, grade level second and affiliation last. The frame ratio factor cells were: grade level by urbanicity by affiliation (thirteen sampling strata affiliations). The results of the frame factors were reviewed by Census Bureau staff. If collapsing was necessary (e.g., where a cell contains QED schools, but no SASS schools), urbanicity was collapsed first, grade level second and affiliation last.

For private area frame schools, the noninterview adjustment cells were: census region by affiliation (3 levels) by grade level. Within census region, the unable-to-contact schools were another cell by themselves. If the factor was less than 1.5 and there were at least 15 schools in the cell, no collapsing was necessary. If collapsing was necessary, grade level was collapsed first, and affiliation was collapsed last. The frame ratio factor cells were: grade level by urbanicity. The results of the frame factors were reviewed by Census Bureau staff. If collapsing was necessary (e.g., where a cell contains QED schools, but no SASS schools), urbanicity was collapsed first and grade level was collapsed last.

Administrator and Private Teacher Demand and Shortage Questionnaires

The public and private administrator weighting, as well as the private teacher demand and shortage weighting was done the same way as the school questionnaire weighting described above. Since the respondents for each of these surveys were different, the weighting process was done separately for each questionnaire. The sum of the administrator weights may not equal the sum of the school weights because some schools do not have administrators.

The final weight for the district data is:

(Basic Weight) X (sampling Adjustment Factor) X (LEA Noninterview Factor) X (Frame Ratio Adjustment Factor)

Where: Basic Weight is the inverse of the probability of selection of the LEA. See section below for more details.

Sampling Adjustment Factor is an adjustment that accounts for unusual circumstances that affect the LEA's probability of selection, such as merger, split or duplication.

Noninterview Adjustment is an adjustment that accounts for total LEA nonresponse. It's the weighted (basic weight*sampling adjustment factor) ratio of total eligible in-scope LEAs to the total responding in-scope LEAs, computed within cells.

Frame Ratio Adjustment Factor This factor adjusts the sample estimates to known QED totals. It's the ratio of the total number of QED noncertainty LEAs to the weighted sample estimate of the total number of noncertainty LEAs in the frame, computed within cells.

Cells

The noninterview and frame ratio adjustments are computed within cells.

The noninterview adjustment cells were: state by LEA enrollment size class by percent minority enrollment size class. If the factor was less than 1.5 and there were at least 10 LEAs in the

cell no collapsing was done. Otherwise, cells were collapsed (percent minority enrollment first and enrollment last). The frame ratio adjustment cells were: state by LEA enrollment size class level by urbanicity. The results of the frame factors were reviewed by Census Bureau staff. If collapsing was necessary (e.g., where a cell contains QED LEAs, but no SASS LEAs), urbanicity was collapsed first and LEA enrollment was collapsed last.

LEA Basic Weights

Given the complexity of the sampling scheme, LEA basic weights are not straight forward. There are three situations that need discussion: LEAs with school, LEAs without schools and LEAs in Delaware, Nevada and West Virginia which have two sets of basic weights.

LEAs with Schools

The LEA sample was not selected directly through the LEA. Instead, the LEAs were selected through the school (i.e., the LEAs associated with the school sample comprised the LEA sample). The basic weight, therefore, is more complicated then normal.

The basic weight for LEA k is the inverse of the probability of selection ($P_k(sel)$):

$$P_k(sel)=1-(1-P_k(Elem)) \times (1-P_k(Sec)) \times (1-P_k(Comb))$$

where: $P_k(Elem)$ is the probability of selecting LEA k in the elementary school stratum. This equals the sum of the school selection probabilities for the elementary schools in LEA k. If the sum is greater than one then $P_k(Elem)$ is set equal to one.

$P_k(Sec)$ is the probability of selecting LEA k in the secondary school stratum. This equals the sum of the school selection probabilities for the secondary schools in LEA k. If the sum is greater than one then $P_k(Sec)$ is set equal to one.

$P_k(Comb)$ is the probability of selecting LEA k in the combined school stratum. This equals the sum of the school selection probabilities for the combined schools in LEA k. If the sum is greater than one then $P_k(Comb)$ is set equal to one.

LEAs Without Schools

The basic weight for LEAs without schools was computed using the usual methodology for a single stage probability proportionate to size sample.

LEA Basic Weights for Delaware, Nevada and West Virginia

After the school selection in Delaware, Nevada and West Virginia, most LEAs were in the sample. Simulations showed that these states' LEA estimates would have a high variance in spite of most LEAs being in sample. To resolve this situation, all LEAs not selected in these three states were added-on to the sample. The basic weights for all LEAs in these states were therefore equal to one.

No schools, administrators or teachers were selected for schools that comprise these add-on LEAs. If an LEA estimate requires such data the basic weights equaled to one are not appropriate. As an example, say someone wants to know how many LEAs have schools with enrollments less than 100 students and suppose there are m total LEAs. This question can only be answered for the LEAs which have schools selected, say there were n such LEAs. Using the basic weights equal to 1 would produce an estimate equaled to n, if all LEAs have at least one such school. The other m-n LEAs would be missing from the estimate.

To handle LEA estimates requiring school, administrator or teacher data a second set of weights is available. The second set of weights uses the basic weight derived through the school selection described in the formula above. With these weights, the add-on schools have weight zero and are being represented by one of the n schools. Hence, data from the add-on LEAs are not required.

Summarizing, when LEA estimates are produced solely using LEA data then the first set of weights, with basic weight equal to 1, are appropriate. However, if LEA estimates require administrator, school or teacher data, then the second set of weights should be used.

Teacher Questionnaire

The final weight for public and private teachers is:

(Basic Weight) X (Sampling Adjustment Factor) X (School Noninterview Factor) X (Bilingual Oversampling Factor) X (Late Mail Return Adjustment Factor) X (Teacher-Within-School Noninterview Adjustment Factor) X (Frame Ratio Adjustment Factor)

Where:

Basic Weight is the inverse of the probability of selection of the teacher in either the basic SASS sample or the bilingual/ESL supplement. If a teacher is selected in both samples then the basic SASS sample weight is used.

Sampling Adjustment Factor is an adjustment that accounts for unusual circumstances that affect the schools probability of selection, such as merger, split or duplication. It's the same factor used in the school weight.

School Noninterview Adjustment is an adjustment that accounts for schools that did not have teachers selected because teacher lists were not provided by the school. It's the weighted (school basic weight*school sampling adjustment factor) ratio of total eligible in-scope schools to the total in-scope schools providing teacher lists, computed within cells.

Bilingual Oversampling Factor is an adjustment that accounts for the fact that bilingual teachers have two chances of selection. The adjustment is the inverse of the probability of selection of a teacher being selected in either basic or bilingual samples divided by the basic

weight.

Late Mail Return Adjustment Factor	is an adjustment that accounts for the fact that some sampled teachers were not selected for nonresponse followup. See below for more details.
Teacher-within-school noninterview adjustment factor	is an adjustment that accounts for sampled teachers that did not respond to the survey. It's the weighted (product of all previously defined components) ratio of the total eligible teachers to the total eligible responding teachers, computed within cells.
Frame Ratio Adjustment Factor	This factor adjusts the sample estimates to known frame totals of number of schools. These factors are the same used in the school weight. See that section for more information.

Late Mail Return Factor

As stated before, not all non-mail return teachers were eligible for a telephone followup. The late mail return factor reflects this in the weight. The factor equals one for teachers responding by mail. For teachers not responding by mail by the close out date, and selected for telephone followup, the inverse of the probability of selection for telephone followup (1.75) is the late mail return factor. Some teachers responded by mail, even though they responded after the mail close-out date and were not selected for followup. Using the 1.75 late mail return factor, these teachers are being represented by telephone followup teachers and should be excluded from the sample. To make use of this data and to avoid a double counting of these mail return teachers, an appropriate proportion of the telephone followup teachers were randomly chosen not to get the late mail return factor.

Cells

The school noninterview and the teacher within-school noninterview adjustments are computed within cells. The frame ratio adjustments are the same as those used in the school weight. The cells for the frame adjustments are described in the school weight section.

Public

For public schools, the school noninterview adjustment cells were: state by school level by enrollment size class by urbanicity. If the factor was less than 1.5 and there were at least 15 schools in the cell no collapsing was done. Otherwise, cells were collapsed (enrollment first, urbanicity second and level third). The teacher within-school noninterview adjustment cells were: state by field of teaching by experience level (new vs. experienced teacher) by school urbanicity. If the factor was less than 1.5 and there were at least 15 teachers in the cell, no collapsing was done. Otherwise, cells were collapsed (urbanicity first, experience level second and field of teaching third).

Private

For private list frame schools, the noninterview adjustment cells were: census region by affiliation (Catholic, other religious and nonsectarian) by grade level. If the factor was less than 1.5 and there were at least 15 schools in the cell, no collapsing was done. If collapsing was done, urbanicity was collapsed first, grade level second and affiliation last. The teacher within-school noninterview adjustment cells were: census region by affiliation (three levels) by field of teaching by experience level by urbanicity. If the factor was less than 1.5 and there were at least 15 teachers in the cell, no collapsing was done. If collapsing was done, urbanicity was collapsed first, teaching experience was collapsed second, field of teaching was collapsed third and affiliation was collapsed last.

For private area frame schools, the school noninterview adjustment cells were: census region by affiliation (three levels) by grade level. Within census region, the unable-to-contact schools were another cell by themselves. If the factor was less than 1.5 and there were at least 15 schools in the cell, no collapsing was done. If collapsing was done, grade level was collapsed first, and affiliation was collapsed last. The teacher within-school noninterview adjustment cells were: census region by affiliation (three levels) by field of teaching. Teachers from unable-to-contact schools were treated like another affiliation

in the cells. If the factor was less than 1.5 and there were at least 15 teachers in the cell, no collapsing was done. If collapsing was done, field of teaching was collapsed first, and affiliation was collapsed last.

Item Response Rates and Imputation

The item response rates (e.g., number sampled units responding to an item divided by the number of responding sampled units) for the SASS surveys ranged from 11 percent to 100 percent. Tables 12 and 13 provide a brief summary of the item response rates.

Imputations were calculated for most missing items on the teacher demand and shortage and school files. Such imputations are flagged so that an analyst may exclude them, if desired. A sequential hot deck methodology was used for the imputation. This "nearest neighbor" approach matches the nonrespondent school or district with the most similar respondent in the same stratum. The variables used to match within a stratum were urbanicity, percent minority and enrollment. A response from a particular item was then imputed in one of two ways: the matched school or district's response was directly assigned to the nonresponse or it was used as an adjustment factor with other data the nonrespondent reported.

On the public school file, all items were imputed. On the private school file items 7 and 35 were not imputed. On both the public and private teacher demand and shortage file items 3, 11, 12, 13 and 28 were not imputed. A copy of an individual questionnaire can be obtained by looking at a NCES survey report or tape documentation from that questionnaire or requesting a copy from the NCES.

Due to time constraints, the administrator and teacher files do not contain imputations for missing items. Instead, a missing data code is used to indicate missing items. At some later time when resources permit, new administrator and teacher files will be issued that contain imputed values.

Because the administrator and teacher files were not imputed, care must be taken when looking at estimates. Totals will be underestimates because missing values will be assumed to have a zero value. For averages, missing values will assume the average value of the variable of interest. This may represent an underestimate or overestimate.

Table 12. -- Item response rates

Survey	Range of item response rates	Percent of items with response of 90% or more	Percent of items with a response less than 75%
Teacher Demand and Shortage Survey			
Public	40-100%	74%	12%
Private	16-100%	70%	18%
Administrator Survey			
Public	70-100%	86%	2%
Private	72-100%	89%	2%
School Survey			
Public	43-100%	64%	11%
Private	11-100%	56%	8%
Teacher Survey			
Public	64-100%	90%	1%
Private	60-100%	89%	1%

The item response rates in this table are unweighted and do not reflect additional response loss due to complete questionnaire refusal.

Table 13. -- Items with response rates less than 75 percent

Teacher Demand and Shortage	Public	9b-9d, 9f, 10b-10d, 10f and 11b
	Private	3, 9b-9d, 9f, 10b-10d, 10f, 11b and 22d
Administrator Survey	Public	18
	Private	18
School Survey	Public	16b(13)-(14), 24a-24d, 31a-31e, 32b-32d
	Private	19b(13)-(14), 29e(1)-(2), 34a-34e, 35b(1-27), 35c(28) and 35d(1-27)
Teacher Survey	Public	14c and 28a(6)
	Private	14c and 28a(6)

On the public use tape, items 9 and 10 from both Public and Private Teacher Demand and Shortage Survey; item 32 from the Public School Survey; item 35 on the Private School Survey; and item 28 on both Public and Private Teacher Survey have been deleted because of data reporting problems.

A copy of an individual questionnaire can be obtained by looking at a NCES survey report or tape documentation from that questionnaire or requesting a copy from the NCES.

Variance Estimation

Each SASS public use file includes a set of replicate weights designed to produce balanced half-sampled replicated variance estimates. The balanced half-sampled technique was used because software to produce such variance estimates is relatively common. The formula for the variance of a statistic Y is given below.

$$\text{Variance } (Y) = 1/n \sum_{r} (Y_r - Y)^2$$

where: Y_r is the estimate of Y using the r^{th} set of replicate weights.

n is the number of replicates.

Below is a brief description of how the replicates were formed. The first step in this process was to exclude all nonresponding units from the process.

Replicates for Public Schools

The public school file was placed into replicates by first forming 48 variance strata within each state. Each variance strata contained at least two schools which were alternately divided into two half-samples. To form the variance strata, certainty schools were placed in their own variance stratum where each certainty school was assigned to each half-sample. The noncertainty schools within a state/school level sampling stratum were sorted by the school's order of selection. Pairs of schools were then systematically placed into consecutive variance strata. When the sample size within a sampling stratum was larger than 96 (48 X 2), contiguous variance strata were collapsed, so that the size of each variance stratum was as equal as possible, and until there were exactly 48 variance strata in the sampling stratum. Some variance strata, therefore, have more than two schools in them. The variance stratum numbering for the next sampling stratum started where the previous sampling stratum left off. When there was an odd number of noncertainty schools within a sampling stratum, one variance stratum was assigned an odd number of cases and adjustments were made to the replicate weights to account for this odd number of cases. After the variance strata were assigned, an orthogonal matrix was used to form the 48 replicates.

Private School Replicates

The sampling strata for private school were too numerous to use the method described for public schools. Instead of forming variance strata within sampling strata, it was done within collapsed strata. The collapsed strata were census region by school level by affiliation (13 groups).

For list frame and certainty area frame PSUs, the following was done to form variance stratum half-samples:

> Within each collapsed stratum, noncertainty schools were sorted by sampling stratum by order of selection. Pairs of schools were then consecutively placed into 48 variance strata, each element of a pair being assigned to different half-samples. If a collapsed stratum had an odd number of noncertainty schools then one of the variance strata had an odd number of schools and an adjustment was made to account for this. Certainty schools were assigned to each half-sample.

For noncertainty area frame PSUs, the following was done to form variance stratum half-samples:

> Within each PSU stratum, Westat had already paired the PSUs into half-samples. These pairings were consecutively assigned to variance strata, after sorting the pairs by the PSUs order of selection.

After the variance strata were assigned, an orthogonal matrix was used to form the 48 balanced half-sampled replicates.

Administrator Replicates

Since nonresponding schools were excluded from the replication process, using the school replicates for administrators presents a problem. Responding administrators without responding schools would not have an assignment in the school replicates. To resolve this problem, administrators were placed into replicates independently, using the same school algorithm. Instead of excluding nonresponding schools, nonresponding administrators were excluded.

Teacher Replicates

Since some responding teachers' schools may not have responded to the SASS, the school replicates described above are not appropriate for teachers. Such teachers were not assigned to a variance stratum. For this reason, all responding SASS teachers were placed into replicates independently from the school

replicates. The first step in forming teacher replicates is to create a single school record when at least one teacher in a noncertainty school responded to SASS. These records were then placed into variance strata using the school variance strata algorithm described above. The responding teachers within a school were then assigned to the school's variance stratum half-sample. Teachers from certainty schools were placed into the same variance stratum, but were split into different half-samples. After the variance strata were assigned, an orthogonal matrix was used to form the 48 balanced half-sampled replicates.

LEA Replicates

To reflect the fact that LEAs were selected through the school, its important to form LEA replicates using the school replicate procedure. Since some LEAs may not have any responding schools, the school replicates are not appropriate for the LEA. Such LEAs were not assigned to a school variance stratum. The LEAs, therefore, were independently placed into replicates, using the school replicate procedures. First, a school record was formed for each responding school within a responding noncertainty LEA. Then, these records were placed into replicates using the school replicate algorithm. For each of these school replicates, an LEA was placed into the corresponding LEA replicate if any of the schools associated with the LEA were in that particular school replicate. Certainty LEAs were placed into all replicates.

LEAs without schools were sorted by stratum and order of selection. Pairs of LEAs were then systematically placed into consecutive variance strata and each element of a variance strata were assigned to alternating half-samples. After the variance strata were assigned, an orthogonal matrix was used to form the 48 replicates.

Replicate Weights

For school, administrator and teacher replicates, the nonzero replicate weights were either: 1) twice the final sampling weight for noncertainty sample units; or 2) the final SASS weight for certainty sample units.

The noncertainty LEAs' replicate weights were assigned using the same LEA weight formula described in the weighting section, the only difference being that, for the basic weight calculation, each school stratum's selection probability for the LEA was divided by two. This is appropriate since the half-sample LEA selection probability within a school stratum is half of the respective full sample probability. Certainty LEA's replicate weights were the SASS final weight.

Cautions

Replicated variance estimates assume sampling is done with replacement. For SASS, this was not the case. Unless the sampling rate is extremely high, the variance estimates should only be a slight overestimate. None of the public school sampling rates were high enough to provide a large variance overestimate. However, for some of the small private school affiliation strata, the sampling rates were high enough to possibly produce a large variance overestimate. For this reason, consideration was given to adjusting the private school replicate weights to better estimate the variance. The adjustment considered was the one appropriate for simple random sampling (i.e., 1-n/N where n and N are the sample and universe sizes, respectively).

An analysis was done to see if such an adjustment would improve the variance estimates. It was determined that it was unlikely that any private school estimate would be greatly reduced by an appropriate finite population correction. For more information on the analysis, see appendix 3.

Frame Evaluation

When the first SASS estimates were produced inconsistencies were noticed between the SASS and the Common Core of Data (CCD) estimates. Some of these inconsistencies were caused from differences between the QED file (SASS frame) and the CCD file (the NCES's file of public schools and LEAs). These differences are discussed below.

School Estimates

1) The definition of a school is different between the QED and CCD. QED defines a school as a physical location, while the CCD defines the school as an administrative unit. If two schools with separate principals reside in the same building, then QED (i.e., SASS) defines this as one school. CCD defines this as two schools. This definitional difference is most noticeable in Nebraska, where the CCD count was 44 percent larger than the SASS estimate. In North Dakota, the CCD count was 42 percent larger than the SASS estimate; in South Dakota, CCD was 38 percent larger than the SASS estimate; and in Montana, CCD was 15 percent larger than the SASS estimate. These are the states with the largest differences. Other states may be affected by this definitional difference, but to a much smaller extent. On a national basis there are 6 percent more CCD schools than SASS schools.

2) For private schools, the SASS estimate of total number of schools is smaller than the number of schools represented on the frame. The differences are due to a combination of frame schools being out of scope, no longer in existence or duplicated on the frame, as discovered after the SASS collection. Table 14 provides a comparison.

Table 14. -- Comparison between private school frame counts with SASS estimates of number of schools

Source	Number of schools on frame	Estimated number of schools from SASS
Total	31,848	26,807
List frame excluding association add-ons	22,600	19,884
Association add-ons	1,586	1,018
Area frame (weighted count)	7,662	5,905

LEA and School Estimates

3) In Nebraska, QED excluded small (elementary grade only) LEAs. A match with CCD revealed 275 such LEAs were missing from the QED file. The missing LEAs represent 2800 students with an average of 10.2 students per LEA. Since these LEAs are excluded from the SASS frame, they are not represented in SASS estimates. The schools associated with these LEAs are also not represented in SASS school estimates. These missing schools are contributing to the 44 percent QED underestimate for Nebraska stated above.

LEA Estimates

4) In Montana, QED collapsed LEAs with the same mailing address into one LEA. Usually, this was a single elementary school being collapsed with a high school. In Montana, CCD had 45 percent more LEAs than QED. All of this difference can be explained by the collapsing. It is unknown if other states are affected by this collapsing procedure.

5) In Louisiana, Arizona and Virginia, the CCD number of LEAs was approximately 20 percent larger than QED. This was caused by some of the sampled QED LEAs being out of scope (i.e., LEAs that don't hire teachers).

Teacher Estimates

6) In some states, the FTE or full-time equivalent (i.e., one full-time equivalent is equal to the amount of time a person would normally spend serving full-time on an assignment) teacher counts between the school and teacher file were not consistent. Part of the inconsistency can be attributed to two sources:

a) The number of teacher head counts reported on the school file, on average, is higher than the number of teachers on the teacher frame given to Census by the school. In the average state, there were 5 percent fewer teachers on the teacher frame. This would cause head count or FTE estimates from the teacher file to be underestimates if the school did not include all teachers on the teacher frame.

b) We believe that schools had problems providing FTE counts because it was not unusual for a school to report the same number of teacher head counts as FTE teachers when the school reported some part-time teachers. Given the way the questions were worded, it's possible this is alright, but given its frequency, it's unlikely. In an average state, 19 percent of the schools reported the same number of head counts as FTEs when some part-time teachers were reported. Since this seems to indicate that the reported FTE counts are higher than they really are SASS FTE estimates from the school file are likely to be overestimates.

Appendices

Appendix 1

Minimizing the School Overlap between the NAEP and NELS Samples

This appendix describes how the original SASS selection probabilities, if no unduplication was required, were adjusted so that the number of overlap schools between SASS and NAEP or NELS samples is minimized without changing the overall probability of selection (i.e., the SASS school probability of selection averaged over all possible NAEP, NELS and SASS samples should equal the original selection probability). To do this required knowledge about the NAEP and NELS selection probabilities. With this information for each QED school, the probability of not selecting a school in NAEP or NELS can be determined. If the SASS original selection probability divided by the probability of not selecting the school in NAEP or NELS is less than one, and the school was selected in NAEP or NELS then the school was eliminated from the frame. If the SASS original selection probability divided by the probability of not selecting the school in NAEP or NELS is greater than or equal to one, and the school was selected in NAEP or NELS then the school remained on the frame. However, the selection probability was reduced to minimize its chance of selection. Other QED schools not selected in NAEP or NELS had their selection probabilities adjusted to account for the unduplication.

Since the overall probability of selection was the original school selection probability, the basic weights (weights without adjustment for nonresponse) are the reciprocal of the original school selection probability. There is no need to use the adjusted selection probabilities when weighting the school sample.

The details of this process are described below. First, required terminology and sets of schools will be defined. Next, the definition of the adjusted probabilities of selection (conditional selection probabilities) will be defined. Selecting the SASS sample with these conditional selection probabilities will maintain the original selection probabilities over all possible NEAP, NELS and SASS school samples, while minimizing the overlap.

Terminology

A - NAEP sample

E - NELS sample

S^2 - SASS sample

i - school

F_s - schools in the S^2 frame

F_{ae} - schools in the A or E frames

NS - not selected

S - selected

P_{si} - original S^2 selection probability for school i without regard to the NAEP and NELS samples

P_{ai} - is NAEP selection probability for school i

$P_i(E \mid i \text{ NS in A})$ - is NELS selection probability for school i given that school i is not selected in the NAEP sample

$P_i(NS)$ - probability of school i not being selected in the A or E samples

Important Sets of Schools

$W = \{i \in F_s \mid i \text{ does not match any school in } F_{ae}\}$

$Z = \{i \in F_s \mid i \text{ matches a school in } F_{ae}\}$

$X = \{i \in Z \mid P_{si}/P(NS) < 1\}$

$Y = \{i \in Z \mid P_{si}/P(NS) \geq 1\}$

Conditional Selection Probabilities

In order to compute the SASS conditional selection probabilities $P_i(NS)$ must be computed first. To do this, the NAEP and NELS selection probabilities (P_{ai} and $P_i(E \mid i \text{ NS in A})$) must be obtained from the NAEP and NELS contractors. Given these probabilities $P_i(NS)$ can be computed as follows:

$$P_i(NS) = P(\text{ i NS in A or E})$$
$$= P_i(\text{NS in A}) * P_i(\text{ NS in E} \mid \text{ i NS in A})$$
$$= (1-P_{ai}) * (1-P_i(E \mid \text{ i NS in A}))$$

This probability will be used to compute the SASS conditional selection probabilities described next.

If $i \in W$ then

$$P_i(S \text{ in } S^2 \mid A \text{ and } E) = P_{si}$$

If $i \in X$ then

$$P_i(S \text{ in } S^2 \mid A \text{ and } E) = \begin{cases} P_{si}/P_i(NS) & \text{if i NS in A or E} \\ 0 & \text{otherwise} \end{cases}$$

If $i \in Y$ then

$$P_i(S \text{ in } S^2 \mid A \text{ and } E) = \begin{cases} 1 & \text{if i NS in A or E} \\ (P_{si}-P_i(NS))/(1-P_i(NS)) & \text{if i S in A or E} \end{cases}$$

The SASS sample was selected using these conditional selection probabilities. It is easy to verify that these probabilities will preserve the initial selection probabilities (P_{si}) while minimizing the NAEP and NELS sample overlap.

Appendix 2

An Allocation of Schools and Teachers which Minimizes Loss of Relative Precision for Specific Estimators

Background

Estimates of school and teacher characteristics are obtained via a stratified design. Stratification variables consist of states x school level x sector. There are 51 states (including the D.C.), three levels (elementary, secondary and combined schools) and two sectors (public, private) indexed by i,j, and k respectively.

Four classes of school characteristics and four classes of teacher characteristics are of interest. Specifically, these are the U.S. level, state comparisons, public v.s. private comparisons, and elementary v.s. secondary v.s. combined comparisons. Each of these estimates can be put into the same general form. Denote a specific school and teacher characteristic by "x" and "y", respectively. Denote strata means by x_{ijk} and y_{ijk}. Denote school and teacher stratum weights by W_{ijk} and U_{ijk}, respectively. An estimate of the following classes looks like:

$$X = \Sigma\Sigma\Sigma_{ijk} W_{ijk} x_{ijk} \text{ for schools}$$

and $$Y = \Sigma\Sigma\Sigma_{ijk} U_{ijk} y_{ijk} \text{ for teachers}$$

For example, to estimate the U.S. total take

W_{ijk} = # schools in strata i,j,k.

U_{ijk} = # teachers in strata i,j,k.

Similarly, a State average would be based on

$$W_{ijk} = \frac{\text{\# schools in strata ijk}}{\text{\# schools in State i.}}$$

$$U_{ijk} = \frac{\text{\# teachers in strata ijk}}{\text{\# teachers in State i.}}$$

Denote S_{ijk} = # schools in strata ijk and

T_{ijk} = # teachers in strata ijk

The specific weights used are in the following table of strata weights:

		Schools		Teachers	
		W_{ijk}		U_{ijk}	
Class	C	Average	Total	Average	Total
U.S.	1	S_{ijk}/S_{+++}	S_{ijk}	T_{ijk}/T_{+++}	T_{ijk}
State:i	2	S_{ijk}/S_{i++}	S_{ijk}	T_{ijk}/T_{i++}	T_{ijk}
Level:j	3	S_{ijk}/S_{+j+}	S_{ijk}	T_{ijk}/T_{+j+}	T_{ijk}
Sector:k	4	S_{ijk}/S_{++k}	S_{ijk}	T_{ijk}/T_{++k}	T_{ijk}

When referring to estimates of a specific class, "c"; the corresponding weights will be denoted by W_{cijk} and U_{cijk}

Variances

$$V(X) = \Sigma\Sigma\Sigma_{ijk} W^2_{ijk} V(X_{ijk})$$

and

$$V(y) = \Sigma\Sigma\Sigma_{ijk} U^2_{ijk} V(Y_{ijk})$$

$V(x_{ijk})$ is approximated by: (Assuming variances of measurements with the same level and sector are equal from state to state)

$$(1-n_{ijk}/N_{ijk})\ V_{jk}/n_{ijk}$$

$V(Y_{ijk})$ is approximated by:

$$(1-n_{ijk}/N_{ijk})\ Q_{1jk}/n_{ijk} + (1-m_{jk}/M_{jk})\ Q_{2jk}/(n_{ijk}\ m_{jk})$$

Where n_{ijk} and N_{ijk} denote, respectively, the number of schools sampled and the total number of schools in strata ijk and m_{jk} and M_{jk} denote, respectively, the sampled number of teachers/school and the average number of total teachers/school in strata j,k (assumed equal for states).

An estimate of V_{jk} was obtained as follows:

To approximate the actual design, a variance term of the following form is needed:

$$V_{jk} = DEFF_{jk}\ S^2_{jk},$$

where: S^{2jk} is the population variance.

$DEFF_{jk} = V_{jk}/s^2_{jk}$ (the variance under the desired design divided by the variance under simple random sampling)

Since (omitting subscripts)

$$CV = SE(EST)/EST$$ (the standard error of an estimate divided by the estimate)

for proportions: $SE(EST) = DEFF\ S/n$,
an estimate of S^2 is:

$$S^2 = CV^2 \times n \times EST^2/DEFF$$

and so,

$$V_{jk} = n_{jk} CV^2_{jk}\ EST^2_{jk}$$

The calculations for totals are the same since estimates of totals consist of weighted means. From simulations, CV, DEFF and EST are obtained for the school variables.

The between and within teacher variances Q^2_{1jk} and Q_{2jk} were simply the variances obtained from Census which estimate the population variance of a pps sample with replacement.

Separate Allocations For Each Estimate

The eight variances to be minimized for a fixed cost are:

$$V(X_c) = \Sigma\Sigma\Sigma_{ijk}\ W^2_{cijk}\ V(x_{ijk})$$

$$V(Y_c) = \Sigma\Sigma\Sigma_{ijk}\ U^2_{ijk}\ V(y_{ijk})$$

denote: CS_k as the cost to sample a school in sector k, C_{public} includes the LEA cost and $C_{private}$ is just the school cost, but with an allowance to cover response to LEA questions.

CT as the marginal teacher cost (the cost to sample a single teacher once the school is in sample).

GC the total variable cost for the survey

The problem is to choose n_{ijk} schools and m_{jk} teachers/school in stratum ijk to minimize $V(X_c)$ or $V(Y_c)$ subject to the cost constraint:

$$\Sigma\Sigma\Sigma_{ijk}\ (CS_k\ n_{ijk} + CTn_{ijk}\ m_{jk}) = GC = \text{Total Cost}$$

Closed form solutions to obtain an allocation which minimizes only one variance at a time can be obtained by using Lagrange multipliers to incorporate the cost restriction (See Cochran,

Chapter 10 for example)

a) separate school estimates: For an estimate of class "c", the optimal allocation is:

$$n_{ijk} = \frac{GC}{(C_k + CT)} \cdot \frac{(W^2_{ijk} V_{jk} (C_k + CT))^{1/2}}{\sum_{ijk} (W^2_{ijk} V_{jk} (C_k + CT))^{1/2}}$$

and, of course, $m_{jk} = 1$, since teachers characteristics are of no interest here.

b) separate teacher estimates:

For an estimate of class "c", the optimal allocation is:

$$m^2_{jk} = \frac{CS_k}{CT} \cdot \frac{Q_{2jk}}{(Q_{1jk} - Q_{2jk}/M_{jk})}$$

, a trade-off of between/within costs and variances

$$n_{ijk} = \frac{GC}{(C_k + m_{jk} CT)} \times \frac{(W^2_{ijk} (Q_{1jk} + 1 - m_{jk}/M_{jk}) Q^2_{2jk} (C_k + m_{jk} CT))^{1/2}}{\sum\sum\sum_{ijk} (W^2_{ijk} (Q_{1jk} + 1 - m_{jk}/M_{jk}) Q^2_{2jk} (C_k + m_{jk} CT))^{1/2}}$$

Compromise Allocation

The idea is to pick a single allocation so that the resulting increase in variance for each estimate is minimized on average. This is the final allocation used. This idea is quantified as follows.

Obtain the minimum variance under the optimal allocation specific to each estimate, as if that was the only estimate needed. The separate allocations needed to obtain these allocations have been outlined above in the section titled "Separate Allocations for Each Estimate". Call these variances $VO(X_1)...,VO(X_c)$, $VO(Y_1)...$ $VO(Y_c)$.

Choose a final allocation which minimizes the sum of the variance increases, i.e., minimize:

$$\frac{V(X_1)}{VO(X_1)} + \frac{V(X_2)}{VO(X_2)} + \ldots + \frac{V(X_c)}{VO(X_c)} + \frac{V(Y_1)}{VO(Y_1)} + \ldots + \frac{V(Y_c)}{VO(Y_c)}$$

This can be rewritten as:

$$\Sigma\Sigma\Sigma_{ijk} \; (SH^2_{ijk} \; V(x_{ijk}) + TH^2_{ijk} \; V(y_{ijk})$$

where

$$SH^2_{ijk} = \frac{W^2_{1ijk}}{VO(X_1)} + \ldots + \frac{W^2_{cijk}}{VO(X_c)}$$

and

$$TH^2_{ijk} = \frac{U^2_{1ijk}}{VO(Y_1)} + \ldots + \frac{U^2_{cijk}}{VO(Y_c)}$$

which is a linear combination of the strata variances.

A closed form solution isn't possible in this case. An iterative solution is employed which uses the number of teachers/school obtained for the separate allocations and updates it.

The update is performed as follows:

STEP 1) obtain initial estimate of m_{jk} and construct teacher fpc and relative state allocations.

STEP 2) minimize variance function with respect to n_{ijk} and m_{jk} given the teacher fpc and relative State allocation.

STEP 3) Based on the new set of m_{jk} which minimize the variance function, obtain a new teacher fpc and relative State allocations.

STEP 4) go to step 2) and repeat until convergence.

Teacher Comparisons

To account for an interest in particular fields of teachers, the following adjustment was made. For "f" fields of teaching, the increase in variance (if 1/f of the teachers are in each field) can be summarized as:

$$V(y_{ijk}) = (1-n_{ijk}/N_{ijk})\ Q_{1jk}/n_{ijk} +$$

$$(1-m_{jk}/M_{jk})\ Q_{2jk}\ f/(n_{ijk}\ m_{jk})$$

Further, the components of the variance related to teachers in secondary or combined schools was multiplied by f to reflect the fact that field of study estimates for this group may be needed. In other words, when constructing TH, we actually used:

$$TH^2_{ijk} = \frac{V^2_{1ijk}}{VO(Y_1)} + \frac{V^2_{2ijk}}{VO(Y_2)} + \frac{f \times V^2_{3ijk}}{VO(Y_3)} + \frac{V^2_{4ijk}}{VO(Y_4)}$$

when j = (secondary or combined).

Variances

Based on the optimal allocation, variances can be computed. Along with variances related to the estimates mentioned, LEA variances are calculated.

Appendix 3

Effect of a Finite Population Correction on SASS Variance Estimates

Replicated variance estimates assume sampling is done with replacement. For SASS, this was not the case. Unless the sampling rate is extremely high, the variance estimates should only be a slight overestimate. None of the public school sampling rates were high enough to provide a large variance overestimate. However, for some of the small private school affiliation strata, the sampling rates were high enough to possibly produce a large variance overestimate. For this reason, consideration was given to adjusting the private school replicate weights to better estimate the variance. The adjustment considered was the one appropriate for simple random sampling (i.e., 1-n/N, where n and N are the sample and universe sizes, respectively).

Some simple simulations were performed to measure the effect of such an adjustment when the sample was selected probability proportionate to size. First, a school frame was assumed. Then, all possible school samples were generated and an estimate of total schools was produced for each sample. In each sample, 66 percent of the schools were selected. From this, the true standard error was computed.

Next, all possible replicate standard error estimates were computed for each sample. These standard error estimates were averaged to produce an average replicated standard error estimate.

Finally, the average replicated estimate was adjusted appropriately to reflect a simple random sample finite population correction factor (adjusted replicated estimate). With this information, it was possible to measure the error in both the unadjusted and adjusted replicated standard error estimates.

The three tables below show the results:

Table 15. -- Effect of finite population correction (fpc) where the distribution of probabilities is unequal and skewed

School frame			
Elementary		Secondary	
School	Probability	School	Probability
1	0.9	4	0.9
2	0.9	5	0.9
3	0.2	6	0.2

Two out of three schools were selected within each stratum using a probability proportionate to size sampling scheme. The probability column above provides the measures of size.

From the set of all possible samples, the following numbers can be computed.

(a) True std error	(b) Average replicate std error	Ratio of b to a
2.199887	2.459549	1.118033

Unadjusted, the replicated standard error overestimates the true standard error by 12 percent. If the replicated standard error is adjusted by the square root of (1-n/N), where n and N are the sample and universe sizes, respectively then the following numbers are obtained.

(c) fpc adjusted replicate std error	Ratio of c to a
1.420021	0.645497

Now, the adjusted replicated standard error is underestimating the standard error by 35 percent.

Table 16. -- Effect of finite population correction (fpc) where the distribution of probabilities is unequal and not skewed

School frame			
Elementary		Secondary	
School	Probability	School	Probability
1	0.5	4	0.5
2	0.7	5	0.7
3	0.8	6	0.8

Two out of three schools were selected within each stratum using a probability proportionate to size sampling scheme. The probability column above provides the measures of size.

From the set of all possible samples, the following numbers can be computed.

(a) True std error	(b) Average replicate std error	Ratio of b to a
0.462910	0.707106	1.527525

Unadjusted, the replicated standard error overestimates the true standard error by 53 percent. If the replicated standard error is adjusted by the square root of (1-n/N), where n and N are the sample and universe sizes, respectively then the following numbers can be computed.

(c) fpc adjusted replicate std error	Ratio of c to a
0.408248	0.881917

Now, the adjusted replicated standard error is underestimating the true standard error by 12 percent.

Table 17. -- Effect of finite population correction (fpc) where the distribution of probabilities is almost equal and not skewed

School frame			
Elementary		Secondary	
School	Probability	School	Probability
1	0.64	4	0.64
2	0.67	5	0.67
3	0.69	6	0.69

Two out of three schools were selected within each stratum using a probability proportionate to size sampling scheme. The probability column above provides the measures of size.

For the set of all possible samples, the following numbers can be computed.

(a)	(b)	
True std error	Average replicate std error	Ratio of b to a
0.066387	0.113328	1.707076

Unadjusted, the replicated standard error overestimates the true standard error by 71 percent. If the replicated standard error is adjusted by the square root of (1-n/N), where n and N are the sample and universe sizes, respectively then the following numbers can be computed.

(c)	
fpc adjusted replicate std error	Ratio of c to a
0.065430	0.985581

Now, the adjusted replicated standard error is underestimating the standard error by 1 percent.

The tables show that the adjusted standard errors underestimate the standard error, by as much as 35.5 percent (1-.645), if the selection probabilities are unequal and skewed. Since the underestimate can be so large, it's probably unwise applying an fpc adjustment.

Another argument against applying an fpc adjustment is the effect the area frame sampling rate has on the overall sampling rate. The 551 area sample schools represents 24 percent of the total number of private schools. These schools come from a sample of 75 out of 2,497 PSUs, a sampling rate of 3 percent. On average, one might expect that 24 percent of an affiliation's estimate would come from the area sample, with a 3 percent sampling rate. This should significantly reduce the overall sampling rate of affiliations with high list frame sampling rates. A similar situation arises if schools from other list frame strata with low sampling rates, report belonging to an affiliation with a high list frame sampling rate. Table 18 below shows the impact of the area sample on the affiliation estimates with the highest list frame sampling rates.

The table shows that except for Friends, military schools and Christian international schools, a significant part of the standard error comes from the area sample, which has a low overall sampling rate. Therefore, adjusting for a high list frame sampling rate should have minimal impact on the total variance. The high variance for Friends, military schools and Christian international schools is caused by schools from other list frame strata reporting they belong to Friends, military schools or Christian international schools. The variance contribution from the Friends and Military schools list frame stratum is zero since they were all selected with certainty. The variance is solely coming from schools from other list strata that make up part of the estimate.

Because the overall affiliation sampling rate is greatly reduced by the low sampling rate of the area frame and some other list strata, and the fact that the proposed adjustment would produce a possibly large underestimate of the variance, the variance replicates were not adjusted for the high sampling rates. It is unlikely that any private school estimate would be greatly reduced by an appropriate finite population correction.

Table 18. -- Estimate of number of schools by list and area stratum

School	List frame		Area frame	
	Estimate	Standard error	Estimate	Standard error
Friends	73	11	2	2
Military schools	51	19	0	0
Exceptional children	199	44	109	52
Christian international	308	49	3	3
Episcopal	319	31	27	5
Montessori	467	54	216	106
Jewish	514	23	90	52

United States
Department of Education
Washington, D.C. 20208–5651

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

FOURTH CLASS BOOK RATE