

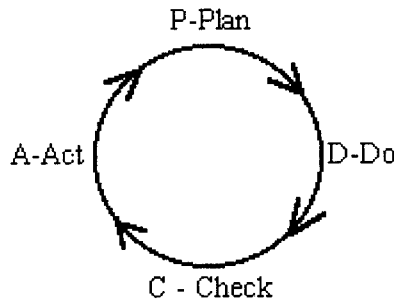
REINTERVIEW: A TOOL FOR SURVEY QUALITY IMPROVEMENT

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,

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

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

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		Total
	With characteristic	Without characteristic	
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 Jacques 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						
<i>Bachelors Degree</i>	79.5	-	NA	7.5	0.6	6.8*
<i>Masters Degree</i>	8.9	2.2	3.8*	4.3	1.1	3.7*
Administrator						
<i>Bachelors Degree</i>	98.5	-	NA	20.3	1.3	14.5*
<i>Masters Degree</i>	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)
<i>Dept Head or curriculum coordinator</i>	61.1	26.5	3.4*	23.5	13.2	1.8*
<i>Assist Principal or program director</i>	29.4	23.4	0.7	14.7	8.6	1.7*
<i>Guidance Counselor</i>	36.1	23.1	1.0	7.6	5.3	0.8
<i>Athletic Coach</i>	45.0	14.4	3.3*	16.5	6.6	2.8*
<i>Sponsor for Student clubs</i>	83.1	31.7	4.5*	31.2	15.9	3.3*
<i>Other - Specify</i>	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)
<i>Ungraded</i>	79.1	41.5	2.1*	11.1	8.6	0.9
<i>Kindergarten</i>	16.3	2.3	3.1*	8.5	1.1	3.5*
<i>1st grade</i>	17.0	3.5	2.9*	8.5	1.7	3.1*
<i>2nd grade</i>	16.3	3.5	2.8*	8.1	1.7	3.0*
<i>3rd grade</i>	17.8	3.5	3.0*	8.9	1.7	3.3*
<i>4th grade</i>	16.3	4.6	2.5*	8.1	2.3	2.7*
<i>5th grade</i>	16.3	4.6	2.5*	8.1	2.3	2.7*
<i>6th grade</i>	16.4	4.9	2.5*	8.1	2.3	2.7*
<i>7th grade</i>	16.0	9.4	1.3	7.8	4.0	1.7*
<i>8th grade</i>	16.7	9.3	1.4	8.1	4.0	1.8*
<i>9th grade</i>	11.1	3.7	1.9*	5.6	1.7	2.2*
<i>10th grade</i>	9.0	4.9	1.1	4.4	2.3	1.2*
<i>11th grade</i>	9.0	2.4	1.9*	4.4	1.1	2.0*
<i>12th grade</i>	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)
<i>Ungraded</i>	49.9	34.9	0.8	6.5	7.3	0.4
<i>Kindergarten</i>	5.7	2.8	1.3	2.8	1.4	1.3
<i>1st grade</i>	5.7	5.5	0.1	2.8	2.8	0.0
<i>2nd grade</i>	4.8	2.8	0.9	2.4	1.4	1.0
<i>3rd grade</i>	5.7	4.6	0.4	2.8	2.3	0.4
<i>4th grade</i>	6.1	7.3	0.4	3.0	3.7	0.5
<i>5th grade</i>	5.2	5.5	0.1	2.6	2.8	0.2
<i>6th grade</i>	4.8	6.6	0.6	2.4	3.0	0.6
<i>7th grade</i>	3.6	6.5	1.0	1.7	3.2	1.1
<i>8th grade</i>	4.0	5.5	0.6	2.0	2.8	0.6
<i>9th grade</i>	4.1	7.0	1.0	2.0	3.2	0.9
<i>10th grade</i>	4.3	6.1	0.6	2.8	0.6	0.6
<i>11th grade</i>	2.8	6.2	1.1	1.3	2.8	1.1
<i>12th grade</i>	2.3	7.3	1.9*	1.1	3.2	1.1

* significant at 0.10 alpha