

RESPONSE ERROR REINTERVIEW OF THE 1999-2000 SCHOOLS AND STAFFING SURVEY¹

David Miller² and Sharon Ennis³

ABSTRACT

The Census Bureau uses response error analysis to evaluate the effectiveness of survey questions. For a given survey, we select questions for analysis that are deemed critical to the survey or considered problematic from past analysis. Questions that are new to the survey or that are revised are prime candidates for reinterview. Reinterview is a new interview where a subset of questions from the original interview are re-asked of a sample of the survey respondents. For each reinterview question, we evaluate the proportion of respondents that give inconsistent responses. We use the "Index of Inconsistency" as our measure of response variance. Each question is labeled low, moderate, or high in response variance. In high response variance cases, the questions are put through cognitive testing, and modifications to the question are recommended.

For the response error analysis of the Schools and Staffing Survey (SASS) sponsored by The National Center for Education Statistics (NCES), we also investigate possible relationships between inconsistent responses and characteristics of the schools and teachers in the survey. Results of this analysis can be used to change survey procedures to improve data quality.

KEY WORDS: Response Error; Response Variance; Reinterview; Regression

1. INTRODUCTION

The Schools and Staffing Survey (SASS) is a collection of surveys conducted periodically by the National Center for Education Statistics (NCES) with the help of the Census Bureau. SASS consists of a School Survey, a Teacher Survey, an Administrator Survey, a Library Survey, and a Public School District Survey, all using one integrated sample design. The integrated design allows researchers to examine relationships among the different components of the educational system. These surveys collect data on a wide range of school-related matters.

SASS has been conducted in 1987-88, 1990-91, 1993-94, and 1999-2000. The Census Bureau has conducted a Response Error (RE) Reinterview of the schools' and teachers' surveys in conjunction with every round of SASS since its inception in 1987-88. The reinterview provides estimates of response variance for selected SASS questions by re-asking a subsample of the surveys' questions to a fraction of the surveys' respondents. We tend to select questions that are new to the survey, modified questions, and questions that seem possibly problematic. High response variance indicates a problematic question and moderate response variance also suggests some problems with reliability. This paper documents the 1999-2000 RE Reinterview of SASS and

1

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

² David E. Miller, United States Bureau of the Census, Washington, DC, United States, 20233

³ Sharon R. Ennis, United States Bureau of the Census, Washington, DC, United States, 20233

the results of our analysis of this reinterview data.

2. METHODOLOGY

The Census Bureau mails questionnaires to respondents. We mailed 13,451 SASS questionnaires to schools in 1999-2000 and 67,114 questionnaires to teachers. Some nonresponse follow-up is conducted by phone. Self-response is required for the teachers' surveys. The school principal is usually the respondent for the schools' surveys, but in some instances other school employees answer the questions.

2.1 Measures Used to Estimate Response Variance

The **gross difference rate (GDR)** and the **index of inconsistency (index)** are the principal measures of response variance used for categorical data. The **GDR** for an individual answer category is the percentage of responses that fall in the category in the original interview but not in the reinterview. For a single category, one-half of the GDR is the simple response variance. The **index** is a relative measure of simple response variance. It estimates the ratio of simple response variance to total variance for a question answer.

Overall estimates of the GDR and the index for a question, the **aggregate GDR** and the **aggregate index**, apply to questions with three or more answer categories. The aggregate GDR is the percentage of responses that change between the original interview and the reinterview. The aggregate index is similar to the index of inconsistency, but applies to the entire question rather than a specific answer category. It is an average index of inconsistency across all categories for the question.

An aggregate index of zero means responses were in perfect agreement, but an index of 100 does not mean that all of the respondents changed answers. Rather, it means that we saw what we could expect if there were no relationship between original and reinterview answers beyond chance agreement.

The GDR is more difficult to interpret than the index of inconsistency. Large GDRs indicate serious response variance in the data. Unfortunately, a small GDR is no guarantee of good consistency. In a low-frequency category, even a small GDR can represent high response variance relative to total variance. If this is the case, the index of inconsistency will tell us.

Pearson's correlation coefficient provides a measure of data reliability for continuous variables. When all the response variance model assumptions are met, the index is approximated by one minus the correlation coefficient between the original and reinterview responses ($1 - \rho$). We use the approximation $1 - \rho$ for quantitative data.

Use this rule of thumb to interpret the index of inconsistency, the aggregate index, and the Pearson's correlation coefficient:

Index Value	Response Variance Level	Interpretation
Less than 20	Low	Usually not a major problem
Between 20 and 50	Moderate	Somewhat problematic
Greater than 50	High	Very problematic

Any of these factors may cause high response variance:

- The methods used to collect the data may need improvement or the question may be unclearly written.
- The concept itself may not be measurable.
- Respondents may not be able to provide reliable information to the level of detail asked.

2.2 Limitations

2.2.1 Possible Bias

2.2.1.1 Nonresponse Bias

These reinterviews are separate surveys and are subject to nonresponse bias. A reinterview nonrespondent is someone who completed their questionnaire for the regular SASS survey, was selected for reinterview, but didn't respond to the reinterview questionnaire. Nonresponse bias exists here if the response variance of the nonrespondents differs from that of the respondents for the selected reinterview questions. Below are the reinterview response rates:

	Public Schools	Private Schools	Public School Teachers	Private School Teachers
n	3012	1146	2588	1270
Response Rate	75.9%	72.8%	70.6%	70.0%

2.2.1.2 Survey Mode Bias

In past versions of SASS, we did conduct a CATI reinterview on selected CATI original cases. However, due to budget constraints, we conducted the reinterviews solely by mail in 1999-2000. Some of the interviews in the original surveys were Computer Assisted Telephone Interview (CATI) cases and some were completed by follow-up in Census Regional Offices (ROs). For the schools' surveys, 6,994 out of the 11,043 (63.3%) completed surveys were mail questionnaires and the remainder were CATI or RO cases. In the teachers'

surveys', 41,487 of the 49,184 (84.4%) completed surveys were mail questionnaires. CATI and RO cases were ineligible for reinterview because the reinterview must replicate the original survey. However, by limiting our reinterview universe to those who responded to the original questionnaire by mail, we may be introducing bias into our reinterview results. The degree of this bias depends on the difference between mail and non-mail (CATI and RO) respondents in terms of their consistency in answering the survey questions⁴.

2.2.2 Capability of Reinterview

Reinterview programs allow us to detect problems in the questions, but usually cannot identify causes of response error, nor correct the problems.

3. RESULTS

We evaluated response variance in 95 questions from the 1999-2000 SASS Public School Survey and 81 questions from the 1999-2000 SASS Private School Survey. A majority of these questions were common to both surveys. For "mark all that apply" questions, each category was treated as a separate question. In the public school reinterview 42 percent of the questions displayed high response variance, suggesting poor reliability. Response variance was moderate for 45 percent of the questions. Of the 81 questions analyzed from the Private School Survey, 33 percent displayed high response variance and 47 percent showed moderate response variance.

We evaluated response variance in the same 57 questions from both the 1999-2000 SASS Public and Private School Teacher Surveys. For public school teachers, 44 percent of the 57 questions evaluated displayed high response variance. There was moderate response variance for 42 percent of the questions analyzed. For private school teachers, 26 percent of the 57 questions evaluated displayed high response variance and 54 percent displayed moderate response variance.

Survey	Total Evaluated	High	Moderate	Low
Public School	95 (100%)	40 (42%)	43 (45%)	12 (13%)
Private School	81 (100%)	27 (33%)	38 (47%)	16 (20%)
Public School Teacher	57 (100%)	25 (44%)	24 (42%)	8 (14%)
Private School Teacher	57 (100%)	15 (26%)	31 (54%)	10 (18%)

⁴ A comparison of response variance by respondent mode in the 1993 National Survey of College Graduates (NSCG) found that there tended to be higher response variance for CATI respondents than for mailout (Bushery, Brick, Severynse, and McGuinness, 1995).

4. MODELING DISCREPANCIES

As mentioned earlier, the function of RE reinterview is to detect problematic questions, but it has been generally unable to find causes of response error. In light of that, we expanded the role of reinterview with our analysis this year. We did this by exploring relationships between discrepancies (inconsistent responses between original interview and reinterview) and characteristics of the respondents (schools and teachers) for questions that showed high or moderate response variance. Our thinking, when we began this aspect of our analysis, was that we might be able to give our sponsor (NCES) ideas about causes of inflated response variance for specific questions and groups of respondents.

We performed a question-by-question analysis for the moderate and high response variance questions. We had in mind the long-term objective of modifications of questions if we uncovered specific problems. If a particular question is vague for a group or groups of respondents, then the question can be improved (not the respondents) and we wanted to find these situations if indeed they exist.

We ran regression analyses on every high or moderate response variance reinterview question in all of the reinterview surveys (Public School Reinterview, Private School Reinterview, Public School Teacher Reinterview, and Private School Teacher Reinterview). In our regression models for the categorical questions, the dependent variables were zero for a discrepancy and one for a match between original and reinterview. For the quantitative questions the dependent variables were the numerical differences between the original response and reinterview response. The regressors differed from the schools to the teachers.

Regressors for the School Questions - urbanicity, region, high school/non-high school, original interview respondent, number of students, student/teacher ratio, percent minority students, type of school, percent full time teachers, percent minority teachers, graduation rate, school for troubled students (yes/no), Title I school (yes/no), religious orientation (yes/no), accreditation (yes/no), tuition, lowest annual teachers' salary, highest teachers' salary, and the number of computers per student.

Regressors for the Teacher Questions - part-time/full-time, Master's Degree (yes/no), teaching assignment field, number of students, salary, gender, race, region, type of school, urbanicity, elementary/secondary/both.

We fit models for each regressor and each moderate or high response variance question, one regressor at a time, using simple linear regression for the quantitative response questions and logistic regression for the categorical response questions. We ran thousands of these regression models and, using an $\alpha = .01$, obtained 215 significant results.

4.1 Deceptive Results

When examining some of the individual significant regressor-question pairs, we found many that did not go along with what we were seeking. Many of them were significant due to differences in the regressor categories for the true response to that particular question, instead of being caused by a disparity in the effectiveness of the question for the regressor categories. It is easier to see with an example:

Question 16 in the Public School Questionnaire - *Does this school have any special requirements for admission other than proof of immunization, age, or residence?*

Regressor - High School/Non-High School⁵; **p-value** <.0001; n=1168

⁵ We consider a school a high school if they answer that they have students in one or more of grades 9-12. They are labeled a non-high school if they answer that they do not.

	Original (yes) ⁶	Reinterview (yes) ⁷	GDR
High School	13.1%	19.5%	14.4%
Non-High School	5.8%	8.7%	6.8%

As you can see from the GDR column, high school respondents were more than twice as likely to give inconsistent responses in the original and reinterview. This is why we get a significant result for the high school/non-high school regressor. It is also evident from the first two columns that having admissions requirements is a much more rare characteristic for non-high schools than for high schools. Rare characteristics tend to have few discrepancies because they are dominated by ‘no’ answers in both the original and reinterview. By the same token extremely common characteristics (near 100 percent) have few discrepancies because they are dominated by ‘yes’ answers in both surveys.

HS			Non-HS		
Reinterview	Original		Reinterview	Original	
	Yes	No		Yes	No
Yes	50	57	Yes	24	30
No	22	419	No	12	554

Of the 620 non-high schools that responded to this question in both the original interview and reinterview, 554 (89.4 percent) responded ‘no’ on both surveys.

In instances such as these, it can be this rare characteristic factor that leads to the disparate GDRs rather than a difference in the effectiveness of the question for the different regressor groups. One way to check this is to examine the indexes of inconsistency by regressor category. Recall that the index of inconsistency is a relative estimate of response variance. It divides the estimate of simple response variance (GDR/2) by an estimate of the total variance for the question answer. Dividing by our estimate of the total variance⁸ adjusts for this rare characteristic factor, since the total variance is smaller the further p is from one-half in either direction.

	Estimate of Simple RV	Estimate of Total Variance	Index	90% C.I.
HS	.0721	.1377	52.4%	(43.7, 62.3)
Non-HS	.0333	.0629	50.2%	(38.7, 64.6)

The indexes are not different enough to conclude there is a significant difference between the two regressor groups. So, it should not be inferred that question 16 is particularly problematic for high school respondents.

⁶ Percentage of respondents answering ‘yes’ to this question in the original interview.

⁷ Percentage of respondents answering ‘yes’ to this question in the reinterview.

⁸ Estimate of total variance = $(p_1q_2 + p_2q_1)/2$, where p_1 is the proportion of ‘yes’ answers in the original interview and p_2 is the proportion of ‘yes’ answers in the reinterview.

The large difference in GDRs appears to be due to the fact that more high schools have admissions programs than non-high schools.

Our list of significant results from our regression analyses is filled with such situations - where there is a disparate discrepancy rate but similar indexes of inconsistency among the regressor categories.

4.2 Notable Results

We have found some situations with significantly different indexes across regressor categories. Here is one of them:

Question 22 in the Public School Questionnaire - *Has this school implemented the following?*

a) *Scheduling of class periods to create extended instructional blocks of time (block scheduling)*

- 1 9 Yes
- 2 9 No

Regressor - High School/Non-High School; **p-value <.0001**; n=1170

	Original (yes)	Reinterview (yes)	GDR
High School	51.5%	50.8%	9.0%
Non-High School	40.3%	40.9%	19.6%

More than twice the rate of non-high school respondents switched answers as opposed to high school respondents. The high school proportion of yes answers is closer to .5 than the non-high school, but block scheduling does not seem to be a very uncommon characteristic for non-high schools either. Now, we look to the indexes:

	Estimate of Simple RV	Estimate of Total Variance	Index	90% C.I.
HS	.0449	.2499	18.0%	(14.2, 22.6)
Non-HS	.0979	.2412	40.6%	(35.2, 46.2)

Note that the confidence intervals do not overlap, so we conclude that this question is more problematic for non-high school respondents. The 18.0% index from the high school respondents would make this a low response variance question on our scale, but the 40.6% for non-high schools is at the upper end of the moderate response variance range.

Here is another instance of a question and regressor where the indexes were significantly different across regressor categories:

Question 40b in the Private School Questionnaire - Which of the following best describes the organization of classes in core subjects (math, science, social studies, English/language arts) for regular students in grades 9 -12?

- 1 9 Classes in ALL core subjects are differentiated by student ability level.
- 2 9 Classes in SOME core subjects are differentiated by student ability level.
- 3 9 Classes in cores subjects are NOT differentiated by ability level.
- 4 9 Not applicable; only one class per grade

Regressor - Accreditation (Yes/No) ; **p-value <.0001**; n=138

	Original (yes)	Reinterview (yes)	GDR
Accredited	(14.3%, 67.0%, 12.5%, 6.3%)	(13.4%, 63.4%, 15.2%, 8.0%)	25.0%
Not Accredited	(7.7%, 34.6%, 30.8%, 26.9%)	(19.2%, 42.3%, 23.1%, 15.4%)	69.2%

With four response categories, it is a little bit more difficult to see if the difference between the true values of the response between accredited and non-accredited schools is a factor here. The GDR for not accredited schools stands out, but we are dealing with a small sample size (n=26 for non-accredited schools). Let us look to the indexes:

	Estimate of Simple RV	Estimate of Total Variance	Index	90% C.I.
Accredited	0	.2660	47.0%	(34.8, 61.5)
Not Accredited	.3462	.3633	95.3%	(70.5, 100.0)

Despite very large confidence intervals, due to our small sample sizes, we can see that the indexes for accredited and non-accredited schools are significantly different. It appears that non-accredited schools have a more difficult time in answering this question.

We do not see why non-accredited schools would have a more difficult time with this question, nor do we see why the public school ‘block scheduling’ question would cause more problems for non-high schools. However, we cite these as two possible examples of what we sought out to find. It will require investigation by people with knowledge of the subject matter to determine if these examples can lead to improved questions.

4.2.1 Less Suitable Results

There were several other instances of significantly different indexes across regressor categories for questions, but they all seemed to be less likely to be fruitful. We made this determination by looking at the questions and regressors and concluding that it wasn’t sensible for these particular questions to be less clear to different groups of these regressors. These particular results may be interesting to NCES or worth noting, but seem unlikely to lead to modified questions. They involved male teachers not being as consistent in answering questions about the degree of seriousness of certain problems as female teachers, teachers with Master’s degrees not answering as consistently as those without Master’s degree on a question about computer usage in the classroom, and other examples of instances that don’t seem to point to problems in the questions.

4.2.2 Conservative Comparison Tests on Indexes of Inconsistency

Our test for differences in indexes of inconsistency uses a conservative estimate of the variance of the index. This may have lessened our number of significant results when comparing indexes.

5. CONCLUSIONS

We have provided our sponsor NCES with estimates of the response variance for SASS questions that we have reinterviewed. For a couple of these questions, we have pinpointed sources of inflated response variance that may lead to improved questions. Perhaps most importantly, we learned that when running our regression analyses, the significant results are not necessarily highlighting specific problems. Rather they are often demonstrating relationships between a regressor and the true value of a response to a question. Since this is not the aim of our analysis, we use comparison tests on indexes of inconsistency to attempt to uncover relationships between respondent characteristics and response variance.

REFERENCES

- Bushery, J., Brick, J., and McGuinness, R. (1995). "How Interview Mode Affects Data Reliability," Proceedings of the Section on Survey Research Methods, American Statistical Association.
- Bushery, J., Schreiner, I., and Sebron, G. (1998). "Response Variance in the 1993-1994 Schools and Staffing Survey: A Reinterview Report," unpublished report, Washington, DC: U.S. Bureau of the Census