Some Aspects Of The Non-Response Bias Study On The 2003-2004 School And Staffing Survey (SASS)[1]

Robyn Sirkis, US Census Bureau
Bac Tran, US Census Bureau
Phyllis Singer, US Census Bureau
Washington, D.C. 20233-8700
robyn.b.sirkis@census.gov

Abstract

Nonresponse bias can lead to the misinterpretation of published data. The National Center for Education Statistics requested that the Census Bureau conduct a Nonresponse Bias Study for the 2003-2004 School and Staffing Survey. Our discussion focuses on portions of the methodology applied and results, as well as future endeavors for nonresponse bias studies.

Keywords: Nonsampling, Nonresponse, Bias, Substantive Difference, SASS

1. Introduction

The primary purpose of the nonresponse study is to detect nonresponse bias in order to mitigate its effects on data. Nonresponse can cause bias, which can lead to misinterpretation of the published results. Nonresponse bias is a function of the response rate and the differences in responses between respondents and nonrespondents. A low response rate is a possible signal of nonresponse bias. Two types of nonresponse are unit and item nonresponse.

The Office of Management and Budget (OMB) has proposed guidelines that require nonresponse bias studies for surveys where the expected unit response rate is below 80 percent or the expected item nonresponse rate is below 70 percent. In those cases, a nonresponse bias analysis should be done to measure how accurately the information describes the target population of interest.

The National Center for Education Statistics (NCES) established an agreement with the U.S. Census Bureau to conduct a nonresponse bias study on the 2003-2004 School and Staffing Survey (SASS) using

an 85-percent cut-off (NCES 2002). In this paper, the 85-percent cutoff is used for teacher listing forms and an 87-percent is used cut-off for teachers. The nonresponse bias study developed by NCES and Census consisted of a nine-step procedure. This paper will discuss public and private teacher listing forms and teachers for four of these nine steps.

2. SASS Background

The purpose of SASS is to collect information necessary for a complete picture of American elementary and secondary education. The data collected permit detailed analyses of the characteristics of schools, principals, teachers, and students. SASS 2003-2004 is the fifth administration of this survey. For more details about SASS and nonresponse analyses of SASS see (Tourkin et al 2004) and (Bose 2001), respectively. See (Groves and Brick 2005) for more analysis of nonresponse bias. Details about definitions of response rates, cooperation rates, refusal rates, and contact rates can be found at www.aapor.org/pdfs/standarddefs_4.

Some of the key characteristics measured by SASS are teacher supply and demand; descriptive characteristics and staffing patterns of schools; workplace conditions and policies of schools and public school districts; education and work experience of teachers and principals; the principals' and teachers' opinions on the adequacy of facilities and support; demographic characteristics of students, teachers, and principals; and student participation in programs such as the National School Lunch Program.

2.1 Survey design

SASS uses a stratified probability sample design. Public and private schools were stratified based on certain characteristics. Then the teachers within the schools were stratified and sampled based on their characteristics.

---

[1] This report is released to inform interested parties of research and to encourage discussion. The views expressed on methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

Public schools received the Public School Questionnaire, Public School Principal Questionnaire, and the Public School Library Media Center Questionnaire. Public school districts received the School District Questionnaire. Private schools received the Private School Questionnaire and the School Principal Questionnaire.

The selected public and private schools received the teacher listing forms. The teacher listing forms collected the full list of teachers from the school, along with information on grade levels, subject matter taught, demographic characteristics and full- or part-time teaching status. The teachers were selected from the sample of teacher listing forms. Public teachers received the Public School Teacher Questionnaire. Private teachers received the Private School Teacher Questionnaire.

2.2 Sampling Frame and Sample Size

The universe for the 2003-2004 public school sample was the NCES 2001-2002 Common Core of Data. The universe for the private school sample was the 2001-2002 Private School Survey.

The sample included 10,368 public teacher listing forms, 3,622 private teacher listing forms, 53,188 public teachers, and 9,947 private teachers. This included interviews, noninterviews, and out-of-scope cases. About 15 percent of the teachers were selected from the teacher listing forms.

2.3 Method of Data Collection

The 2003-2004 SASS collected data using Field Representatives (FRs). An FR visited a school and administered a Computer Assisted Personal Interviewing (CAPI) instrument that determined whether the school was in-scope. The FR entered information on each teacher from a school provided roster or listing. The sample of teachers was then selected from the rosters. The FR distributed the paper questionnaires and was responsible for followup.

The schools were eligible if they provided classroom instruction in an academic subject. The teachers are eligible if they taught a class.

3. Methodology of the SASS 2003-2004 Non-Response Bias Study

The nonresponse bias study developed by NCES, and the Census Bureau consisted of a nine-step procedure. Some of these steps are similar to steps in (Brick and Bose 2001). This paper discusses Steps 1, 2, 3, and 6. The nine-steps are as follows:

Step 1: Compare response rates of each group to the overall response rates. The groups are determined by the answer categories of the frame variables.

Step 2: Generate and compare the unit response distributions to the frame using base weights.

Step 3: Identify the characteristics of nonresponse through logistic regression modeling.

Step 4: Use auxiliary data to validate the existence of nonresponse bias.

Step 5: Examine the last 5 percent of responses (late/difficult) to see whether the household and person-level demographic characteristics of late cases differ from the earlier interviews. Also, check to see if critical survey estimates would be different without late/difficult cases.

Step 6: Generate and compare the unit response distributions to the frame using final weights.

Step 7: Generate item response rates.

Step 8: Compare item response status by frame variables.

Step 9: Identify the characteristics of nonresponse of each item through logistic regression modeling.

3.1 Description of Data

The data were analyzed for the teacher listing forms and the teachers. The data for the public teacher listing forms and public teachers were grouped into those funded by the Bureau of Indian Affairs (BIA) and those not funded by the Bureau of Indian Affairs (Non-BIA), as requested by the NCES. The data were analyzed for the:

- Public teacher listing forms:
  - Public (BIA, non-BIA) and
  - Private
- Teachers:
  - Public (BIA, non-BIA) and
  - Private.

The teacher listing form files contained the school-level variables. The teacher files contained both the school-level variables and the teacher-level variables. Items analyzed for the public teacher listing forms (BIA and Non-BIA) included state, level of instruction (elementary, secondary, combined), locale of the school (rural, urban), and enrollment (number of students). Items analyzed for the private teacher listing forms were the same as the public teacher listing forms, except association membership was substituted for the state. Items analyzed for the teachers included the school-level variables plus teacher experience, teacher race, teacher subject, and full or part-time teaching status. Continuous variables were categorized using publication categories.

3.2 Weighting

The data were weighted for the teachers and the teacher listing forms. The three types of weights were:

- Basic weights,
- Replicate weights, and
- Final weights.

The basic weights were used to reflect the sample design at the unit level. The replicate weights were used to estimate the variances. The final weights were used to reflect the sample design and adjust for nonresponse.

In SASS, unit nonresponse is adjusted by using weighting cells, where the adjustments are defined as (weighted eligible/weighted interviews) for each cell. The cell definitions vary by questionnaire and sector. A separate adjustment is made for each stage of selection. For example, teachers are adjusted for teacher listing form nonresponse as well as sample teacher nonresponse. Weighting cells are defined using variables important to the survey estimation. For example, public teacher listing form nonresponse uses state, grade level, locale, and enrollment.

3.3 Response Rates

Response rates were computed using the basic weights for Steps 1, 2, and 3. The response rates were computed using the final weights for Step 6. Response rates were defined as the total number of sample units that responded divided by the total number of eligible sample units.

The potential for nonresponse bias was considered to be low if the response rates by answer category and the overall unit response rate were not significantly different. The potential for nonresponse bias was considered to be high if the response rates by answer category and the overall unit response rate were significantly different.

3.4 Hypothesis Tests

The NCES wanted to know when a significant difference in response rates (compared to the frame or compared to the unit response rate) would have an impact on the data. The NCES and the Census Bureau developed the concept of a substantive difference. Included in the calculations were standard errors and coefficients of variation (CVs).

3.4.1 Substantive Difference Test

The substantive difference test was developed to determine when significantly different response rates could have an affect on the data. Items were substantively different if the following conditions were all met.

- T-Statistic is significant,
- Relative Difference $\geq$ 10 percent,
- Absolute Difference is $\geq$ 1 percent,
- CV $\leq$ 15 percent, and
- n $\geq$ 30.

The absolute difference is the absolute value of the difference of two rates, as follows:

$$|\text{Rate overall} - \text{Rate at each category}|$$

The relative difference is the ratio of the absolute difference to the overall rate, as follows:

$$\text{Absolute difference/Rate overall}$$

3.4.2 Reason Not Noteworthy Test

The reason(s) an item with a significant t-test fails the substantive difference test is/are the reason(s) not noteworthy. The reason not noteworthy is indicated by assigning numbers 1-4. The numbers are as follows:

1=relative difference < 10 percent,
2 = absolute difference < 1 percent,
3 = CV > 15 percent, and
4 = n < 30.

3.5 Four Steps of the Nonresponse Bias Study

This section describes steps 1, 2, 3, and 6 as conducted for the analysis in this paper. This is only part of the overall analysis in the SASS 2003-2004 nonresponse bias study.

3.5.1 Step 1

The first step involved comparing the response rates of each group, broken down by frame variable, to the overall response rates. The groups were determined by the answer categories of the frame variables. The idea is that if the response rates for the groups are substantively different from the overall unit response rate, then this might be an indication of nonresponse bias.

The statistics that were included in step 1 are the:

- Unweighted number of eligible units,
- Estimate of the weighted response rate,
- Standard error of the response rate,
- T-statistic,
- Substantive difference test (indicated by shading and bolding in the tables), and
- Reason not noteworthy.

The unweighted number of eligible units is the sample size for each group. The estimates of the weighted response rates for each answer category, and for the whole sample, are computed. The t-statistic tested the response rate for the answer category against the overall response rate.

3.5.2 Step 2

The second step involved generating and comparing sample response distributions and sample estimates

to the frame distribution and frame estimates. Using the basic weight, which included the probability of selection and adjustments for the sample design, the distribution (estimates) of the sample respondents was compared to the distribution (estimates) of the sampling frame using any relevant frame variables adjusted for out-of-scope sample units. The frame distributions were adjusted for out-of-scope units by subtracting the weighted estimates of out-of-scope units in the sample from the frame totals. This produces a more realistic representation of the eligible sampling frame. The differences in estimates may signal sampling error and nonresponse error.

The statistics that were included in step 2 are the:

- Weighted proportion of the frame distribution adjusted for ineligible units,
- Standard error of the frame distribution adjusted for ineligible units,
- Proportion of sample distribution adjusted for ineligible units,
- Standard error of sample distribution adjusted for ineligible units,
- T-Statistic,
- Substantive difference test (indicated by shading and bolding in the tables), and
- Reason not noteworthy.

The weighted proportion of the frame distribution adjusted for ineligible units was estimated by weighting the ineligible units from the sample. The eligible units in category of the frame were obtained by subtracting the estimated ineligible units from all the units in category of the frame.

3.5.3 Step 3

The third step involved identifying the characteristics of nonresponse by using logistic regression modeling. Models were fit to the response status in order to identify predictive characteristics of nonresponse. The logistic model was used with the weight and sample design structure.

The Hosmer-Lemeshow Goodness-of-Fit Test Statistics and the R-square statistics were used to assess the model fit from the output from PROC LOGISTIC in SAS. The Hosmer-Lemeshow goodness-of-fit test statistics applied if the sample size was greater than or equal to 400. The R-square statistics applied if the sample was less than 400. The 5-percent significance level and 95-percent

confidence interval were used throughout the analysis.

Explanatory variables analyzed for the BIA and Non-BIA teacher listing forms were the state, census region, enrollment, level of instruction, locale of the school, and percentage of ethnic students. Explanatory variables analyzed for the private teacher listing forms were the association membership, census region, enrollment, level of instruction, and locale of the school.

The explanatory variables analyzed for the BIA and Non-BIA teachers were the school-level variables and teacher experience, teacher race, teacher subject, and teacher full or part-time teaching status. The explanatory variables analyzed for the private teachers were the school-level variables and teacher experience, teacher race, teacher subject, and teacher full or part-time teaching status.

3.5.4 Step 6

The sixth step involved generating and comparing sample response distributions and sample estimates to the frame distribution and frame estimates using final weights. The procedures for Steps 2 and 6 differed only in which weights were used.

The main idea was to determine if the weighting procedure was able to eliminate any potential bias that was identified in Step 2. Testing was conducting using final replicate weights to perform t-tests. If the final weights were created appropriately then there would not be a significant difference between the values.

3.6 Tables

The output for the teacher listing forms and teachers were displayed in main tables and additional tables. The main tables for the teacher listing forms consisted of the four school-level variables. The main tables for the teachers consisted of the school-level variables and the four teacher-level variables. The tables included standard errors. The CVs were not on the tables, but were included in the calculations for the substantive difference test. Examples of the tables are in the Appendix.

Additional tables were created for states (public) or association memberships (private) with low response rates. The additional tables for states or association memberships with low response rates contained the

same variables as the main tables. There were additional tables for the public teacher listing forms, public teachers, private teacher listing forms, and private teachers.

4. Results

The results indicated that many variables were statistically significant in Steps 1, 2, and 6, but not many of them satisfied the substantive difference test. Satisfying the substantive difference test was an indication of possible nonresponse bias. Summarizing Steps 1, 2, and 6, the following variables had answer categories satisfying the substantive difference test:

- State for BIA teachers in Steps 1 and 2, non-BIA teachers in Steps 1 and 6, and non-BIA teacher listing forms in Step 1;
- Association membership for private teachers in step 6 and private teacher listing forms in steps 1 and 6;
- Teacher Race for non-BIA teachers in Steps 1 and 2;
- Level of instruction for non-BIA teacher listing forms in Step 2;
- Locale of the school for private teacher listing forms in Step 2; and
- School enrollment non-BIA teacher listing forms in Step 2.

For step 3, which deals with logistic modeling, we have not found a model to fit the data yet. We will do more research in the future.

5. Limitations

There were limitations of both the nonresponse bias study process and the data. The factor adjustment did not seem to be appropriate when the response rate was low in category. Some of the statistics, such as the variance, could be unreliable when analyzing small groups.

The private school universe was not truly a universe. It was composed of the list frame and the area frame. The list frame is a universe. However, the area frame is a sample and thus not a universe.

There were limitations from some of the steps in the nonresponse bias study not analyzed in this paper. The analysts were unable to match the data with external data in order to obtain auxiliary variables for analysis. The analysis of the last 5 percent was not

implemented as expected due to the slowness of the paper transmission process. Also, the date stamp was not filled out by the respondents.

6. Conclusions

There needs to be additional research to determine how to minimize the potential bias that can result from nonresponse error. Cognitive reinterview will be conducted to determine if the wording, layout, or some other property of the questionnaire lead to nonresponse. Latent class analysis will be used to help model the nonresponse. Latent class analysis may help determine the mechanism of missing data and then estimate the missing cell counts. Once the model fits well, 'true' values under the model associated with the nonresponse will be estimated. Nonresponse bias is a function of those 'true' values.

7. Acknowledgements

The authors wish to thank Cornette Cole, Betty Jackson, Randy Parmer, Dennis Schwanz, and Toni Warner for sharing their knowledge about SASS and nonresponse bias studies.

8. References

Bose, J. (2001), "Nonresponse Bias Analyses at the National Center for Education Statistics," in *Proceedings of Statistics Symposium 2001: Achieving Data Quality in a Statistical Agency: A Methodological Perspective*.

Brick, J. M., and Bose, J. (2001), "Analysis of Potential Nonresponse Bias," Proceedings of the Annual Meeting of the American Statistical Association. August 5-9, 2001.

Groves, R. M., and Brick, J. M. (2005), "Practical Tools for Nonresponse Bias Studies." The Joint Program in Survey Methodology. June, 2005.

OMB, "Proposed Standards and Guidelines for Statistical Surveys," on the web at www.whitehouse.gov/omb/inforeg/proposed_sta ndards_ for_statistical_surveys.pdf.

NCES (2002), NCES Standard 4-4, on the web at nces.ed.gov/statprog/2002/std 4_4.asp.

Tourkin, S. C., Pugh, K. W., Fondelier, S. E., Parmer, R. J., Cole, C., Jackson, B., Warner, T., Weant, G., Walter, E., Gruber, K., and Zhao, L. (2004), "1999-2000 Schools and Staffing Survey (SASS) Data File User's Manual," on the web at nces.ed.gov/pubs2004/2004303.pdf.

Appendix – Tables

The tables below are portions of the tables submitted to the NCES for their nonresponse analysis. The tables illustrate sponsor requirements.

Cells in the t-statistic column of each table are highlighted if the t-statistic is significant. Rows are set in bold-face type if the value also satisfied the substantive difference test.

Table 1. Example of a Main Table from Step 1: Portions of the Table for BIA Teachers[2]

Response Rate Units		Unweighted Number of Eligible Units	Weighted Response Rate	Standard Error	T-statistic (Overall RR compared to Unit RR)	Reason not Noteworthy [3]
	Total [1]	683	92.3362	0.8708		
State	Arizona	162	97.0009	1.0565	-3.4071	1
	New Mexico	160	98.2598	0.4447	-6.0583	1
	South Dakota	108	96.6460	1.5247	-2.4546	1
	All Other States	**253**	**82.2775**	**2.1111**	**4.4046**	
Enrollment	Less than 100	42	97.6594	2.8920	-1.7625	
	100 to 199	171	92.4186	1.7046	-0.0431	
	200 to 499	335	91.1175	1.3675	0.7517	
	500 to 749	102	94.5442	2.0941	-0.9736	
	750 to 999	33	92.0895	4.5495	0.0533	
	Small town or rural	541	94.3943	0.8962	-1.6470	

---

[2] The entire BIA teacher table includes level of instruction, locale of the school, and the four teacher-level variables. The state was grouped into four categories.

[3] The reason not noteworthy column shows numbers 1 through 4 (even though only 1 is shown) if the value is statistically significant but does not satisfy the substantive difference test. The value of 1 indicates that the relative difference was less than 10 percent but the t-statistic was significant.

Table 2. Example of an Additional Table from Step 1: Portions of Additional BIA Teachers Table for "All Other States"[4]

	Response Rate Units	Unweighted Number of Eligible Units	Weighted Response Rate	Standard Error	T-statistic (Overall RR compared to Unit RR)	Reason not Noteworthy
All Other States	Total	253	82.2775	2.1111	0.0000	
	Elementary	**79**	**93.6268**	**2.1583**	**-3.7591**	
	Secondary	**74**	**93.1071**	**3.6030**	**-2.5933**	
	Combined	**100**	**68.0521**	**4.5668**	**2.8275**	
	Central city	16	76.2422	11.9312	0.4981	
	Urban fringe or large town	**55**	**68.8730**	**5.8638**	**2.1508**	
	Small town or rural	182	88.1761	2.3090	-1.8853	
	Less than 100	16	93.6051	6.9000	-1.5698	
	100 to 199	90	89.8339	2.7380	-2.1855	1
	200 to 499	114	75.9220	3.5360	1.5432	
	500 to 749	33	82.2765	6.7718	0.0001	

Table 3. Example of Main Table from Step 2: Portions of Table for Private Teacher Listing Forms [5]

		Frame Distribution (adjusted for ineligible units)		Interviewed Distribution		T-statistic (Frame versus Interviews)	Reason not Noteworthy
		Proportion	Standard Error	Proportion	Standard Error		
Association Membership	Catholic - Parochial	0.1477	0.0021	0.1533	0.0035	-1.3697	
	Catholic - Diocesan	0.1000	0.0015	0.1040	0.0030	-1.1830	
	Catholic - Private	0.0367	0.0007	0.0359	0.0017	0.4437	
	Amish	0.0222	0.0006	0.0246	0.0011	-1.8362	
	Assembly of God	0.0140	0.0009	0.0153	0.0014	-0.7742	
	Baptist	0.0673	0.0021	0.0596	0.0035	1.9016	
	Episcopal	0.0129	0.0004	0.0117	0.0010	1.1349	
	Jewish	0.0277	0.0011	0.0198	0.0013	4.5928	2
	Lutheran Church - Missouri Synod	0.0395	0.0020	0.0442	0.0050	-0.8585	
	Wisconsin Evangelical Lutheran Synod	0.0134	0.0003	0.0146	0.0005	-2.1217	1,2

---

[4] The BIA teachers had an additional table for the "all other states" answer category because the response rate of 82.3 percent was less than the 87-percent guideline. The entire table also includes the four-teacher level variables. Only low-responding states (public) and low-responding association memberships (private) have additional tables.

[5] The entire private teacher listing form table includes level of instruction, locale of the school, and school enrollment, as well as association membership. The association membership variable has more answer categories than shown on this example. There are significant, but not substantively different, answer categories shown on this part of the table. The tables for Step 6 have the same headings as this table but were filled using final weights rather than base weights.