NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

The Working Paper Series was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series.

U. S. Department of Education
Office of Educational Research and Improvement

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association

Working Paper No. 97-41 December 1997

Contact: Steve Kaufman
Surveys and Cooperative Systems Group
(202) 219-1337
e-mail: steve_kaufman@ed.gov

U. S. Department of Education
Office of Educational Research and Improvement

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Surveys and Cooperative Systems Group
Paul D. Planchon
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association*, Working Paper No. 97-41. Project Officer, Steve Kaufman. Washington, D.C.: 1997.

December 1997

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Ruth R. Harris at (202) 219-1831 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5654.

Samuel S. Peng
Acting Director
Statistical Standards and Services Group

This page intentionally left blank.

Table of Contents

Preface

The five papers contained in this volume were presented at the 1997 American Statistical Association (ASA) meeting in Anaheim, California (August 10-14). This is the fifth collection of ASA papers of particular interest to users of NCES survey data published in the ***Working Papers Series***. The earlier collections were Working Paper 94-01, which included papers presented at ASA meetings in August 1992 and August 1993 and the ASA Conference on Establishment Surveys in June 1993, Working Paper 95-01, which included papers from the 1994 ASA meeting, Working Paper 96-02, which included papers from the 1995 ASA meeting, and Working Paper 97-01, which included papers from the 1996 ASA meeting.

A list of SASS methodological papers and reports is included in the following pages for readers who wish to learn more about the Schools and Staffing Survey.

SASS Methodological Papers and Reports Reference List

Bushery, J., Royce, D., and Kasprzyk, D. (1992). "The Schools and Staffing Survey: How Reinterview Measures Data Quality." In *Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 7-12. NCES Working Paper No. 94-01. Washington, D.C.: National Center for Education Statistics.

Chaney, B. (1994). *The Accuracy of Teachers' Self-Reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey.* NCES Working Paper No. 94-04. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Cole, C., Parmer, R., and Schwanz, D. (1997). *Improving the Mail Return Rates of SASS Surveys: A Review of the Literature.* NCES Working Paper No. 97-18. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Fink, S. (1994). *Schools and Staffing Survey 1990-91: SASS Cross-Questionnaire Analysis.* NCES Working Paper No. 95-03. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Holt, A., and Scanlon, B. (1994). *QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates.* NCES Working Paper 95-02. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Isaacs, J., Garet, M., and Sherman, J. (1997). *Collection of Private School Finance Data: Development of a Questionnaire.* NCES Working Paper No. 97-22. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Jabine, T. (1994). *Quality Profile for SASS: Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS).* NCES 94-340. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: Government Printing Office.

Jackson, B., and Frazier, R. (1996). *Improving the Coverage of Private Elementary-Secondary Schools.* NCES Working Paper No. 96-26. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Jenkins, C. (1992). "Questionnaire Research in the Schools and Staffing Survey: A Cognitive Approach." In *Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 19-24. NCES Working Paper No. 94-01. Washington, D.C.: National Center for Education Statistics.

Jenkins, C. (1997). *Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year.* NCES Working Paper No. 97-10. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

__________, and Von Thurn, D. (1996). *Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey.* NCES Working Paper No. 96-05. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Kasprzyk, D., Gruber, K., Salvucci, S., Saba, M., Zhang, F., and Fink, S. (1994). "Some Data Issues in School-Based Surveys." In *Schools and Staffing Survey (SASS): 1994 Papers Presented at the 1994 Meeting of the American Statistical Association,* pp. 32-37. NCES Working Paper No. 95-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Kaufman, S. (1992). "Balanced Half-Sample Replication with Aggregation Units." In *Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 25-30. NCES Working Paper No. 94-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

__________. (1993). "A Bootstrap Variance Estimator for the Schools and Staffing Survey." In *Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 74-79. NCES Working Paper No. 94-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

__________. (1994). "Properties of the Schools and Staffing Survey's Bootstrap Variance Estimator." In *Schools and Staffing Survey (SASS):1994 Papers presented at the 1994 Meeting of the American Statistical Association,* pp. 12-23. NCES Working Paper No. 95-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

__________. (1997). "Estimating the Variance in the Presence of Imputation Using a Residual." In *Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association,* pp. 45-50. NCES Working Paper 97-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Li, B. and Scheuren, F. (1996). "Improved GLS Estimation in NCES Surveys." NCES Working Paper No. 96-02. U.S. Department of Education, Office of Educational Research and Improvement. Washington, D.C.: National Center for Education Statistics.

Monaco, D., Salvucci, S., Zhang, F., Hu, M., and Gruber, K. (1997). *An Analysis of Total Nonresponse in the 1993-94 Schools and Staffing Survey (SASS).* NCES 98-243. U.S. Department of Education, Office of Educational Research and Improvement. Washington, D.C.: National Center for Education Statistics.

Mullens, J., and Kasprzyk, D. (1996). *The Schools and Staffing Survey: Recommendations for the Future.* Conference Proceedings. NCES 97-596. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: Government Printing Office.

Parmer, R., Shen, P-S., and Tan, A. (1992). "Mail Versus Telephone Response in the 1990-91 Schools and Staffing Surveys." In *Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 13 -18. NCES Working Paper No. 94-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Royce, D. (1994). "The Results of the 1993 Teacher List Validation Study (TLVS)." In *Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys.* NCES Working Paper No. 94-06. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Salvucci, S., Weng, S., and Kaufman, S. (1995). *Design Effects and Generalized Variance Functions for the 1990-91 Schools and Staffing Survey (SASS).* NCES 95-342. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington. D.C.: National Center for Education Statistics.

Salvucci, S., Walter, E., Conley, V., Fink, S., Saba, M., and Kaufman, S. (1997). *Measurement Error Studies at the National Center for Education Statistics.* NCES 97-464. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: Government Printing Office.

Scheuren, F., and Li, B. (1995). *Intersurvey Consistency in NCES Private School Surveys.* NCES Working Paper No. 95-16. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Scheuren, F., and Li, B. (1996). *Intersurvey Consistency in NCES Private School Surveys for 1993-94.* NCES Working Paper No. 96-27. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Scheuren, F., Monaco, D., Zhang, F., Ikosi, G., Chang, M., and Gruber, K. (1996). *An Exploratory Analysis of Response Rates in the 1990-91 Schools and Staffing Survey (SASS)*. NCES 96-338. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Shen, P-S., and Fisher, R. (1993). "Adjusting for Nonresponse Bias of Correlated Items Using Logistic Regression." In *Schools and Staffing Survey, (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 80-84. NCES Working Paper No. 94-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

__________, Parmer, R., and Tan, A. (1992). "Characteristics of Nonrespondents in the Schools and Staffing Surveys' School Sample." In *Schools and Staffing (SASS) Papers Presented at Meetings of the American Statistical Association,* pp. 31-36. NCES Working Paper 94-01. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Smith, W., Ghosh, D., and Chang, M. (1995). "Optimal Periodicity of a Survey: Alternatives under Cost and Policy Constraints." In *Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association,* pp. 29-34. NCES Working Paper No. 96-02. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Smith, W., Ghosh, D., and Chang, M. (1997). *Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis.* NCES Working Paper No. 97-14. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

Zukerberg, A., and Lee, M. (1997). *Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form.* NCES Working Paper No. 97-23. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington, D.C.: National Center for Education Statistics.

APPLYING MASS IMPUTATION USING THE SCHOOLS AND STAFFING SURVEY DATA

Steven Kaufman, National Center for Education Statistics; Fritz Scheuren, Ernst and Young
Steven Kaufman, Room 422d, 555 New Jersey Ave. N. W., Washington, D.C. 20208

Key Words: Nearest Neighbor Imputation, Bootstrap Variance, Simulation

1. Introduction

During the past few years the authors have been trying to find a methodology which produces a set of survey weights for which the estimates of number of schools, students and teachers agree for both a sample survey and its frame. The frame estimates are produced for the same time period as the survey collection. Of course, the "control" is only achieved for a specified set of cells.

If there were only one estimate, say number of schools, then a simple raking procedure would do the job. However, we require the agreement of three independent estimates simultaneously. Raking does not converge for this problem.

Scheuren and Kaufman (1996) found a solution using a generalized least squares methodology (GLS). Using a multivariate ratio adjustment first, weights less than 1 for the most part become minimal; and the implementation become manageable. The problem with our "solution" is that for cells not controlled for, the original weights quite often produce closer agreement than the GLS weights. In the 1996 paper, it was proposed that a mass imputation procedure (Kovar and Whitridge 1995) might provide better results.

Mass imputation is where the survey respondents are used as donors to impute back to the entire frame. If the survey respondent data were mass imputed to the frame for all data elements except the schools, students and teacher elements then the desired consistency would be achieved for all estimation cells. All weights would equal 1 and the survey estimates for schools, students and teachers would equal the frame total.

The principal problem with this version of mass imputation is the difficulty in variance estimation for survey variables other than school, student and teachers. Since values are assigned to the entire frame, standard variance procedures produce a zero variance. A variance procedure that measures the imputation variance is required to compute the mass imputation variance estimates.

Shao and Sitter (1996) proposed a methodology for measuring the imputation variance. It works for general estimates coming from any sample design and imputation methodology. The methodology calls for generating bootstrap samples of both respondents and nonrespondents. The original imputation procedure is applied to each bootstrap sample; and the distribution of bootstrap estimates is relied on for inference.

This paper investigates the magnitude of the precision gains that we thought mass imputation promised. Additionally, a variance estimator for the mass imputation, motivated by the Shao and Sitter methodology, is developed. (One potential problem, for example, with the Shao and Sitter methodology is the assumption that probabilities of being a nonrespondent are equal within an imputation cell. With mass imputation, this need not necessarily be true.)

The precision gains of mass imputation and its proposed variance estimator are tested through a simulation study. Frame variables are used, so that true values for all sample imputed estimates are known. The "respondents" or donors are determined using a single stage probability proportionate to size sample design, similar to the Schools and Staffing Survey. Given these respondents:

(1) A mass imputation is performed and compared with the standard Horvitz-Thompson estimator.
(2) Further, an estimate of the true mean square error (MSE) of the mass imputation estimate is compared to the Horvitz-Thompson variance.
(3) Additionally, the proposed mass imputation variance is computed and compared to an estimate of its true variance. Since all sample estimates can be obtained exactly, estimates of the true variance are computed using the simple variance of the selected simulation sample estimates.
(4) Finally, the case when nonrespondents are selected completely at random will be investigated by simulating a sample design following this assumption.

2. NCES Applications for Mass Imputation

The motivating application for mass imputation is the GLS problem described in the introduction. However, two other applications are possible. For example, NCES uses indirect estimation procedures to produce state private school enrollment figures. The estimation procedure applies an adjustment factor to each known private school. The adjustment reflects results from a survey that measures the number of schools missing from our lists. Since this is a mass imputation procedure, the results of this paper can be useful in the variance estimation of these state estimates.

Another potential application of mass imputation

is in the Center's data warehousing project. Here, the object is to link past and current Center surveys across program areas. If sample surveys are mass imputed to their frame, then the linkage problem is reduced to linking the frames, eliminating the need to link all the sample surveys. Again, the variance estimator proposed here might be useful.

3. **Imputations**

3.1 Nearest Neighbor Imputation. -- The nearest neighbor imputations used in this paper are done within imputation cells, after the schools have been sorted by the number of students per school. The imputation cells are state/school level/urbanicity. There are three school levels - elementary, secondary and combined; also three levels of urbanicity - central city, urban fringe/large town and rural/small town. After the file is sorted, it is accessed sequentially using the nearest responding school as the donor for a nonresponding school. For a particular unit i, we represent the imputed value for variable y_i by $y_{i\bullet}$.

The imputation described above is used two ways with different file sortings before the imputations are determined. The first imputation sorts the file in ascending order within imputation cell. This will be referred to as ascending imputation. The second imputation is done by determining the imputations in ascending, as well as descending order. Each time an imputation is required, a random 50/50 selection is used to determine which imputation is used in the estimate.

3.2 Mass Imputation. -- In mass imputation the sample weights associated with a probability sample s are ignored. Instead, it is assumed that the entire frame is in the sample, but the only units that respond are the units in s. Estimates are produced by assigning all units on the frame a weight of 1 and using the units in s as the donors to impute all the other frame units. The nearest neighbor imputation, described above, was used in the mass imputation process. After the imputation, estimates were computed as though the entire frame responded. If the imputation process is "good," then there may be some efficiency gains, compared to the usual Horvitz-Thompson estimator.

4. **Sample Selection (Mass Imputation Donors)**

Two sample designs will be used. The main design studied employed the square root of the number of teachers in a school as a PPS measure of size; the second (subsidiary) design selected units with equal probability within an imputation cell. The first is used to test the mass imputation procedures under the SASS sample design, while the second is used for comparison to verify the importance of the missing completely at random assumption.

4.1 The Schools and Staffing Sample Design. -- The Schools and Staffing Survey (SASS) is a stratified probability proportionate to size (PPS) sample of elementary, secondary and combined schools. The selection is done systematically using the square root of the number of teachers per school as the measure of size. State-by-school level cells define the stratification. Before systematic selection, schools are sorted to provide a good geographic distribution. Sample allocations are designed to provide reliable state estimates. In this simulation study, four small States were studied. The sample state sizes ranged from 72 to 196 schools. The sampling rates ranged from 14 to 42 percent of each state's school population.

In order to eliminate the SASS design effects from systematic sampling and high sampling rates, the simulation split each state/school level stratum into a number of substrata (h) so that exactly two schools are selected within each substratum **with replacement**. The original SASS sample sizes by state were, however, maintained.

4.2 The Equal Probability Sample Design. -- The Shao - Sitter variance methodology assumes that nonrespondents are missing completely at random. To test the importance of this assumption in the SASS setting, the sample selection procedure described above was modified to select each school in a stratum with equal probability. Within each state, the sample sizes again remained the same, but the allocation and stratification boundaries were altered to achieve the desired equal selection probabilities. Again, two units will be selected within each stratum **with replacement.**

5. **Mass Imputation Bootstrap Variances**

To generate the bootstrap variance estimator for an estimate θ, the following is done:

(1) A bootstrap sample s^* is generated by selecting 2 units from s within each of the h stratum. The selection is done with equal probability and with replacement.

(2) Then s^* is sorted by the imputation cell and one bootstrap unit is randomly selected within each imputation cell and eliminated from s^*. This is done in an attempt to produce a more unbiased variance estimate. The Shao - Sitter procedure does this by selecting n-1 units within each stratum. In the Shao and Sitter setting, this is appropriate since variability is introduced through the sampling mechanism. In the mass imputation setting, there is only an imputation variance. Therefore, the appropriate place to reduce the sample size seemed to be where the imputation process begins – the imputation cells. To verify that the imputation cell

is the appropriate place to reduce the bootstrap sample size by one, a simulation was done, reducing the sample size in the stratum controlling the donor selection. (See Table 4)

(3) The bootstrap mass imputation is generated by doing the mass imputation procedure on the original frame using the units determined in step 2 as donors.

(4) Using the results from step 3, compute the bootstrap estimate θ^* the same way θ is calculated.

(5) Repeat steps (1) to (4) B times, producing bootstrap estimates θ_j^*, j equaling 1 to B.

(6) The simple variance of the θ_j^* is our bootstrap variance estimate ($V^*(\theta)$).

6. Simulations

There were 2,000 simulations performed for each sample design described above. Mass imputations and bootstrap variances are computed for each simulation. The estimate and analysis statistics used in the simulation are described below.

6.1 Estimates. -- Four mass imputation estimates per state are computed:

$$\hat{y}_{Me1} = \sum_{k \in N} S_k p_{ke1}; \hat{y}_{Me2} = \sum_{k \in N} S_k p_{ke2};$$

$$\hat{y}_{Mm} = \sum_{k \in N} S_k p_{km}; \text{and } \hat{y}_{Mh} = \sum_{k \in N} S_k p_{kh}.$$

S_k is the known number of students in school k

p_{ke1} : student proportion in school k grades pre - kindergarten to 3

p_{ke2} : student proportion in k grades 4 to 6

p_{km} : student proportion in k grades 7 to 9

p_{kh} : student proportion in k grades 10 to 12

It is assumed that S_k is known for all k and that only the p's require collection. Therefore, when k is not selected to be a responding unit, a nearest neighbor donor's p will be applied to S_k.

Additionally, four Horvitz-Thompson estimates are computed within each state:

$$\hat{y}_{e1} = \sum_{k \in s} w_k S_k p_{ke1}; \hat{y}_{e2} = \sum_{k \in s} w_k S_k p_{ke2};$$

$$\hat{y}_{m} = \sum_{k \in s} w_k S_k p_{km}; \text{and } \hat{y}_{h} = \sum_{k \in s} w_k S_k p_{kh}.$$

where: w_k is the inverse of the selection probability and s is the set of all selected schools.

6.2 Simulated Variance and Bias Estimates. -- The two variance estimates computed within each sample and averaged across samples are:

$$V^*(\hat{y}_\bullet) = 1/B \sum_{j=1}^{B} (\hat{y}^*_{\bullet j} - \bar{y}^*_\bullet)^2$$

$\hat{y}_\bullet$: mass imputation estimates described above,

$\hat{y}^*_{\bullet j}$: a bootstrap estimate of $\hat{y}_\bullet$,

$\bar{y}^*_\bullet$: the average of the bootstrap estimates $\hat{y}^*_{\bullet j}$.

Estimates of the true variance of the mass imputation estimate ($\hat{y}_\bullet$) and the Horvitz-Thompson estimate ($\hat{y}$) are provided below:

$$V_T(\hat{y}_\bullet) = 1/n \sum_{s=1}^{n} (\hat{y}_{\bullet s} - \bar{y}_{\bullet s})^2$$

$\hat{y}_{\bullet s}$ and $\bar{y}_{\bullet s}$ are the value of $\hat{y}_\bullet$ for the s^{th} simulation and the average of the $\hat{y}_{\bullet s}$, respectively.

$$V_T(\hat{y}) = 1/n \sum_{s=1}^{n} (\hat{y}_s - \bar{y}_s)^2$$

$\hat{y}$: Horvitz-Thompson estimate,

$\hat{y}_s$: s^{th} Horvitz-Thompson estimate,

$\bar{y}_s$: average of the Horvitz-Thompson estimates.

The bias of the mass imputation estimate ($\hat{y}_\bullet$) is estimated by:

Bias($\hat{y}_\bullet$)= $\bar{y}_{\bullet s} - \bar{y}_s$

7. Analysis Statistics from Simulations

Four tables are provided at the end of this paper which provide summaries of our simulation results. Three key analytic statistics have been used:

(1)To evaluate the imputation methodology, the relative bias of the estimated standard error (RBS) is computed.

$$RBS = \left(\sqrt{\hat{V}(\hat{y}_\bullet)} - \sqrt{V_T(\hat{y}_\bullet)}\right) / \sqrt{V_T(\hat{y}_\bullet)}$$

(2) The relative precision of the mass imputation estimate (RPS) is given by:

$$RPS = \sqrt{\left(V_T(\hat{y}_\bullet) + \text{Bias}^2(\hat{y}_\bullet)\right)} / \sqrt{V_T(\hat{y})}$$

(3) The relative bias of the mass imputation estimate (RBE)

$RBE = (\bar{y}_{\bullet s} - \bar{y}_s) / \bar{y}_s$.

7.1 Table 1 Overall Comparisons. -- Table 1 displays how mass imputation might work in the SASS setting using ascending imputations. The answer to the question of whether this approach is satisfactory is "no." The precision of mass imputation relative to the Horvitz-Thompson clearly gives the Horvitz-Thompson estimator the advantage. The mass imputation estimator only once has a large efficiency gain over Horvitz-Thompson (18.4 percent). Seven times there was not much difference between the two estimation methods. In these cases, the gains ranged from -11.5 to +12.4 percent. On the other hand, there

were eight times when mass imputation had a big precision loss. In these latter cases, the lost was between -15.3 and -61.1 percent. Overall, mass imputations perform poorly for these estimates in the states tested.

7.2 Table 1 Standard Error Comparisons. -- Table 1 also demonstrates with ascending imputations, through the relative bias of the standard error (RBS) values shown there, that our variance procedure underestimates the standard error. Most of the time the absolute bias is less than 10 percent. The reason for the underestimate may be the way the bootstrap sample sizes were reduced by one. Some theoretical work may be needed to provide an unbiased estimator. Another possible reason (See Table 2 below), is the inappropriateness of the completely missing at random assumption for the nonrespondents.

7.3 Table 2 Standard Error Comparisons. -- The purpose of table 2 is to investigate (using ascending imputations) the robustness of the missing completely at random assumption of the variance procedure. Comparing the relative biases of the standard error in tables 1 and 2 shows that the standard errors are reasonably close to the true values, but still consistent underestimate. When the missing completely at random assumption is violated (table 1), there are more large underestimates than when the assumption is not violated. This indicates the missing completely at random assumption is not critical in this setting, but cannot be completely ignored.

7.4 Other Points from Table 2. -- Table 2 also demonstrates another point. Since the selection and allocation are designed to produce an equal probability sample, one might expect the Horvitz-Thompson estimator to be inefficient estimating total numbers of students. That being the case, the mass imputation might compensate for the deficiencies in the sample design and be more efficient then the Horvitz-Thompson estimator. This does not appear to be true. The mass imputation does much worse than the Horvitz-Thompson estimator five times. In those cases, the loss of precision ranged from -21.2 to -43.1 percent. The mass imputation has large precision gains five times. The gains ranged from +19.4 to +33.4 percent. However, it should be noted that mass imputation with the equal probability design performed better than the unequal probability design. Therefore, there is some truth to the assertion made at the beginning of this paragraph.

7.5 Table 3 Ascending and Descending Imputation. -The purpose of table 3 is to determine whether mass imputation might work in the SASS setting when both ascending and descending imputations are used. The answer to this question is still "no." Twice the mass imputation was much better than Horvitz-Thompson. In these cases, the gains were +23.2 and +27.2 percent. Eleven times there was not much difference between the two estimation methods. In these cases, the gains ranged from −14 to +11.2 percent. Three times, mass imputation had a big precision loss between -17.1 to -21.9 percent.

Still, as might be expected intuitively, using the ascending/descending imputation works much better than just ascending imputation. This is seen by comparing Tables 1 and 3. Using the ascending imputation, mass imputation is reasonably close or better than Horvitz-Thompson eight times, while the ascending/descending is reasonably close or better thirteen times. One reason for this is that the ascending/descending imputation is generally less biased. Since both imputation methodologies have small biases, this is not the main reason for the difference.

The main reason for the difference is a smaller variance. Now there are two ways to reduce this variance: (1) reduce the variability of the donor enrollment counts or (2) reduce the variability of the weights. Since donor enrollment counts are the same for both imputation methodologies, the reduction in variance is coming entirely from reducing the variability of the weights. Since the ascending/descending mass imputation imputes values from both sides of the missing data at an expected 1 to 1 rate, the implicit weight for this processes will be (more or less) the moving average of the individual ascending and descending imputation weights. Since the units are sorted by size, the weights should be increasing as you go down the file. Therefore, the moving average of the weights should be less variable than the individual weights.

7.6 Table 4 On Bootstrap Bias Issues. -- Table 4 displays what happens when reducing the bootstrap sample sizes by one at the sampling stratum level, rather than at the imputation cell level as earlier. As can be seen in the relative bias of the standard error from table 4, the standard errors are overestimated by +24 to +48 percent. This shows that reducing the bootstrap sample size at the imputation cell level, although a slight underestimate, is better than doing it at the sample stratum level. Again, some work on the theory would seem to be needed here.

8.0 Conclusions And Areas For Future Study

8.1 Some Basic Conclusions. -- A lot was learned from the simulation work discussed here:

(1) We remain convinced, for example, of appeal of being able to make greater use of the frame variables to improve estimation; however, we now have a much greater appreciation of the practical

difficulties in an actual implementation of such a procedure.:

(2) Broadly speaking, for the states and variables used in the present analysis, the bias in the mass imputation estimator, no matter how conducted, is relatively small. This suggests that any of the nearest neighbor imputation variants employed here would be sufficient for a mass imputation process.

(3) However, while at times the mass imputation estimator did outperform the Horvitz-Thompson estimator, it just never performed better overall.

(4) Of the two methods of imputation employed, the ascending/descending method was clearly superior to just using the ascending method alone.

(5) As our work on the robustness of the "missing at random" assumption demonstrated, the sample design and the imputation procedure cannot be treated independently.

(6)The variance estimation procedure proposed in this paper seems to work reasonably well. Most of the time, it underestimates the variance slightly. Occasionally, though, the variance is greatly underestimated, especially when the selection probabilities within imputation cells are unequal.

(7) The proposed variance procedure does not appear to be unbiased, although some ad hoc adjustments seem clearly better than others (e.g., adjusting at the imputation cell rather than stratum level).

8.2 <u>Some Next Steps and Second Thoughts</u>. -- Without question we were disappointed with the performance of mass imputation. While not a failure, so far it has not delivered on our expectations. Some conjectures about why:

(1) One possible reason is that, even though the imputation took school size into account, there were not enough large schools selected in the equal probability design to measure the large school's distribution appropriately. As noted already, users of mass imputation must take the sample design into account when determining an appropriate imputation.

(2) Another possibility is that a better imputation procedure needs to be used. Fixing either of these possibilities requires designers very knowledgeable about the data being imputed. Clearly, just using a relatively efficient general imputation procedure, like nearest neighbor, does not guarantee good performance of the mass imputation estimator.

(3) In doing the imputations, no control was introduced on the number of times a donor case was used. If done, this, all by itself, might have improved our results dramatically. We conjecture that the cases where extremely poor results were obtained would have been lessened.

(4) Theoretical work on nearest neighbor imputation, given at these meetings, also is a place to look for ideas for improvements (Chen and Shao 1997). In addition, theoretical work seems required to find an unbiased variance methodology. Even so, given the general difficulty of variance estimation for indirect estimates, the variance procedure described here may be applicable to the indirect estimation problem stated in section 2. We are less sure, by the way, about the application of mass imputation to NCES's data warehousing work.

(5) If the ascending/descending mass imputation is used when the donors are selected with equal probability, then the mass imputation may well outperform Horvitz-Thompson. However, there was not the time to do this simulation for the present paper.

(7) While we started off our work determined to better the GLS procedures studied earlier, no direct comparison was made here with a comparable GLS estimator. We conjecture that mass imputation at best may be not much better than a "wash' when imputing based on a single variable; but that as the dimensionality of the information used from the frame grows, mass imputation may yet show it value.

Table 1 -- Relative Precision (RPS), Bias (RBS) of the Mass Imputation Standard Error and Relative Bias (RBE) of the Estimator using SASS Sample Design and Ascending Imputations

State	Est.	Standard Error		Estimate
		Relative Precision	Relative Bias	Relative Bias
2	$\hat{y}_{Me1}$	100.2	-5.7	0.1
	$\hat{y}_{Me2}$	89.7	-8.8	-0.2
	$\hat{y}_{Mm}$	112.4	8.4	-2.0
	$\hat{y}_{Ms}$	88.5	5.5	2.7
9	$\hat{y}_{Me1}$	136.2	-17.6	-3.3
	$\hat{y}_{Me2}$	126.0	-18.0	-0.1
	$\hat{y}_{Mm}$	137.3	-16.5	3.0
	$\hat{y}_{Ms}$	81.4	-6.9	1.9
10	$\hat{y}_{Me1}$	132.2	-3.3	-3.3
	$\hat{y}_{Me2}$	115.3	-3.8	3.8
	$\hat{y}_{Mm}$	118.9	3.6	0.9
	$\hat{y}_{Ms}$	99.4	11.6	-.04
24	$\hat{y}_{Me1}$	161.1	-9.7	-4.5
	$\hat{y}_{Me2}$	108.0	-19.3	-0.5
	$\hat{y}_{Mm}$	156.1	-9.4	5.3
	$\hat{y}_{Ms}$	104.4	-10.8	1.3

Table 2 -- Relative Precision (RPS), Bias (RBS) of the Mass Imputation Standard Error and Relative Bias (RBE) of the Estimator with Equal Probability Selection of Donors and Ascending Imputations

State	Est.	Standard Error		Estimate
		Relative Precision	Relative Bias	Relative Bias
2	$\hat{y}_{Me1}$	92.3	-10.6	0.1
	$\hat{y}_{Me2}$	72.9	-10.0	0.5
	$\hat{y}_{Mm}$	110.7	-5.3	1.5
	$\hat{y}_{Ms}$	80.6	-2.2	-2.9
9	$\hat{y}_{Me1}$	123.2	-11.7	-0.8
	$\hat{y}_{Me2}$	106.0	-16.6	0.0
	$\hat{y}_{Mm}$	112.3	-12.6	-0.2
	$\hat{y}_{Ms}$	66.6	-2.9	1.6
10	$\hat{y}_{Me1}$	133.2	-0.1	-1.5
	$\hat{y}_{Me2}$	98.5	0.4	2.1
	$\hat{y}_{Mm}$	121.2	-2.9	2.0
	$\hat{y}_{Ms}$	96.6	12.4	-3.0
24	$\hat{y}_{Me1}$	143.1	-13.3	-2.4
	$\hat{y}_{Me2}$	80.4	-14.1	-0.3
	$\hat{y}_{Mm}$	128.7	-13.2	3.1
	$\hat{y}_{Ms}$	68.6	-8.0	0.8

Table 3 -- Relative Precision (RPS), Bias (RBS) of the Mass Imputation Standard Error and Relative Bias (RBE) of the Estimator using SASS Sample Design and Ascending and Descending Imputations

State	Est.	Standard Error		Estimate
		Relative Precision	Relative Bias	Relative Bias
2	$\hat{y}_{Me1}$	92.6	-1.0	0.9
	$\hat{y}_{Me2}$	88.8	-6.0	-0.3
	$\hat{y}_{Mm}$	103.2	14.7	-1.8
	$\hat{y}_{Ms}$	72.8	21.9	0.3
9	$\hat{y}_{Me1}$	109.9	-17.6	-1.4
	$\hat{y}_{Me2}$	110.5	-18.3	0.2
	$\hat{y}_{Mm}$	109.1	-15.0	1.0
	$\hat{y}_{Ms}$	76.8	-1.9	1.1
10	$\hat{y}_{Me1}$	121.9	0.5	-0.1
	$\hat{y}_{Me2}$	108.2	1.9	0.5
	$\hat{y}_{Mm}$	114.0	3.1	0.0
	$\hat{y}_{Ms}$	94.4	15.4	-0.4
24	$\hat{y}_{Me1}$	120.3	-13.7	-1.4
	$\hat{y}_{Me2}$	92.1	-20.9	-0.2
	$\hat{y}_{Mm}$	117.1	-13.4	1.7
	$\hat{y}_{Ms}$	93.9	-4.8	0.3

Table 4 -- Relative Bias (RBS) of the Mass Imputation Standard Error, Adjusting the Bootstrap Sample Size at the Donor Selection Stratum Level using Ascending Imputations

State	Est.	Relative Bias	State	Est.	Relative Bias	State	Est.	Relative Bias	State	Est.	Relative Bias
2	$\hat{y}_{Me1}$	36.4	9	$\hat{y}_{Me1}$	24.3	10	$\hat{y}_{Me1}$	43.8	24	$\hat{y}_{Me1}$	28.9
	$\hat{y}_{Me2}$	30.6		$\hat{y}_{Me2}$	24.6		$\hat{y}_{Me2}$	42.6		$\hat{y}_{Me2}$	24.1
	$\hat{y}_{Mm}$	30.5		$\hat{y}_{Mm}$	25.2		$\hat{y}_{Mm}$	32.2		$\hat{y}_{Mm}$	29.6
	$\hat{y}_{Ms}$	30.5		$\hat{y}_{Ms}$	46.0		$\hat{y}_{Ms}$	47.6		$\hat{y}_{Ms}$	27.5

References

Chen, J. and Shao,J. (1997), Bias and variance of Survey Estimators Based On Nearest Neighbor Imputation, presented at the 1997 Anaheim Joint Sataistical Meetings.

Hidiroglou M. et al (1995), "Weighting and Estimation in Business Surveys," *Business Survey Methods* edited by Cox et al, New York: John Wiley & Sons, pp. 477-500.

Kovar, J. and Whitridge, P. (1995), "Imputation of Business Survey Data," in Cox, Binder, Chinnappa, Christianson, Colledge, and Kott, eds. *Business Survey Methods*, Wiley: New York.

Business Survey Methods edited by Cox et al, New York: John Wiley & Sons, pp. 447-502

Scheuren F. and Kaufman S. (1996), "Where Will It All End? Some Alternative SASS Estimation Research Opportunities," in *Proceedings of the Survey Research Methods Section, American Statistical Association, pp.* 423-428

Shao J. and Sitter R. (1996), "Bootstrap for Imputed Survey Data, *Journal of the American Statistical Association*, Vol. 91, No. 435, pp. 1278-1288.

THE EFFECT OF MODE OF INTERVIEW ON ESTIMATES FROM THE 1993-94 SCHOOLS AND STAFFING SURVEY (SASS) PUBLIC SCHOOL TEACHER SURVEY

Cornette L. Cole, Robert C. Abramson, Randall J. Parmer, Dennis J. Schwanz
Cornette L. Cole, U. S. Bureau of the Census, Washington, D.C. 20233

KEY WORDS: CATI, Significance tests, sign rank tests, mode of interview

I. Introduction

To reduce the cost of data collection and to improve the efficiency, data was collected for the 1994 Schools and Staffing Survey's (SASS) Public School Teacher Survey by mail, telephone, and personal visits.

During the selection of sample teachers, a split-panel design was used where two-thirds of the teacher sample was randomly assigned for mail nonresponse follow-up interviewing from centralized computer-assisted telephone interviewing (CATI) facilities and the remaining one-third was assigned for telephone follow-up interviewing from decentralized facilities. The teachers were randomly assigned a teacher follow-up mode flag value of 1, 2, or 3 as indicated below. We used this teacher follow-up mode designation flag to separate teacher records and formed the CATI and nonCATI treatment groups we used for this analysis.

1 CATI
2 nonCATI
3 CATI (these records were initially held for possible sample reductions)

Data from the 1994 Public Teacher Survey was collected primarily through self-administered questionnaires, where sample teachers completed the questionnaires and returned them by mail. About 69% of the total interviews were mail returns.

Telephone calls were made from either CATI or decentralized facilities to teachers who did not return their questionnaires by mail. Personnel from the Census Bureau's Field Division were asked to determine the workload capacity of the centralized telephone interviewing (CATI) facilities. From the teachers who hadn't returned their questionnaires by mail, the indicated number of teacher records with follow-up mode designation flags which indicated that they had been designated for CATI follow-up interviewing, was sent to the CATI facilities to be interviewed.

The remaining CATI-designated cases, that is, those that CATI couldn't handle, were sent to be interviewed by decentralized telephone interviewing, along with mail nonrespondents previously designated for this follow-up mode. About 19% of the total interviews were completed in CATI interviews and 12% in decentralized (NON-CATI) telephone interviews.

A very small number of interviews, couldn't be interviewed through either of these telephone methods, and were completed during visits to schools by Census Bureau field representatives.

To be certain there was no bias in survey estimates because we used different modes in the telephone follow-up of mail nonrespondents, we initiated this study to compare the data we collected in CATI interviews with those collected in decentralized telephone interviews.

A. The Schools and Staffing Survey

The SASS is a periodic survey sponsored by the National Center for Education Statistics (NCES) and conducted by the U. S. Bureau of the Census. The SASS provides data on the policies and conditions of public and private elementary and secondary schools, principals, libraries, librarians, teachers and students in the United States.

The school, principal, library, librarian, teacher, and student samples were selected so that data from each of the components could be linked. For the 1993-94 school year, about 13,000 schools, 67,000 teachers, 7,600 libraries and librarians, and 6,900 students were selected[1] for SASS as follows:

- Private and public sample schools were selected first.
- All principals from SASS sample schools were in sample for the School Administrator Survey,
- A sample of teachers was selected within each of the SASS sample schools for the Teacher Survey.

---

[1]S . Kaufman et al. (1996). *1993-94 Schools and Staffing Survey: Sample Design and Estimation.* NCES 96-089. U. S. Department of Education, Office of Educational Research and Improvement. Washington, D. C.: National Center for Education Statistics.

- A subsample of SASS sample schools was selected for the Library and Librarian Surveys.
- And a subsample of SASS sample schools and teachers was selected to participate in the Student Record Survey.

B. Public School Teacher Survey Sampling Procedure

The sample of teachers for the Public School Teacher Survey was selected from SASS sample schools. Each sample school was asked to provide a list of its teachers with the information below for each teacher:

- whether the teacher was new (less than three years experience) or experienced,
- the teacher's race and ethnicity,
- whether he or she was considered a Bilingual or English as a Second Language (ESL) teacher
- his or her main field of teaching

Within each sample school, sample teachers were classified into one of the following five strata in the hierarchical order listed below. For example, if a teacher is both API and bilingual the teacher was assigned to the API stratum.

(1) Asian or Pacific Islander (API)
(2) American Indian, Aleut, or Eskimo (AIAE)
(3) Bilingual
(4) New
(5) Experienced

Within each school and teacher stratum, teachers were selected with equal probability. From the lists of teachers provided by the schools, 56,736 public school teachers were selected.

C. Estimation

The weight used to produce estimates of public school teacher characteristics was a product of the following weight and factors:

- <u>Basic Weight</u> - the inverse of the probability of selection

- <u>School Sampling Adjustment Factor</u> - an adjustment to the school's probability of selection to account for school mergers, splits, and duplicates

- <u>School Nonresponse Adjustment Factor</u> - an adjustment to account for teachers whose schools did not provide a list of its teachers

- <u>Teacher Within School Noninterview Adjustment Factor</u> - an adjustment that accounts for teacher nonrespondents

- <u>Frame Ratio Adjustment Factor</u> - a factor which adjusts teacher estimates to the total universe count of teachers from the public school sample frame

- <u>Teacher Adjustment Factor</u> - an adjustment which makes estimates of the weighted number of teachers from the SASS School and Teacher Survey consistent

II. Methodology

A. Estimates for the Analysis

A flag was assigned after the interviews were completed to indicate the actual mode of interview. The flag to indicate which telephone mode should be used to follow up mail nonrespondents was assigned <u>prior</u> to the initial questionnaire mailing. <u>*We used the follow-up mode designation flag to separate interviewed teacher records and form a CATI and NON-CATI treatment group for our analysis*</u>. Therefore, the treatment name (CATI or NON-CATI) is not necessarily an indicator of how the interview was actually completed.

<u>*The CATI treatment*</u> comprises all teachers who were designated for mail nonresponse follow-up interviewing from CATI facilities. Their interviews were actually either returned by mail or completed by telephone from CATI facilities.

<u>*The NON-CATI treatment*</u> comprises all teachers who were assigned for mail nonresponse follow-up interviewing from decentralized facilities. Their questionnaires may have actually been returned by mail or completed in interviews from decentralized facilities.

Recall that two-thirds of the teacher sample was assigned for CATI follow-up interviewing and the remaining one-third for telephone follow-up interviewing from decentralized facilities. To insure that the estimates we produce from records in our CATI and NON-CATI treatments would be approximately equal to the estimates we got for the entire sample, we increased the teacher basic weights on records in the CATI treatment by 1.5 and those in the NON-CATI treatment by 3.0. Then,

- We processed the CATI and NON-CATI data sets (separately) through the same weighting procedure used to weight the regular Public School Teacher Survey data.
- We further separated the reweighted teacher records

by actual mode of interview within each treatment.

- We finally produced CATI and NON-CATI treatment estimates for the Public School Teacher Survey questionnaire items.

We made CATI vs. NON-CATI comparisons of estimates for the following groups of teacher records:

Comparison Group 1: All interviews (all modes)

All teachers who were designated for telephone follow-up from centralized facilities (regardless of the mode the interview was completed in)

vs.

those who were designated for telephone follow-up from decentralized facilities (regardless of the mode the interview was completed in)

Comparison Group 2: Mail interviews only

Teachers who were designated for telephone follow-up from centralized facilities who returned their questionnaires by mail

vs.

those who returned their questionnaires by mail who had been designated for telephone follow-up from decentralized facilities

Comparison Group 3: Interviews completed during telephone follow-up only.

Teachers who were designated for telephone follow-up from centralized facilities and their interviews were completed in centralized telephone interviews

vs.

those who were designated for telephone follow-up from decentralized facilities and their interviews were completed in decentralized telephone interviews

Our primary interest was in Comparison Group 3. For these respondents, the telephone follow-up mode flag and the flag which indicates the actual interview modes have the same value.

B. Computing the Variances for the Analysis

We used the Balanced Repeated Replication (BRR) method in *WESVAR*[2] to compute variances for each estimate. The *WESVAR* BRR procedure uses replication techniques to calculate variances for estimates using:

$$v(\theta) = \frac{1}{G'}\sum_{k=1}^{G'}(\theta_k - \theta)^2$$

where,
θ = the estimator for the teacher questionnaire item
$v(\theta)$ = the variance of the estimate
G' = the number of replicates

The replicate weights we used to compute the variance estimates in *WESVAR* were computed using the same replicate factors used to calculate variance estimates for the regular 1993-94 SASS publication estimates.

C. Comparing Treatment Estimates

We evaluated the magnitude of the differences between CATI and NON-CATI estimates to see if they were statistically significant. We formed the null hypothesis,

$$H_o: \Theta_{CATI} = \Theta_{NONCATI}$$

which says an estimate produced using the records of CATI teachers (θ_{CATI}) is the same as that produced using the records of NON-CATI teachers ($\theta_{NON\text{-}CATI}$).

To test the hypothesis, we used the 'z' statistic:

$$z = \frac{\theta_{cati} - \theta_{non-cati}}{\sqrt{var(\theta_{cati}) + var(\theta_{non-cati})}}$$

where,

- θ is the estimate of the teacher characteristic of interest, $var(\theta)$ is its variance,
- the numerator is the difference between the CATI and NON-CATI estimates and
- the denominator is an estimate of the standard error of the difference.

A negative value for the z meant the NON-CATI estimate was higher, while a positive z value meant the CATI estimate was higher. Results of the significance tests are presented in Section III.

D. Evaluating the Distribution of the Differences between Treatment Estimates

Our significance tests evaluated the magnitude of the difference between CATI and NON-CATI estimates

[2]Westat, Inc., *The WESVAR SAS Procedure, Version 1.2*, Rockville, MD: Westat, Inc.

individually for each teacher questionnaire item. We used the sign rank test to evaluate the distribution of the differences across the items.

We used the SAS *PROC UNIVARIATE*[3] procedure to perform the Wilcoxon Signed-Rank Test. We assumed each difference was equally likely to be positive or negative and that the distribution of the differences is symmetrical. We tested the hypothesis that the median of the differences between the CATI and NON-CATI estimates is zero. The following steps are involved in the test:

- The absolute values of the differences are assigned ranks by magnitude, from smallest to largest, then the positive and negative signs are restored to the ranked values.

- The totals of the ranks with negative signs and those with positive signs are calculated.

The Wilcoxon signed rank statistic S is computed in SAS as follows:

$$S = \sum r_i^+ - \frac{n(n+1)}{4}$$

where,

- S is a sum of scaled binomial distributions
- r_i^+ is the rank of $|x_i|$ after discarding values of $x_I = 0$ and x_i is the difference between the CATI and NON-CATI estimates ($|\theta_{CATI} - \theta_{nonCATI}|$)
- n is the number of nonzero x_i values and
- the sum is computed over the values of x_i greater than zero.

The significance level of S is computed as:

$$\textit{Significance level} = S\,\frac{\sqrt{n-1}}{\sqrt{nV-S^2}} \quad \textit{where,}$$

$$V = \frac{n(n+1)(2n+1)-0.5\sum t_i(t_i+1)(t_i-1)}{24}$$

The sum is calculated over differences tied in absolute value and t_i is the number of tied values with the i^{th} difference.

[3] SAS Institute Inc., *SAS Procedures Guide, Version 6, Third Edition*, Cary, NC: SAS Institute Inc.

SAS outputs a probability or p-value that is a measure of the strength of the evidence against the null hypothesis. If the p-value is less than the significance level of the test, which in our case is 0.10, the null hypothesis should be rejected. The smaller the p-value, the stronger the evidence for rejecting the null hypothesis.

III. Results

A. Tests of Significance

At the $\alpha = .10$ level of significance, we expect no more than 10 percent of the estimates within a group would be significantly different. Table 1 provides a summary of the results of our significance tests. The table shows that for all three groups, more than 10 percent of the comparisons yielded statistically significant results.

Table 1 also shows that the interviews completed during telephone follow-up (Comparison Group 3) had the higher proportion of significant differences. In this group, we are comparing the responses of teachers who were actually interviewed from CATI facilities with those interviewed from decentralized telephone facilities. We see in Table 1 that there were about twice the proportion of significant differences between these respondents than mail respondents.

B. Items with significant differences

Most of the significant differences were between the responses of teachers in the two treatments to the series of questions labeled "Perceptions and Attitudes Toward Teaching" (Section E of the 1993-94 Public School Teacher questionnaire).

In general, NON-CATI treatment estimates were higher for categories of items which have negative connotations, while CATI treatment estimates were higher for responses which suggest these teachers had a more positive outlook.

The NON-CATI treatment estimate was higher for items which say

- More of these teachers believed their principal did not enforce student rules, he did a poor job of getting resources, and he did not let them know what was expected of them.
- More of them reported they had little influence or no control over the curriculum, textbooks, homework, and over teacher evaluations and
- More of them said they would remain in the same school system, but would teach at another school the

next year.

One the other hand, the CATI treatment estimates were higher for items which suggest the attitudes of most of these teachers were more positive.

- More of them reported their principals let them know what was expected from them and their schools' administrations treated them fairly and were supportive.
- More of them said their principals enforced school rules and backed them when they needed him to.
- More of them planned to continue teaching at the same school the next school year.

Responses to the question *"If you could go back to your college days and start over again, would you become a teacher or not?"* summarizes the contrast in attitude between teachers in the two treatments. The CATI estimate was higher for the category *'certainly would become a teacher'* and the NON-CATI estimate was higher for the category *'chances about even for or against'*.

C. Sign Rank Tests

In Table 2 below, probability values ($PR \geq |S|$) for the two-tailed tests are shown. Each p-value is greater than 10 percent, indicating that the hypothesis that the median of the differences between CATI and NON-CATI estimates is zero should not be rejected.

Table 2 also shows that the group consisting of interviews completed during telephone follow-up had the highest p-value. Thus, there is no evidence favoring the rejection of the hypothesis about the distribution of the differences for these respondents.

Our assumption for the sign-rank test was that each difference was equally likely to be positive or negative and that the distribution of the differences was symmetrical. The test results say there were about an equal number of differences with the CATI estimate higher as those with the NON-CATI estimate higher.

We stated earlier that most of the significant differences between the responses of teachers in the treatments were to attitude and perception items. Table 3 shows the two-tailed p-values from the sign-rank tests we performed using only the categories of attitude and perception items which have negative connotations.

Each p-value is less than 10 percent, suggesting we should reject the null hypothesis. The median difference for these items alone is different from zero and the distribution of the differences is skewed. This result agrees with our observation that NON-CATI treatment estimates were higher for these types of items.

IV. Conclusions

By randomly assigning teacher sample records between telephone modes, we gave each teacher record a chance of being assigned to CATI or decentralized telephone for follow-up. Our tests show that teachers within a treatment provided similar responses to attitude and perceptions questions and estimates of these responses were statistically different between treatments.

There are two possible explanations. One is that there was some periodicity in way the teacher records were ordered. This ordering resulted in the teachers assigned to the same treatment having similar characteristics. Another is that the assignment was truly random, but we were unlucky in the assignment, and the results of the assignment are due to the natural variability between teachers in the treatments.

There were also a higher proportion of significant differences between the responses of teachers in the third comparison group, the group with teachers interviewed during telephone follow-up. The majority of the significant differences for this group were to attitude and perception items, the same as we saw between CATI and NON-CATI respondents in the other two analysis groups.

Also, as seen in the other two groups, CATI treatment respondents reported more optimistic answers and the NON-CATI respondents reported more pessimistic answers.

Unlike the CATI and NON-CATI respondents in the other two groups, in the third group we isolated records by both designation mode and interview mode:

	Actual Interview Mode	
	CATI Treatment	NON-CATI Treatment
Group 1	all interviews	all interviews
Group 2	Mail	Mail
Group 3	CATI	Decentralized telephone and a very small number of CATI cases

The increase in the proportion of significant differences

between respondents interviewed by telephone may be attributable to data in the CATI group being collected by CATI and data in the other primarily by decentralized telephone interviews. This result suggests that the attitude data we collected in CATI and decentralized telephone interviews were different.

LIMITATIONS

We cannot attribute the differences we observed between the responses of teachers in the analysis groups we formed for this study solely to the mode in which their interviews were completed or to the method we used to assign teacher records for mail nonresponse follow-up. There are other errors, such as those due to estimation, coverage, processing, nonresponse, etc., which may have influenced our results.

TABLE 1
Questionnaire Items with Significant Differences

Interview Mode	Proportion of the Differences between CATI and NON-CATI Treatment Estimates that were Statistically Significant
All Interviews	18%
Mail Returns	14%
Telephone Follow-up Interviews	29%

TABLE 2
Results of Sign Rank Tests --- All Items

Interview Mode	P-value (PR ≥ |S|)
All Interviews	0.4092
Mail Returns	0.3782
Telephone Follow-up Interviews	0.7112

TABLE 3
Results of Sign Rank Tests --- Negative Responses To Attitude and Perception Items Only

Interview Mode	P-value (PR ≥ |S|)
All Interviews	0.0001
Mail Returns	0.0012
Telephone Follow-up Interviews	0.0019

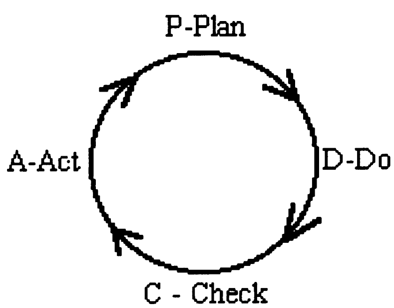
REINTERVIEW: A TOOL FOR SURVEY QUALITY IMPROVEMENT

Patricia Feindt, Irwin Schreiner, John Bushery, U.S. Bureau of the Census
Patricia Feindt, U.S. Bureau of the Census, Washington, DC 20233

Key Words: **Reinterview, quality improvement, response variance, interview mode**

1. Introduction

This paper discusses how reinterview programs can be a key component in efforts to improve survey data quality. The process of improving survey data quality is a continuous one and is analogous to the Plan-Do-Check-Act (PDCA) Cycle. This approach was first developed by Walter Shewhart and later referred to in Japan as the Deming Cycle (Sytsma).



This cycle describes the Census Bureau's efforts to continuously improve questionnaire design. The "Plan" is the cognitive research that goes into the development of the questions, the "Do" is the administering of the questionnaire in a survey setting, the "Check" is the reinterview evaluation, and finally, the "Act" is revising those questions with poor reliability. The cycle is then repeated in the next round of the survey to evaluate and further improve the revised questionnaire.

The Schools and Staffing Survey (SASS) provides an example of this cycle. The SASS is an integrated set of surveys including the Administrator, School, and Teacher Surveys. The surveys measure critical aspects of teacher supply and demand, the composition of the administrator and the teacher work force, and the general status of teaching and schooling in public and private elementary and secondary schools. The National Center for Education Statistics (NCES) sponsors the SASS. The Census Bureau first conducted the SASS during the 1987-88 school year and again during the 1990-91 and 1993-94 school years.

During each of these surveys the Census Bureau also conducted a reinterview to measure response variance for the administrator, school, and teacher surveys. By comparing original interview and reinterview responses, one can obtain a measure of response variance. In each reinterview, the reinterviewers re-asked a subset of questions from the original questionnaire. The questions selected were critical to the survey or suspected to be problematic. The results inform data users of the reliability of the questions and identify those that are problematic.

Generally, after problem questions are identified, cognitive research and other questionnaire design methods are used to make improvements. Then, a reinterview study in the next round of the survey can assess how much the revised questions improved reliability. The NCES and Census Bureau went through this process between the 1987-88 and the 1990-91 SASS and again between the 1990-91 and 1993-94 SASS.

While the results shown in this paper are from the SASS, this process has also been used in the National Household Education Survey (NHES) conducted by Westat for the NCES. Brick, et al. (1997) used this process on two Head Start questions identified as problematic from the reinterview in the 1991 survey. The questions were revised and reinterviewed again in the 1993 NHES and the changes made resulted in more consistent responses than the method used in the 1991 survey.

2.1 Reinterview Methodology

All the SASS surveys are conducted by mail, with telephone follow-up of nonrespondents. In 1994, the Census Bureau's Computer Assisted Telephone Interviewing (CATI) centers conducted the telephone follow-up operations.

Except for the 1994 School Survey, each of the SASS reinterview studies completed about 1,000 reinterviews, subsampled from cases completed in the original surveys. The 1994 School Survey completed about 550 reinterviews. Table 2.1a shows reinterview response rates by year for each of the three surveys.

Table 2.1a SASS Reinterview Response Rates

	1988	1991	1994
Administrator	87	94	82
School	87	91	62
Teacher	75	83	73

The 1988 SASS reinterviews were completed by telephone, no matter how the original interview was completed -- an imperfect replication of the original survey conditions. The 1991 SASS School Survey reinterviews exactly replicated the original interview modes -- mail and telephone. However, the 1991 Administrator and Teacher reinterviews were again conducted entirely by telephone. In 1994 all three reinterview studies replicated the original interview mode -- mail and CATI. Table 2.1b illustrates the interview and reinterview modes used in the three SASS surveys.

To determine the effect of question improvements, this analysis compares the response variance on questions changed from one survey to the next. However, changes in reinterview methodology complicated this analysis. The more recent reinterview studies better replicate the original surveys and produce more accurate estimates of response variance. Research by Bushery, Brick, Severynse, and McGuinness (1996) indicates that mail interviews can yield data with lower response variance than CATI interviews. This result suggests that telephone reinterviews of mail interviews likely overstate response variance. To avoid methodological differences confounding year-to-year comparisons, this paper compares only similar mode reinterview estimates of response variance. The changes in reinterview methodology prevent comparisons between the 1988 SASS and the 1994 SASS. These differences also limit the 1991 and 1994 comparisons in the Administrator and Teacher Surveys to those cases interviewed and reinterviewed by telephone. The shift from paper and pencil (PAPI) telephone interviewing in 1991 to CATI in 1994 also may affect the comparisons between these surveys, but the effect on the specific questions compared should be minimal. The effect of the change from PAPI telephone to CATI appears significant in the School Survey, however. Table 2.1c illustrates these comparisons.

This paper report results for 23 questions or subquestions revised after the SASS reinterview. Two were revised between 1988 and 1991 and 21 between 1991 and 1994.

2.2 Analytic Methods Used

Two statistics assess the reliability of reporting in this analysis: the gross difference rate and the index of inconsistency.

The gross difference rate is the percentage of cases with different responses in the two interviews and equals twice the simple response variance.

Table 2.1b SASS Interview and Reinterview Mode

Year/Survey	Original Mode	Reinterview Mode
1988		
Administrator/Teacher	mail	telephone
	telephone	telephone
1991		
Administrator/Teacher	mail	telephone
	telephone	telephone
School	mail	mail
	telephone	telephone
1994		
Administrator/ Teacher/ School[1]	mail	mail
	CATI	CATI

[1] The 1994 SASS School Survey reinterview excluded private schools interviewed by telephone.

Table 2.1c Year-to-Year Reinterview Comparisons

1988 versus 1991	
Administrator	
Full sample:	Original-mail/telephone Reinterview-telephone
Teacher	
Full sample:	Original-mail/telephone Reinterview-telephone
School	No eligible questions
1991 versus 1994	
Administrator	
Partial sample:	Original-telephone Reinterview-telephone
Teacher	
Partial sample:	Original-telephone Reinterview-telephone
School	
Full sample:	Original-mail/telephone Reinterview-mail/telephone

The index of inconsistency is a relative measure of response variability. In some circumstances, it estimates the proportion of the total variability due to random response error. Forsman and Schreiner (1991) give a more detailed discussion of these statistics. Table 2.2 shows the general format of the possible reporting outcomes from the original interview and the reinterview. The gross difference rate and index of inconsistency, formulated using the cells of this table, can be expressed as percentages,

$gdr(\%) = 100\ (b+c)\ /\ n$ and

$index(\%) = gdr(\%)\ /\ (P_o(1-P_r) + P_r(1-P_o))$,

where $P_o = (a+c)\ /\ n$ and $P_r = (a+b)\ /\ n$.

Table 2.2 General Format of Interview-Reinterview Results

Reinterview	Number of cases in Original Interview: With characteristic	Without characteristic	Total
With characteristic	a	b	a+b
Without characteristic	c	d	c+d
Total	a+c	b+d	n = a+b+c+d

The gross difference rate and the index of inconsistency apply to dichotomous questions. Each response category of "mark all that apply" questions is treated as a separate dichotomous variable. Finally, the aggregate gross difference rate and the aggregate index of inconsistency measure response variance in polytymous questions. The aggregate gross difference rate is the percentage of all cases reporting different responses in the two interviews. The aggregate index may be regarded as a weighted average of indexes across all categories of a question. U.S. Bureau of the Census (1985) describes these statistics in more detail.

All observations with missing responses to either the original or the reinterview were excluded from the analysis. Items with too few observations to estimate the index of inconsistency reliably also were excluded. The individual estimates of the index and the gross difference rate were compared using the Z-test. All comparisons were tested for significance at the 0.10 level.

3.1 A Comparison of the 1988 and 1991 Teacher and Administrator Surveys

We compare two problematic 1988 teacher subquestions, *"Bachelor's"* and *"Master's"* from the question, *"Which of the following college degrees have you earned?"* This question was revised and reinterviewed again in 1991. The new 1991 revised question showed significant improvement in response variance. The 1988 question provided a list of possible degrees and asked the respondent to "mark all that apply." In 1991 two "yes/no" questions, *"Do you have a bachelor's degree?"* and *"Do you have a master's degree?"* were asked, with a "mark all that apply" question for the remaining degrees (associate, doctorate, etc.). The response variance was substantially reduced for the two "yes/no" 1991 questions. However, the items that remained "mark all that apply" showed no improvement.

Administrators also were asked about "degrees earned." The revised 1991 "yes/no" questions showed similar improvement as the 1991 Teacher Survey. Table 3.1 shows the index of inconsistency and the gross difference rate (GDR) for these questions, for both the teacher and administrator surveys. A Z-test at significance level 0.10 shows the revised 1991 question has lower response variance than the 1988 question. Bushery et al. (1992) discuss these questions in more detail.

3.2 A Comparison of the 1991 and 1994 Administrator and Teacher Surveys

Five of the seven Administrator and Teacher questions revised and reinterviewed in 1994, improved significantly from 1991. The problematic "mark all that apply" question identified in the 1991 Administrator Survey, *"What other school positions, if any, did you hold before you became a principal?"* provided a list of six positions for administrators to choose from. Table 3.2 shows this list of positions.

The field interviewers were instructed to read the six positions, pausing after each, and mark all that applied. If the response is "yes" to *"Other Specify,"* then they were to fill in the response. If the respondent answered "no" to all six positions, then the *"None"* box was to be marked.

The analogous 1994 question had two parts. First CATI interviewers asked *"Did you hold any other school positions BEFORE you became a principal?"*. If the respondent answered "yes," the next question the CATI interviewers asked was, *"Did you hold the position of ______?"* A list of positions was provided with "yes/no"

boxes. The CATI interviewer had to mark the "yes" box or "no" box.

The 1994 revised questions divided the position, *"Department Head and Curriculum Coordinator"* into two separate positions. In addition, a new position, *"Library Media Specialist"* was added.

Four of the six 1991 subquestions (i.e., positions), showed significant improvement in their indexes in 1994 when the "yes/no" format was used. Five showed significant improvement in their GDRs. We hypothesize this improvement is due to the fact that changing the "mark all that apply" question to a series of "yes/no" questions forced the CATI interviewers to ask each position individually. In 1991 the interview was performed by paper and pencil (PAPI) and did not force the interviewers to ask each position individually. Table 3.2 provides response variance results for each subquestion.

The response variance for the 1994 teacher question, *"What Type of Certificate do you hold in this field?"* showed no improvement. This question was revised in 1994 by adding four more categories. The aggregate index was 39.5 in 1991 and 55.9 in 1994 (z = 1.4).

3.3 A Comparison of the 1991 and 1994 School Surveys

We compare the 1991 School Survey question, *"For what grade levels does this school offer instruction?"* to the 1994 question, *"How many students were enrolled in each of the grades shown on the front page, plus any ungraded levels, around the first of October?"* Both questions provided a list of grades ranging from *"ungraded"* to *"12th grade."* We analyzed 14 subquestions.

The instructions were revised slightly in 1994 by asking respondents not to include prekindergarten, postsecondary or adult education students. The instruction also asked the respondents to refer to their 'official fall report.'

In the CATI portion of the sample, ten of the 14 questions showed significant improvement in their indexes. Thirteen showed significant improvement in their GDRs. See table 3.3 for the results from the CATI portion of the sample.

In the mail portion of the sample, one of the 14 subquestions, *"12th Grade,"* significantly worsened in 1994 from having an index of 2.3 (1.1, 4.8) in 1991 to 7.3 (4.0, 13.5) in 1994 (Z= -1.9) and the rest showed no significant difference. See table 3.4 for results from the mail portion of the sample. Further, the responses to the "grade level offered" and "how many students enrolled" subquestions were not completely consistent. Often respondents failed to mark "grade level offered," but then reported a number of students enrolled. Sebron (1997) examined three grade levels: ungraded, fourth, and tenth. He found that between 37 and 49 percent of the mail respondents reported students enrolled in these grades, but failed to indicate that those grades were offered. Fortunately, the "grade level offered" information for these cases are taken care of during a consistency edit. The CATI part of the sample did not experience this inconsistency. The CATI instrument forced interviewers to answer the "grade level" subquestion before entering the number of students enrolled.

4. Conclusions

A series of "yes/no" questions almost always generates more reliable responses than a single "mark all that apply" question. The payoff in reliability provides some assurance that respondent burden is worth the extra effort. This result supports other work that suggest data quality is better when a series of "yes/no" questions are asked. Rasinski found that item nonresponse is lowered when individual "yes/no" questions are asked, rather than a "mark all that apply" question (Rasinski, Mingay, Bradburn, 1994)

The PDCA cycle has achieved some success in improving data quality in the SASS. This success has been limited because few questions have been reinterviewed a second time after being assessed by cognitive methods. Census and the NCES should develop a more comprehensive plan for continual questionnaire improvement.

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

References

Bushery, John M., Brick, J. Michael, Severynse, Jacqueline, and McGuinness, Richard A., (1996) "How interview mode affects data reliability," American Statistical Association, Proceedings of the Section on Survey Research Methods, Chicago, pp 600-604.

Bushery, John M., Royce, D., Kasprzyk, D., (1992) "The Schools and Staffing Survey: How reinterview measures data quality," American Statistical Association,

Proceedings of the Section on Survey Research Methods, Boston, pp 458-463.

Forsman, Gösta and Schreiner, I., (1991) "The design and analysis of reinterview: An overview," Measurement Errors in Surveys, Biemer, P. et al, editors, John Wiley & Sons, New York, pp 279-301.

Jenkins, Cleo R., (1992) "Questionnaire Research in the Schools and Staffing Survey: A Cognitive Approach," U.S. Census Bureau.

Rasinski, Kenneth A., Mingay, D., and Bradburn, N. "Do Respondents Really 'Mark All that Apply' on Self-Administered Questions:" *Public Opinion Quarterly*, 58:400-408.

Sebron, Gene A., (1997) "Reinterview Report: Response Variance in the Schools and Staffing Survey," U.S. Census Bureau.

Sytsma, S. (undated). "Practicing Continuous Improvement in the Classroom, "*http://www.sytsam.com/tqmpap.html.*

U.S. Census Bureau, (1985) Evaluating Censuses of Population and Housing, Statistical Training Document ISP-TR-5, Washington, D.C.

U.S. Department of Education. National Center for Education Statistics. *Reinterview Results for the School Safety and Discipline and School Readiness Components,* NCES 97-339 by J. Michael Brick, Lou Rizzo, and Jacque Wernimont. Project Officer: Kathryn Chandler. Washington, DC: 1997.

Acknowledgments

The authors are grateful to Carol Gunlicks, Andrew Zbikowski, and Brian Richards for providing valuable comments and to Gene Sebron for providing data.

Table 3.1 Summary of the Reinterview Reliability for the 1988 and 1991 Administrator and Teacher Survey

Which of the following degree have you earned?		Index of Inconsistency			Gross Difference Rate (%)		
		1988	1991	Z(diff)	1988	1991	Z(diff)
Teacher							
	Bachelors Degree	79.5	-	NA	7.5	0.6	6.8*
	Masters Degree	8.9	2.2	3.8*	4.3	1.1	3.7*
Administrator							
	Bachelors Degree	98.5	-	NA	20.3	1.3	14.5*
	Masters Degree	49.4	11.3	6.7*	9.9	1.7	7.4*

- too few cases to reliably estimate the index.
* significant at 0.10 alpha

Table 3.2 Summary of the Reinterview Reliability of the 1991 and 1994 Administrator Surveys

What other school positions did you hold before becoming a principal?	Index of Inconsistency			Gross Difference Rate (%)		
	1991	1994	Z (diff)	1991	1994	Z(diff)
Dept Head or curriculum coordinator	61.1	26.5	3.4*	23.5	13.2	1.8*
Assist Principal or program director	29.4	23.4	0.7	14.7	8.6	1.7*
Guidance Counselor	36.1	23.1	1.0	7.6	5.3	0.8
Athletic Coach	45.0	14.4	3.3*	16.5	6.6	2.8*
Sponsor for Student clubs	83.1	31.7	4.5*	31.2	15.9	3.3*
Other - Specify	94.6	57.2	3.7*	58.5	24.5	6.6*

* significant at 0.10 alpha

Table 3.3 Summary of the Reliability of the 1991 and 1994 Schools Survey of the Telephone/CATI Sample

For what grade levels does this school offer instruction?	Index of Inconsistency			Gross Difference Rate (%)		
	1991	1994	Z(diff)	1991	1994	Z(diff)
Ungraded	79.1	41.5	2.1*	11.1	8.6	0.9
Kindergarten	16.3	2.3	3.1*	8.5	1.1	3.5*
1st grade	17.0	3.5	2.9*	8.5	1.7	3.1*
2nd grade	16.3	3.5	2.8*	8.1	1.7	3.0*
3rd grade	17.8	3.5	3.0*	8.9	1.7	3.3*
4th grade	16.3	4.6	2.5*	8.1	2.3	2.7*
5th grade	16.3	4.6	2.5*	8.1	2.3	2.7*
6th grade	16.4	4.9	2.5*	8.1	2.3	2.7*
7th grade	16.0	9.4	1.3	7.8	4.0	1.7*
8th grade	16.7	9.3	1.4	8.1	4.0	1.8*
9th grade	11.1	3.7	1.9*	5.6	1.7	2.2*
10th grade	9.0	4.9	1.1	4.4	2.3	1.2*
11th grade	9.0	2.4	1.9*	4.4	1.1	2.0*
12th grade	7.5	2.4	1.6	3.7	1.1	1.7*

* significant at 0.10 alpha

Table 3.4 Summary of the Reliability of the 1991 and 1994 Schools Survey of the Mail/Mail Sample

For what grade levels does this school offer instruction?	Index of Inconsistency			Gross Difference Rate (%)		
	1991	1994	Z(diff)	1991	1994	Z(diff)
Ungraded	49.9	34.9	0.8	6.5	7.3	0.4
Kindergarten	5.7	2.8	1.3	2.8	1.4	1.3
1st grade	5.7	5.5	0.1	2.8	2.8	0.0
2nd grade	4.8	2.8	0.9	2.4	1.4	1.0
3rd grade	5.7	4.6	0.4	2.8	2.3	0.4
4th grade	6.1	7.3	0.4	3.0	3.7	0.5
5th grade	5.2	5.5	0.1	2.6	2.8	0.2
6th grade	4.8	6.6	0.6	2.4	3.0.2	0.6
7th grade	3.6	6.5	1.0	1.7	3.2	1.1
8th grade	4.0	5.5	0.6	2.0	2.8	0.6
9th grade	4.1	7.0	1.0	2.0	3.2	0.9
10th grade	4.3	6.1	0.6	2.8	0.6	0.6
11th grade	2.8	6.2	1.1	1.3	2.8	1.1
12th grade	2.3	7.3	1.9*	1.1	3.2	1.1

* significant at 0.10 alpha

IMPROVING THE COVERAGE OF PRIVATE ELEMENTARY-SECONDARY SCHOOLS

Betty J. Jackson, Nancy R. Johnson, Richard L. Frazier
Betty J. Jackson, Bureau of the Census, Washington, DC 20233

KEY WORDS: Data Collection, Evaluation, Traditional, Kindergarten-terminal

I. Purpose of this Presentation

This paper discusses the "traditional" universe of private elementary and secondary schools developed by the Census Bureau for the National Center for Education Statistics or NCES. This universe was initially developed in 1987, and subsequently updated five (5) times, with the sixth update currently in progress. Results of earlier updates have been previously reported, so this presentation focuses on the most recent updates, in 1995. Key results of the updates, including an analysis of the sources of added schools, their characteristics, and the impact of the adds on the universe will be discussed. Additionally, development of the Kindergarten-Terminal or K-terminal frame and results of the capture/recapture analysis will be discussed.

II. Background

As background, it is useful to provide definitions that pertain to this paper.

A. K-Terminal

A K-Terminal school contains an educational program primarily for 5-year-old children who will enter first grade in the upcoming school year. This includes transitional kindergartens and/or first grades if these children are expected to enter first grade upon completing these programs. Some of these K-Terminal programs may contain nursery or preschool age children. The 1995 PSS estimated approximately 7,300 private K-Terminal schools in the nation.

B. Private School Universe

It is useful to review the definition of the private school universe and how it is used. The private school universe is defined as including all schools that provide educational services for at least one of grades 1-12, have one or more teachers, are not administered by a public agency, and are not operated in a private home.

The private school universe is used in two major data collection efforts:

1) First, all of the schools on this universe are included in the Private School Survey or PSS. PSS is a census of private elementary and secondary schools conducted bi-annually for NCES beginning with the 1989-90 school year. PSS has a two-fold purpose.

    a) First, it generates bi-annual data on the total number of private schools, along with the number of students, teachers, and graduates at these schools.
    b) Second, the results it generates are used to build an accurate and complete list of private schools for NCES to use for other private school surveys.

    The 1995 PSS estimated that there are 27,686 private elementary-secondary schools in the nation.

2) The second major data collection effort using this universe is the Schools and Staffing Survey or SASS. SASS selects a sample of approximately 3,500 private schools from the private school universe.

It is also useful to discuss the methodology for compiling and updating the K-Terminal universe and the traditional private school universe.

C. Traditional Private School Universe

The <u>traditional private school universe</u> consists of two coverage improvement operations -- List Frame updating and an Area Search Frame. <u>List Frame updating</u> is a <u>national</u> coverage improvement operation designed to locate private elementary and secondary schools not already on the existing private school universe. The updating operation uses lists from private school associations, the 50 states and Washington, D.C., and private vendors. <u>Area Search Frame updating</u> is a coverage improvement operation consisting of an independent search in a nationally representative <u>sample</u> of counties. This operation is used to locate private schools still missing from the private school universe resulting after list frame updating.

As mentioned earlier, the private school universe was initially developed in 1987 with Quality Education Data

Incorporated (QED) providing us with a list of private elementary and secondary schools. List Frame updating was the first step in improving the coverage of this universe. For this update, 22 of the largest private school associations in the country were contacted and their lists of schools were requested. These lists were matched to the QED list and eligible non-matched schools were added to the universe.

The next step in improving the coverage of the first Private School Universe was area frame updating. For this update, a national sample of 75 PSUs was selected and field representatives were instructed to use up to ten (10) different sources such as the Yellow Pages to create an independent list of all private elementary and secondary schools in these sample areas. These independent lists were matched to the universe resulting from the list frame updating within each of the sample PSUs. The in-scope schools that did not match were weighted up to represent the schools that were missing from the updated list frame.

Since the initial development and updating of the private school universe in 1987, the universe has been updated every two years. In 1989, the List Frame updating was done using only 12 association lists due to budget constraints. We picked the lists based on the following criteria:

a) not too large
b) had a significant difference in the total number of schools reported between 1987 and 1989.

The Area Frame Updating in 1989 was done in a sample of 120 PSUs.

For updates in 1991, 1993, 1995, and 1997 we used many more lists for the list frame updating. These updates included lists from as many as 44 private school associations, the 50 states and Washington, D.C., QED and Josten's Education Data. For the area frame, we continued to use sets of 120 PSUs as we did in 1989.

D. K-Terminal Private School Universe

In 1993-94, we began to collect information on K-Terminal school programs and build a K-Terminal frame. As lists were collected from the 50 states and Washington, D.C., and Associations for the 1993 list frame updating and from the sample PSUs for the 1993 area frame updating, we began to identify and separate those programs that indicated that they contained at most a kindergarten or were primarily for 5-year-old children.

In 1995, the K-Terminal updating again consisted of a list frame updating and an area frame updating operation. In addition to what was done in 1993, more of an effort was made to contact states or other alternative private organizations to specifically ask for a list of their private kindergartens. This was added to the 1995 operation to evaluate alternative sources for lists of kindergartens and to improve coverage of schools containing kindergartens.

The results of the 1995 updating operations will now be presented. This analysis is done separately for the list frame updating, the area frame updating, and the K-Terminal operations. The 1995 results from both the list frame and area frame are contrasted with the 1993 results. Note that unless otherwise stated the results for 1995 are similar to those for 1993.

III. List Frame Updating Analysis

In 1995, we added about 2,400 in-scope schools to the universe during the "traditional"list frame updating. The corresponding 1993 number was 2,300.

In terms of the **sources** of the adds, 62% came from the state lists and 38% came from the association lists.

Overall, the state lists were most effective with a total of about 1,500 adds. As might be expected, the list from the state of California provided the largest number of adds (about 25% of the total of the state list adds). The next nine (9) states provided another 43% of the adds, such that the top 10 states accounted for about 2/3 of the state adds in 1995.

The list from Arkansas was the most effective list since about 16% of the schools on the list were in-scope adds. The next three most effective state lists had effectiveness rates above 12%. They were Tennessee, Montana, and Georgia.

The four least effective state lists were from Kansas, Iowa, North Dakota, and Oklahoma with no adds.

Association Lists were also effective, adding about 900 schools. We are not able to do any further detailed analysis of adds from association lists because the information was lost.

The **characteristics** of the adds will now be presented. Regarding the **religious orientation** of the added schools:

- 57% of schools were Other Religious schools
- 39% of schools were Nonsectarian schools

- 4% of schools were Catholic schools

In terms of the **grade level** of the added schools:

- 49% of schools were Elementary schools
- 40 % of schools were Combined schools
- 11% of schools were Secondary schools

In terms of the **size** of the added schools, we see that schools added from the list frame were predominantly small schools contributing 69%. The next largest schools contributed 18%, and the larger schools contributed at most 7%.

In terms of the **percent of minority students** at the added schools, we see that schools with the lowest minority percentage contributed the most to the list frame with 35%. Schools with the next highest minority percentage contributed 23%, and schools with the higher minority percentage contributed 32%.

In terms of the **school type** of the added schools, more than half (58%) are **regular elementary/secondary** schools. Each of the other school types contribute at most 18%.

Looking at the **impact** on the universe estimates, we find that, overall list frame adds represented:

- 8% of schools on the universe
- 3% of students on the universe
- 4% of teachers on the universe
- 1% of graduates on the universe

These percentages were close to what they were in 1993 with the exception of graduates where the impact on the universe was 3% in 1993.

The impact varied considerably for the **religious orientations** and showed that the list frame updating had a substantial impact on improving coverage of Nonsectarian and Other Religious schools and very little impact on Catholic schools.

- Nonsectarian schools led the way with 15% impact
- Other Religious schools followed with 10% impact
- Catholic schools had a minimal 1% impact

The impact for the **school grade levels** showed less variation and indicated that the list frame updating had an impact on improving the coverage for all 3 grade levels.

- Combined schools led the way with a 12% impact
- Secondary schools followed with a 9% impact
- Elementary schools were next with a 7% impact

The impact varied considerably for the different **sized** schools. An inverse relationship exists between the size of school and the size of this impact. The smallest schools had a 19% impact and the largest schools had a 1% impact.

The impact varied only slightly for schools with different **percent of minority students**.

In terms of the impact of **schools of different types**, we see that Voc. Tech., Montessori, and Alternative schools had at least a 24% impact whereas the other 4 had at most a 16% impact. This is somewhat different than what it was in 1993.

IV. Area Frame Updating Analysis

In 1995, we identified a weighted estimate of 2,386 in-scope area frame schools during the updating. The corresponding 1993 number was 2,026.

The **characteristics** of the adds will now be presented. Regarding the **religious orientation** of the added schools:

- 62% of schools were Other Religious schools
- 35% of schools were Nonsectarian schools
- 3% of schools were Catholic schools

In terms of the **grade level** of the added schools:

- 47% of schools were Elementary schools
- 49 % of schools were Combined schools
- 4% of schools were Secondary schools.

In terms of the **size** of the added schools, we see that schools added from the area frame were predominantly small schools contributing 77%. The next largest schools

contributed 12%, and the larger schools contributed at most 5%.

In terms of the **percent of minority students** at the added schools, we see that schools with the lowest minority percentage contributed the most to the area frame with 37%. Schools with the next highest minority percentage contributed 29%, and schools with the higher minority percentage contributed 20%.

In terms of the **school type** of the added schools, we see that 8 out of 10 schools are either **regular elementary/secondary or Alternative schools**. Each of the other types contribute at most 7% each.

The characteristics of the area frame adds were somewhat similar to those of the list frame adds for Religious Orientation, Grade Level, Enrollment, Percentage of Minority Students, and Type of School.

Looking at the **impact** of these adds on the universe estimates, we find that, overall, area frame adds represented:

- 8% of schools on the universe
- 3% of students on the universe
- 4% of teachers on the universe
- 1% of graduates on the universe

The impact varied considerably for the **religious orientations** and showed that area frame updating had a substantial impact on improving coverage of Nonsectarian and Other Religious schools and very little impact on Catholic schools.

- Nonsectarian schools led the way with 13% impact
- Other Religious schools followed with 11% impact
- Catholic schools had a minimal 1% impact

The impact for the **school grade levels** showed less variation and indicated that area frame updating had an impact on improving the coverage for all 3 grade levels.

- Combined schools led the way with a 14% impact
- Secondary schools followed with a 7% impact
- Elementary schools were next with a 4% impact

The impact varied considerably for the different **sized** schools. An inverse relationship exists between the size of school and the size of this impact.

The impact of schools with different **percent of minority students** varied only slightly.

In terms of the impact of **schools of different types**, we see that **Alternative and ECC/Daycare schools** have a combined 60% impact whereas the others have at most a 16% impact each. This is somewhat different than what it was in 1993.

The characteristics of the area frame adds were somewhat similar to those of the list frame adds for Religious Orientation, Enrollment, and Percentage of Minority Students.

V. K-Terminal Updating Analysis

Regarding the **religious orientation** of the added schools:

- 74% of schools were Other Religious schools
- 25% of schools were Nonsectarian schools
- 1% of schools were Catholic schools

In terms of the **grade level** of the added schools:

- 30% were Kindergarten only
- 70% were Kindergarten and less

Looking at the **impact** of these K-Terminal adds on the universe estimates, we find that, overall, these adds represented:

- 41% of schools on the K-Terminal universe
- 34% of students on the K-terminal universe
- 32% of teachers on the K-Terminal universe

The impact varied somewhat for the **religious orientations** and showed that the K-Terminal updating had a substantial impact on improving coverage of all 3 religious orientations.

- Nonsectarian schools led the way with 43% impact

- Other Religious schools followed with 37% impact
- Catholic schools had a smaller although significant impact at 25%

The impact for the **school grade levels** showed some variation as well and again indicated that the K-Terminal updating had an impact on improving the coverage for both grade levels.

- Kindergarten only schools led the way with a 58% impact
- Kindergarten and less schools followed with a 36% impact

VI. Capture/Recapture Analysis

In this section, the capture/recapture methodology and how it was used to estimate the number of schools on the 1995 PSS universe is discussed.[4] We will compare the capture-recapture estimate of the number of schools to the final weighted PSS estimate (traditional estimate) of the number of schools to estimate the coverage of private schools on the 1995 PSS universe.

The capture-recapture estimate is based on the following assumptions:

1. The list frame and area frame are independent of one another.
2. There are no out-of-scope records on either frame.
3. There are no duplicate school records.

   AND

4. The probability of observation of a school from a frame has the same expected value for all units.

This can be likened to estimating the number of fish in a pond. There is some unknown quantity (x) of fish. Draw a sample of ten and tag them. The probability of a tagged fish from this first sample is: P(t) = 10/x. Throw the tagged fish back into the pond and draw another sample of ten fish. This time there are 2 tagged fish and 8 untagged fish. Since P(t) is the probability of being tagged in the first capture, 10P(t) should equal the expected number of tagged fish in the recapture. Thus, 10P(t) = 100/x = 2 and solving for x, we estimate that there are 50 fish in the pond.

In the original list frame, 25,300 schools were "captured" and "tagged". Thus, the probability of inclusion in the list frame can be expressed as P(t) = 25,300/x where x is the population of private schools in the United States.

In the subsequent 2nd sample (area frame), 22,247 schools were "captured", of which 19,861 were "recaptured" or "already tagged". The "recaptured" schools were identified during the area search frame matching operation. Any area search frame school that matched to the list frame can be said to have been "recaptured".

So, 22,247P(t) = 22,247*(25,300/x) = 19,861. Solving for x reveals a capture-recapture estimate of private schools equal to 28,339. As noted earlier, the traditional estimate of schools is 27,686.

Thus, when comparing the traditional PSS estimate of schools to the capture-recapture estimate of schools, we estimate that the coverage of schools on the 1995 PSS universe to be 97.7%.

It's likely that the private school coverage has been overestimated based on the violation of assumption 1 (that the two frames are independent of one another) which was violated during the area frame operation.

Based on data presented in this paper assumption 4 is also violated to a certain extent. Violation of this assumption tends to underestimate the under coverage. Concerns about the validity of our coverage estimate due to violation of assumption 4 (that the probability of a school from a frame has the same expected value for all units) can be alleviated by poststratification. This is important particularly if steps are taken to address the violation of assumption 1. Poststratification involves computing a capture-recapture estimate for each one of a set of cells, with cells chosen to be correlated with the likelihood of being captured by a particular frame.

---

[1] A discussion of the model and assumptions as it applies to decennial census data can be found in Wolter, K.M. (1986): Some Coverage Error Models for Census Data. Journal of the American Statistical Association, 81,. 338-346.

Private Schools Survey-Capture/Recapture Estimates by Sets of Poststratification Cells

Sets of Poststratification Cells	Estimate
Typology	28,693
Grade Level	28,446
Religious Orientation	28,636
Religious Orientation Within Grade	28,637
TOTAL	28,339

The above table shows that the capture/recapture estimate from each of the four sets of poststratification cells is fairly close to the total capture/recapture estimate. Thus, the poststratification cells that give us the highest capture/recapture estimate will be used. Using the highest estimate would make it least likely that assumption 4 would be violated. Thus, the estimate used is given by the typology cells (28,693).

Thus, when comparing the traditional PSS estimate of schools to the capture-recapture estimate of schools using typology, we estimate that the coverage of schools on the 1995 PSS universe is 96.5%.

VII. Conclusions

1. First, the "traditional" list frame updating continues to be effective in improving the coverage of private schools - as it added about 9% to the universe.

2. Secondly, since the 1995 area frame estimated that we're still missing 8% of this universe, we need to continue our efforts in this updating to achieve a more complete universe of **all** private schools.

3. Thirdly, coverage improvement operations are especially needed for improving the coverage of small schools, Other Religious and Nonsectarian schools, and non-regular types of schools.

VIII. Additional Analysis

The following additional analysis will be done.

We will be factoring in the costs of these operations to do a cost-benefit analysis. These results can be used to develop future updating strategies for different budget scenarios - such as a tight or reduced budget.

We will look at additional K-terminal analysis.

We will also be analyzing the results of the 1997 PSS updating when they become available.

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:

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

The variables are as defined above.

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

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

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 $\mathrm{MSE}(x) = \mathrm{Var}(x) + \mathrm{B}^2(x)$ 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

Listing of NCES Working Papers to Date

Please contact Ruth R. Harris at (202) 219-1831
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User's Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97-25 (Aug.)	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97-26 (Oct.)	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
97-27 (Oct.)	Pilot Test of IPEDS Finance Survey	Peter Stowe
97-28 (Oct.)	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97-29 (Oct.)	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97-30 (Oct.)	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97-31 (Oct.)	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97-32 (Oct.)	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97-33 (Oct.)	Adult Literacy: An International Perspective	Marilyn Binkley
97-34 (Oct.)	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97-35 (Oct.)	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97-36 (Oct.)	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-37 (Nov.)	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97-38 (Nov.)	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97-39 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97-40 (Nov.)	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
97-41 (Dec.)	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman