1993-94 STUDENT RECORDS SURVEY:SAMPLING AND WEIGHTING CONUNDRA

Randall J. Parmer, Robert C. Abramson, Cornette L. Cole, Lenore A. Colaciello, B. Dale Garrett
U. S. Bureau of the Census, Washington, DC 20233

KEY WORDS: Nonresponse, Mean-Squared Error, Truncation

This paper reports the general results of research undertaken by Census Bureau Staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau.

I. Introduction

The sample for the Student Records component of the Schools and Staffing Surveys (SASS) was selected through a complex four-stage sampling procedure. Three of the four stages of selection relied heavily upon information supplied by the respondent, with the actual selection also being carried out by the respondent. Consequently, some of the information needed to select sample students was not reported or was reported inaccurately. These reporting problems introduced considerable difficulty in the computation of probabilities of selection.

Additional difficulties were encountered during the weighting and as a result of computing initial variances. National estimates of students by race turned out to have extremely high variances. Strategies for dealing with this problem are presented in this paper.

II. Background

A. General Goals

The National Center for Education Statistics (NCES) sponsors the Schools and Staffing Survey in order to provide periodic, timely data on public and private schools, school districts, administrators and teachers. A student component was added to SASS for the first time in 1993-94.

The goal of the student component is to examine the quality of teachers through their students and analyze student characteristics, participation in special programs, and achievement. Data is collected with mailout/mailback of questionnaires and with telephone follow-up of mail nonrespondents.

Feasibility studies were conducted in 1991 and again in 1992-93 to determine the willingness of the school to provide certain data about their students and to test the ability of the school to follow a complex set of sampling instructions. Many problems were discovered in these earlier studies and many lessons were learned. See King and Kaufman (1994) for further discussion of the 1992-93 feasibility study. This paper delves into the general methodological lessons that were learned from conducting the 1993-94 survey.

B. Sampling

1. School Selection

The first stage of sampling for the Student Records Survey was school sampling. Generally, SASS sample schools were subsampled for the student sample. Of the 9134 SASS sample public schools outside Alaska, 551 were selected for the student sample. Schools were stratified by grade level and urbanicity, then sorted by SASS stratum, order of selection and school ID. Schools were selected with probability proportional to the SASS stratum's sampling interval.

All 444 schools with more than 19.5% American Indian enrollment, all 199 Alaska schools, and all 176 Bureau of Indian Affairs (BIA) schools from SASS were selected for the student survey with certainty.

From among the 3315 private SASS sample schools, 381 were chosen for the student sample. Schools were stratified by affiliation and grade level, then sorted by frame and enrollment. Schools were selected with probability proportional to size, similarly to public schools.

2. Teacher Sampling

For each SASS sample school, from 1 to 20 sample teachers had been selected, depending upon the size and make-up of the teaching staff. If the school had more than three sample teachers for SASS, three teachers were systematically selected for the student survey. If the school had three or fewer sample teachers, all teachers were selected for the student survey.

3. Class Period Sampling

Schools were contacted by telephone to verify that sample teachers were eligible for SASS and asked if the teacher teaches self-contained or departmentalized classes. If self-contained, no class period selection was necessary. If departmentalized, the school was asked how many class periods per week were in the school's schedule. We proceeded to select five class periods (one on each day of the week) and asked if each teacher taught any those class periods. If so, we randomly selected one of the class periods that the teacher was teaching. If the teacher wasn't teaching any of those class periods, another set of five class periods was selected and inquiry was made about the teacher's status for those five periods. If the teacher still wasn't teaching any of those five, we asked for the teacher's weekly schedule and randomly selected one of the class periods the teacher was teaching. Thus, ultimately one class period was selected for each sample teacher teaching departmentalized instruction.

4. Student Sampling

For each sample class period from a sample teacher, a

class roster was requested. For each class roster, two students were selected systematically for the student survey.

C. Weighting

1. Basic Weight

The student's basic weight is the inverse of the probability of selection conditioned on the specific set of sample teachers selected for the student sample at the school. The sum of the inverse of conditional probabilities for all sample students in the school are ratio adjusted to the school's enrollment to account for all possible teacher samples, a probability we cannot calculate since we don't know each sample student's entire weekly schedule. Thus, the basic weight is an approximation. The basic weight is expressed as:

$$W_{ki} = \frac{1}{P_{ki}} X \frac{school\ enrollment}{\sum_{i=1}^{6} \frac{1}{P_{ki}}} X\ W_k X\ F_{ki}$$

where:

W_k = basic weight for school K.

F_{ki} = school student subsampling factor.

where:

The students probability of selection is the sum of the probabilities of selecting the student from the teachers (of the three sample teachers at the school) that teach the student.

$$P_{ki} = \sum_{j=1}^{3} P_{kji}$$

and:

P_{kji} = 0 if the j^{th} teacher does not teach student i, or equal to the result of one of the two equations defined below, depending upon whether the j^{th} teacher is departmental or self-contained. The definitions for the variables used to calculate the probability (P_{kji}) for students with departmental teachers are defined as follows:

N_{kji} = the total number of times, within school k, that student i has teacher j each week.

L_{kj} = the total number of periods the sample teacher teaches an eligible class at the sample school per week.

Tp_{kj} = the teacher probability of selection for the student sample adjusted for teachers erroneously classified as not teaching regularly scheduled classes.

S_{kj} = size (enrollment) of the sample class period.

The probability of selecting the i^{th} student from the j^{th} teacher at a school k was dependent upon the probability of selecting the sample class period from the total class periods at school k (if the teacher is classified as departmental), the probability of selecting the teacher from school k, and the probability of selecting the student from the teacher's sample class period.

For students selected from departmental teachers, the formula below was used.

where:

$$P_{kji} = \left[\frac{N_{kji}}{L_{kj}}\right] \cdot \left[\frac{2}{S_{kj}}\right] \cdot TP_{kj}$$

The variables are as defined above.

For students from self-contained teachers, the formula below was used.

$$P_{kji} = \left[\frac{2}{S_{kj}}\right] \cdot TP_{kj}$$

2. Other Factors

Various other factors are applied as part of the weighting process. Most of these we won't discuss in detail here since they are not pertinent to subsequent discussion:

School Nonresponse Adjustment Factor -
accounts for schools that did not participate in either the teacher or student sampling procedures.

First Stage Ratio Adjustment Factor -
adjusts frame counts of enrollment for sample schools to the known frame total enrollment.

Misclassified Teacher Adjustment Factor -
adjusts for teachers reported to not be teaching during student sampling but later reported as teaching in the teacher survey.

Student Noninterview Adjustment Factor -
accounts for sample students for whom their schools did not return questionnaires.

3. Student Adjustment Factor

The Student Adjustment Factor is discussed in greater detail here since it is discussed in detail later in this paper.

The Student Adjustment Factor adjusts for the inconsistency between the estimated enrollment on the school data file and the student sample file. It is computed as the ratio of the weighted number of students from the school data file to the weighted number of

students on the student data file. Factors are computed separately for each cell. The student weight used is the product of all components described in Sections II.C.1. and II.C.2.

Public schools' cells (including BIA) were defined by grade level (elementary, secondary, combined) by enrollment (three categories within each grade level) by race/ethnicity (American Indian or Alaskan Native, Hispanic, Black, white, Asian or Pacific Islander).

Private school cells are defined by affiliation (Catholic, other religious, nonsectarian) by grade level (elementary, secondary, combined) by race/ethnicity (American Indian or Alaskan Native, Hispanic, Black, White, Asian or Pacific Islander). Cells were collapsed if the computed factor was less than 2/3 or greater than 3/2, or if there were less than 15 students in the cell. Collapsing continued until these criteria were satisfied for all collapsed cells.

Public schools were collapsed across race/ethnicity first, then enrollment category, and finally grade level. Private schools were collapsed across race/ethnicity first, then grade level, and finally affiliation.

After collapsing was completed, factors were applied to each interviewed record within a cell.

III. PROBLEMS ENCOUNTERED DURING SAMPLING

Several problems were encountered during the sampling, and many lessons were learned. In the list below, we've attempted to describe the problems,

A. Missing Sampling Information

Some of the information we needed to perform the student weighting procedure were missing for a large proportion of sample students' records. We resolved this problem by several means:

- First, in a clerical operation, we were able to locate some of the missing information from student sampling worksheets. The information either had not been keyed or had been keyed incorrectly on the student data file.
- If a student with missing data had the same teacher or was from the same school as another sample student whose records contained the information we needed, we copied it to the other student's record.
- As a last resort, we filled the missing fields through imputation. Imputation rates generally varied from 1 to 6 percent.

B. Schools Refused To Cooperate With Sampling

Many schools were reluctant to provide us with student names and associated information by telephone. Our only alternative for obtaining an interview for these cases was to make personal visits to the schools. Because of the expense, we were unable to visit some schools, such as those located in remote parts of Alaska and in some areas of California. For these schools, we made an additional attempt to interview them by telephone.

School response rates generally varied from, 80 to 95 percent, with BIA and American Indian schools having higher response rates than other public or private schools. These response rates are for schools that participated in the Student Survey. Numerators were weighted counts of student subsample schools that participated in the student survey by completing any of the six sample students' questionnaires. These results are not indicators of how many students were interviewed, but of how many schools participated in the student survey.

C. Schools Didn't Complete the Survey Questionnaires

A number of Schools agreed to participate in the student sampling operation, but then didn't complete the questionnaires that were sent to them.

Questionnaire response rates generally range from 80 to 95 percent.

The response rates indicate the proportion of eligible student records that were considered to be completed interviews. Eligible counts are the total number of students that were selected for interview and interviewed counts are of those eligible students whose records were completed and returned.

D. Problems With the Sampling Instructions

- We found unrealistic values for some variables on some student records, suggesting that some schools may have misreported sampling information.
- We suspect some respondent schools did not follow our sampling instructions. The sampling instructions may have been too complicated or too time consuming to be understood over the telephone. For example, a respondent school could be required to go through three different sets of class periods for up to three teachers in order to identify one eligible class period per teacher.
- The instructions for selecting sample class periods were difficult to apply to schools with unusual schedules.

E. Duplicate Students

If a student was selected for more than one of a school's sample teachers, instructions were to place an "M", to denote 'MULTIPLE', in a field on the student sampling worksheet. This field had been filled per instruction in many cases, however, the information was never keyed. We were forced to identify duplicate student records through a tedious clerical operation.

F. Timing Problems

Some school schedules conflicted with our sampling schedule. Census Bureau personnel in Jeffersonville had difficulty contacting some schools because the schools were closed for holidays or vacation during the time period we designated for sampling.

G. Teachers Did Not Match School

In a few cases, a school was called and information was requested for a particular teacher, but we were told the teacher was not employed by the school we had telephoned. After investigating, we found that the teacher taught at another school and that the mix-ups were between private and public schools with similar names. The teachers we were trying to locate were public school teachers, but the telephone numbers we had called were for private schools.

H. Some Sample Teachers Were Erroneously Classified as Out-of Scope

During the student sampling operation, some sample teachers were reported to us as not teaching or no longer teaching. These teachers were classified as out-of-scope and no student sampling was conducted for them. Many of these teachers were subsequently discovered to be valid teachers during the teacher survey.

Out-of-scope rates for public and private teachers ranged from 1 to 6 percent.

Teachers considered as any one of the following were classified as out-of-scope:

- Sort-term substitutes
- Student teachers
- Non-teaching specialists: Librarians, nurses, guidance counselors, administrators
- Teacher's aides
- Support staff: Cooks, custodians, Bus drivers, dieticians, secretaries

Teachers were also classified as out-of-scope if the associated sample school had been classified as out-of-scope.

I. Number of Classes Students Take was not Asked

Because we did not ask how many classes each student was taking, we were forced to assume all students in a school took the same number of classes, thus making the weighting biased.

This issue is discussed further in Section IV.B.

IV. Weighting Issues

A. Changes to Probability of Selection

As described in King and Kaufman (1994), the probability of selection of a class period was the sum of a hypergeometric random variable. Later, we noticed that this hypergeometric could be simplified to $1/L_{kj}$, where L_{kj} is the number of class periods per week that the teacher teaches. Intuitively, this makes sense since each of the L_{kj} class periods that the teacher teaches in a week have an equal probability of selection.

B. Identifiable Bias in Probability of Selection

As mentioned in III.I., the estimator used for the student records survey assumes that all students within a school have the same number of class periods per week. We know this assumption is somewhat faulty, but unfortunately the necessary information about each sample student was not collected. Thus, in order to approximate the correct probability of selection, the sum of the weights of sample students were controlled to the school's total enrollment.

V. Variance Issue

A. The Problem

One of the primary purposes of the survey was to measure participation in special programs for American Indian students. The coefficient of variation (CV) in public schools was found to be extremely large for total American Indian students - above 50%. This was deemed to be unacceptably high. In an effort to reduce the variance for these estimates, several changes to the weighting procedure were proposed, including truncating the weights or Windsorizing, changing the collapsing criteria, and changing the order of cell collapsing. These last two options applied to the final step in the weighting procedure which is a ratio adjustment of the student survey totals to the school survey total enrollment - The Student Adjustment Factor.

Ten alternative weighting schemes were devised and implemented to compare their effectiveness in reducing the size of the problem. The ten alternatives are as listed in Table 1.

B. Alternative Schemes for Solving the Problem

As can be seen from Table 1, we tested three alternative schemes for truncating the weights of American Indian students (no maximum, maximum 18,000, maximum 6,000), two alternative factor ranges (0.3 to 3.0, 0.67 to 1.5), two alternative minimum number of records per cell (7,15), and two alternative collapsing orders (race/ethnicity then enrollment then grade level, enrollment then grade level then race/ethnicity). We also tested the effect of truncating weights for Black and Hispanic students (maximum 60,000 for Black, 50,000 or 73,000 for Hispanic) to see if we could lower the variance for those groups as well. Note that not all possible combinations of weighting options were tested. Only the more promising alternatives to the current scheme were

investigated.

C. Why We Investigated Weight Truncation

We decided to investigate the effect of truncating weights due to the strong deviation from a normal distribution exhibited by the weights for some race/ethnic groups, especially for American Indians. In particular, the upper tail of the American Indian distribution is extremely drawn out. The maximum value is 60 times larger than the 99th percentile. For other race/ethnic groups, the maximum value is no more than 4 times larger than the 99th percentile.

The cause of this unusual distribution is related to the source from which American Indian students were selected for sample. Most of them came from BIA schools or schools with a large percentage of Indian students. These schools had been selected with certainty, or at least with a large probability.

Some of the schools with a small percentage of Indian students are large urban schools. Thus some of the American Indian students selected for sample are from the same strata as White students, with the majority coming from strata having much larger probabilities of selection.

D. Criteria for Evaluating the Options

We decided that since much of our weighting methodology involves making bias-for-variance trade-offs, the most appropriate criteria for evaluation should be to look at mean-squared-error as the measure of the quality of our estimates. Thus, we computed $MSE(x) = Var(x) + B^2(x)$ for our alternative weighting schemes to measure the relative quality of each. The root MSE was displayed since it is easier to look at.

E. Results

The root MSE for each weighting option are listed in Tables 2 for each race/ethnic group. Notice that option 10 generally produces the lowest MSE of all the options for all race/ethnic groups except Black, where it is only slightly higher than option 9.

As further evaluation of the results, two characteristics not directly influenced by the weighting scheme were chosen for analysis. Number of students by sex and number of students receiving free or reduced price lunch were chosen to determine the effect of the alternative weighting schemes.

Only coefficients of variation are looked at here rather than mean-squared error since we had no reason to believe these estimates would be differentially biased with respect to the weighting scheme being implemented. Results of this evaluation were inconclusive.

F. Conclusions

Based on the information in Table 2 , we concluded that Option 10 produced the lowest mean-squared error for total students by race. The further evaluation described above was inconclusive, with no option producing clearly superior results. Thus, since option 10 proved superior at controlling race/ethnic totals and was no worse than any other weighting option with regard to uncontrolled data items, we decided to implement option 10 as the final weighting scheme for the student records survey. In accordance with this decision, the following changes were made to the weighting procedure:

- American Indian Students' base weights were truncated at 18,000 if they had originally been higher.
- The lower and upper bounds of the Student Adjustment Factor were expanded from 0.67 and 1.5 to 0.3 and 3.0.
- The order to collapsing cells for the Student Adjustment Factor was changed from race/ethnicity then enrollment and grade level to enrollment first, then grade level, and finally race/ethnicity.

VI. RECOMMENDATIONS FOR IMPROVING THE STUDENT RECORDS SURVEY

Listed below are recommendations we believe should be incorporated in SASS Student Surveys in the future:

- We should ask how many classes each student is taking so we won't have to assume all students in a school take the same number of classes.
- We should be more careful and observant during the transfer of information from student worksheets when creating the student files. Strict attention paid during the construction of student records should reduce some of the missing data problem and should eliminate the need for the clerical transfer of the information to identify multiple records. An automated sampling worksheet with internal edits would eliminate many of the sampling problems.
- We should make the selection of sample class periods more user-friendly. A process that is easier to follow would yield more accurate results.
- Further research is needed into determining an optional weighting scheme that minimizes the mean-squared error for the characteristics of interest.

References

King, Karen E. and Steven Kaufman (1994) , Estimation Issues Related to the Student Component of the SASS: Proceedings of the Section on Survey Research Methods, American Statistical Association.

Abramson, Robert et.al. (1996), 1993-94 schools and Staffing Survey: Sample Design and Estimation, National Center for Education Statistics Technical/Methodology Report, No.96-089.

Table 1: Student Survey Weighting Options

Weighting Option	Student Adjustment Factor Collapsing: Lower and Upper Bounds	Weighting Truncations: Maximum Basic Weight			Student Adjustment Factor Collapsing : Minimum Number of Records Per Cell	Student Adjustment Factor Collapsing: Order of Collapsing*
		American Indian	Hispanic	Black		
1-Original	0.67 to 1.5	None	None	None	15	R,ENRL,GL
2	0.30 to 3.0	None	None	None	15	R,ENRL,GL
3	0.30 to 3.0	18,000	None	None	15	R,ENRL,GL
4	0.30 to 3.0	18,000	73,000	60,000	15	R,ENRL,GL
5	0.30 to 3.0	18,000	50,000	60,000	15	R,ENRL,GL
6	0.30 to 3.0	18,000	None	None	7	R,ENRL,GL
7	0.30 to 3.0	6,000	None	None	7	R,ENRL,GL
8	0.67 to 1.5	18,000	None	None	7	R,ENRL,GL
9	0.67 to 1.5	18,000	None	None	15	ENRL,GL,R
10	0.30 to 3.0	18,000	None	None	15	ENRL,GL,R

* R = race/ethnicity
ENRL = enrollment Category
GL = grade level

Table 2: Root Mean-Squared Error for the Weighting Options by Race/Ethnicity

Root MSE					
Weighting Options	American Indian	Asian	Hispanic	Black	White
1	190,833	639,026	635,810	568,756	999,258
2	215,249	242,837	356,308	155,887	436,831
3	62,627	329,320	279,898	174,453	436,831
4	62,565	329,320	278,854	173,581	436,831
5	62,343	329,320	278,810	173,581	436,831
6	63,835	396,661	263,345	162,358	457,589
7	117,175	396,661	299,905	165,730	457,589
8	79,368	624,259	504,457	513,088	964,551
9	65,398	643,071	180,958	118,212	693,199
10	20,693	222,662	179,254	118,475	288,271