

THE 1991-92 TEACHER FOLLOW-UP SURVEY REINTERVIEW AND EXTENSIVE RECONCILIATION

Cleo R. Jenkins and Angela-Jo Wetzel, Bureau of the Census
Cleo R. Jenkins, Bureau of the Census, Washington, D.C. 20233

Key Words: Measurement Error, Cognitive Research

I. INTRODUCTION

Traditionally, reinterviews have been designed for one (or more) of the following four purposes:

- to detect whether interviewers have deliberately falsified data,
- to evaluate interviewer performance,
- to estimate response variance, or
- to estimate response bias (Forsman and Schreiner, 1991).

Many reinterviews performed by the Census Bureau focus on estimating response variance. Although measuring response variance exposes inconsistencies in respondents' answers between interviews, it does little to explain why the inconsistencies occur.

Consequently, the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation was designed with a new objective in mind. Primarily, it focused on determining the reasons for respondent and instrument errors.

In this paper, we briefly describe the methods that were used to conduct this reinterview, followed by a discussion of both the methodology's benefits and limitations.

II. METHODOLOGY

A. Description of the 1991-92 TFS Reinterview Program

The Census Bureau conducted the 1991-92 TFS a year after collecting information from teachers in the 1990-91 Schools and Staffing Survey (SASS) for the National Center for Education Statistics (NCES). The TFS' purpose was to provide information about teacher attrition and to project teacher demand (Faupel et al., 1992). In general, the Census Bureau conducted the TFS Reinterview and Extensive Reconciliation two to three weeks after the TFS.

Both the TFS and the TFS Reinterview and Extensive Reconciliation contained two components: one for former teachers and another for current teachers. Each component had its own questionnaire (the TFS-2 for former teachers and the TFS-3 for current teachers), asking primarily different questions. The reinterview reasked a subset of questions from the TFS. The NCES chose the questions for reinterview. The Census Bureau offered suggestions, favoring factual over opinionated questions.

The TFS was a mixed-mode survey consisting of a first and second mail questionnaire, succeeded by a

telephone follow-up of mail non-respondents. The TFS Reinterview and Extensive Reconciliation was conducted exclusively by phone.

B. Development of the Extensive Reconciliation Probes

The use of an extensive reconciliation distinguishes this reinterview from others. It contained a series of probes aimed at identifying the reason for response differences and a reconciliation question to determine the correct response.

Closed-ended probes offered respondents specific reasons for differences. They were not the same from question to question, but tailored to each reinterview question. We used closed-ended probes to capture the data efficiently.

Two methods were used to develop the closed-ended probes:

- An expert analysis was conducted in which potential problems with the reinterview questions or possible reasons for differences between the two interviews were identified (see Forsyth and Lessler, 1991, for a discussion of this method).
- The findings of previous cognitive research with the 1990 Field Test Teacher Questionnaire (see Bates and DeMaio, 1990) were used. This information was especially helpful in identifying questions that might be susceptible to misinterpretation.

If the respondent did not choose one of the closed-ended probes, they were asked the open-ended probe: "Or was there some other reason [for the difference]?". The open-ended reasons were professionally reviewed and clerically coded prior to data entry.

C. Reinterview and Extensive Reconciliation Procedure

Working from a paper questionnaire, supervisory field representatives (SFRs) administered the TFS Reinterview and Extensive Reconciliation by phone. The SFRs received their instructions in a home self-study manual. The manual instructed them to first administer all of the reinterview questions. Immediately after completing the reinterview, the SFRs compared the respondents' reinterview responses with their original responses. The original responses had been transcribed to the reinterview questionnaires. Because the original responses were visible during the reinterview, this made it a dependent reinterview.

When a difference between the two responses occurred, the SFRs continued with the extensive reconciliation by asking the series of probes and the reconciliation question.

D. Sample Selection

Our goal was to obtain completed reinterviews for approximately 500 former and 500 current teachers. To achieve this goal, Demographic Statistical Methods Division (DSMD) randomly selected approximately 800 former teachers and 700 current teachers from the TFS sample files. DSMD oversampled to compensate for any non-response from the original interview and the reinterview. The 1992 TFS Reinterview and Extensive Reconciliation achieved a 92 percent completion rate (number of completed reinterviews (1314) divided by the number of eligible reinterview cases (1425)). We obtained completed reinterviews from 685 former teachers and 629 current teachers.

E. Analysis

We used two measures to analyze our reinterview data for this paper.

1. Gross Difference Rate (GDR)

The GDR is the proportion of responses that differ between the original interview and the reinterview. We calculated the GDR before reconciliation for the overall question. The GDR provides a rough idea of how consistently respondents answer a question.

2. Net Difference Rate (NDR)

The NDR is the difference between the percent of original responses in a specific answer category and the percent of reinterview responses in that category. We calculated a NDR after reconciliation for each answer category for a question.

The NDR shows the direction of change in responses for an answer category. We tested each NDR to see if it was significantly different from zero at the 90 percent confidence level. If the NDR is significant and positive, the answer category was over-reported in the original interview. If the NDR is significant and negative, the answer category was under-reported in the original interview.

III. RESULTS AND DISCUSSION

A. Benefits of the Methodology

The reinterview and extensive reconciliation produced some meaningful information from which we were able to make recommendations for either improvements or further research for a number of the TFS questions. We identified 19 of the 49 reinterview questions as problematic. We considered a question problematic if 1) one or more of its answer categories had a significant NDR or 2) it had one or more notable reasons for response differences. Refer to Jenkins and Wetzell (1994a) for a complete analysis of each reinterview question.

In this paper we illustrate two types of problems that we were able to uncover: 1) comprehension and 2) information storage or retrieval.

1. Comprehension Problems

Respondents demonstrated difficulty understanding the meaning of some questions. We illustrate this using two questions: the grade level and the teaching assignment question. We present the original question followed by our recommendations for improving it. We offer the supporting data in a table that includes:

- the GDR before reconciliation,
- each answer category that has an after reconciliation NDR significantly different from zero at the 90% confidence level, and
- the complete list of respondents' answers to the series of probes.

a. The Grade Level Question:

In what grade levels are the students in your classes at THIS school?

The intent of this question is to learn what the grade levels are of all the students that the teacher teaches. Respondents were supposed to mark all grade levels that applied. For our analysis, we considered each of the 16 answer categories shown in Table 1 as a separate question with two possible answer categories: marked and unmarked.

Respondents demonstrated difficulties understanding the wording of this question. The NDRs in column 3 of this table suggest that respondents tended to overreport students in the 4th through 8th grades in the original interview. Respondents' reasons for inconsistent answers given in part 2 shed some light on this result:

- One-third (15) reported misunderstanding some aspect of the question. Specifically, four reported misunderstanding what was meant by "grade level" or "class." Another five were uncertain whether they should report the grade levels of students they sometimes teach or classes with only a few students. Six simply reported misunderstanding the question as a whole.
- Three respondents had difficulty because they taught special students. These respondents either had trouble reporting the equivalent grade levels for the students, or they were not certain whether they should report them as ungraded or in their equivalent graded level.

The reasons respondents gave for differences suggest that if the intent of this question is to learn what the grade levels are of all the students that the

teacher teaches, regardless of whether the student is in a formal "class" or not, then the question should be reworded: **In what grade levels are the students that you teach at THIS school?** This wording eliminates the confusing word "class," the definition of which gives respondents problems. Does a class need to meet regularly to be considered a class? Does it need to be a certain size before it qualifies as a class? Respondents are not certain of the answers to these questions.

b. The Teaching Assignment Question:

Which of the following categories best describes your teaching assignment?

- Regular full-time or part-time teacher**
- Itinerant teacher (i.e., your assignment requires you to provide instruction at more than one school)**
- Long-term substitute (i.e., your assignment requires that you fill the role of a regular teacher on a long-term basis, but you are still considered a substitute)**

In this question, respondents reported having difficulty with the question's wording and the answer categories. Part 3 of Table 2 shows that half (6) of the respondents who gave a reason for inconsistent answers said they misunderstood the question or thought the answer categories were confusing. The NDRs in part 2 of Table 2 suggest that the problem lies with the first two answer categories. Respondents tended to overstate being a regular full- or part-time teacher (1.6%) in the original interview, while understating being an itinerant teacher (-1.5%).

A possible explanation for this is that respondents chose the first answer category because they thought it fit their situation well enough. Perhaps they cued in on the words "full-time or part-time teacher," while overlooking, ignoring, or not understanding the word "regular." Without this word, itinerant and long-term substitute teachers might reasonably mistake themselves for full- or part-time teachers. This behavior of selecting the first response alternative that seems to constitute a reasonable answer is discussed by Krosnick (1991).

The word "itinerant" may be another problem. Cognitive research with the Public School Questionnaire revealed that many respondents did not know what an "itinerant" teacher was (Jenkins et al., 1992a, p. 26). They knew "itinerant" teachers by other names, including traveling, co-op, and satellite teachers.

Based on these results, we suggest the following changes to this question:

- Reorder the answer categories. The itinerant and long-term substitute teachers are more likely to consider themselves regular full- or part-time teachers than vice versa.
- Reword the "itinerant teacher" answer category. State the definition of "itinerant teacher" first, then the technical term in parentheses, instead of vice versa.
- Provide a more comprehensive list of familiar names for itinerant teachers, such as traveling, co-op, or satellite teachers.

Our suggested order and wording are:

- You provide instruction at more than one school (i.e., you are an itinerant, traveling, co-op, or satellite teacher).**
- You fill the role of a regular teacher on a long-term basis, but you are still considered a substitute (i.e., you are a long-term substitute teacher).**
- You are a regular full-time or part-time teacher.**

2. Information Storage or Retrieval Problems

Respondents demonstrated difficulty obtaining information to answer some questions. We illustrate this using two questions: the base year salary and the family income question. Again, we present the original question followed by our recommendations for improving it.

a. The Base-Year Salary Question:

The following questions refer to your before-tax earnings from teaching and other employment from the summer of 1991 through the end of the 1991-92 school year.

Record earnings in whole dollars.

DURING THE CURRENT SCHOOL YEAR--

What is your academic base year salary for teaching in this school?

This question requests a monetary value. The before reconciliation disagreement rate (14.8%) in part 1 of Table 3 shows that respondents had difficulty reporting this value. (According to reinterview instructions, the dollar values disagree if they exceed a \$1,000.00 difference.) Part 2 of Table 3 shows that the predominant reason for monetary differences is that respondents were unsure of the exact amount of their earnings. This suggests that respondents do not have an easily accessible, precise figure stored in

memory to accurately answer this question. It also suggests an inability or unwillingness on the respondent's part to look up appropriate records which may exist.

We discuss these problems further after looking at the results from the next question.

b. The Family Income Question:

Which category represents the total combined income (include your own income) of ALL FAMILY MEMBERS age 14 and older in your household during 1991? Include money from jobs, net business or farm income, pensions, dividends, interest, rent, social security payments, and any other income received by family members in your household.

- less than \$10,000
- .
- .
- .
- .
- \$100,000 or more

This question requests categorical data. The GDR (16.2 percent) in part 1 of Table 4 is the largest of any of the closed-ended questions. Part 2 shows that nearly half (41) of the respondents who gave a reason for inconsistent answers said they were unsure of the exact amount. Again, this suggests that they do not have an easily accessible, precise figure stored in memory to accurately answer the question.

The fact that respondents had difficulties consistently answering an income question whether it requested a monetary value (base-year salary) or categorical data (family income) does not appear simple to solve. Initially we thought that asking respondents either 1) to obtain records to accurately answer the income questions or 2) to stop and think about them more carefully might be possible solutions to this problem. However, we now believe this to be a naive perspective. According to a recent experimental treatment, requiring the use of personal records may decrease response rates and increase follow-up costs without a large enough improvement in answer quality (Marquis, 1993).

We need to have a better understanding of respondents' use of records before we will be able to properly guide this process. Jenkins (1992b) concludes that respondents' use of records is one of the most complex areas of questionnaire research to study, since it requires in-depth knowledge about respondents' records as well as how they use those records. Perhaps asking respondents to gather appropriate records is more feasible with a self-administered

questionnaire than other modes of administration. Certainly this is an area in need of further research.

Since asking respondents to use their records may have a detrimental effect on the data in other ways (i.e., increased nonresponse), the question becomes just how much measurement error in the data can the sponsor tolerate. Although responses to the family income question differ, they do so by a limited amount. A crosstabulation of inconsistent answers between the reinterview and original interview shows that almost 60 percent of them are due to respondents choosing answer categories that are next to each other in the two interviews. For instance, a respondent might choose the answer category \$15,000-\$19,000 in the original interview and \$20,000-\$24,000 in the reinterview, or vice versa.

B. Limitations of the Methodology

We believe the 1991-92 TFS Reinterview and Extensive Reconciliation had shortcomings involving the dependent-type reinterview and the closed-ended probes. Jenkins and Wetzel (in press) contains a complete report of the reinterview and extensive reconciliation's methodology and our recommendations for improving it.

1. The Dependent-Type Reinterview Produced Too Few Differences

In general, the 1991-1992 TFS Reinterview and Extensive Reconciliation produced too few differences. There are fourteen questions from the reinterview and extensive reconciliation that are the same as those from the 1989 TFS Reinterview, and all but two of them have before reconciliation GDRs significantly lower than their 1989 counterpart at the 90% confidence level. Evidence also exists from past research that dependent reinterviewing results in fewer differences (Schreiner, 1980; Koons, 1973).

Because of the low GDRs, our counts for specific reasons for differences are very small at times. This can be seen in the numbers we discuss in the previous section (Results and Discussion).

The 1989 and 1992 surveys had two major differences:

- The 1989 methodology used an independent reinterview, whereas the 1992 methodology used a dependent-type reinterview.
- The 1989 methodology used FRs in both the original and reinterview. In contrast, the 1992 procedures specified that SFRs conduct the reinterview.

We hoped that SFRs would be more likely to ignore the original response than FRs. The data suggest, however, that this was not the case and that the lower GDRs are due to the reinterview's dependency.

2. The Extensive Reconciliation Produced Too Many Open-ended Responses

Approximately 54% of the total number of reasons for differences were open-ended. This unexpectedly high percentage suggests that the series of closed-ended probes did a relatively poor job of providing respondents with adequate reasons for differences in their responses.

3. The Extensive Reconciliation Produced Too Many General Responses

An even larger deficiency with the extensive reconciliation was that respondents did not adequately verbalize the reasons for differences in their answers when the closed-ended questions did not apply. Approximately 43% of the open-ended responses were "don't know" or "misunderstood question." This is a much more serious error than obtaining open-ended responses that could be coded to specific reasons. The general responses led to the omission of useful data.

IV. CONCLUSION

The 1991-92 TFS Reinterview and Extensive Reconciliation represents the Bureau's first attempt to employ an extensive structured reconciliation. The ultimate goal was to identify problematic questions, to identify the sources of the problems, and to offer suggestions for improving the TFS questionnaires.

As demonstrated in this paper, we were able to identify some problem questions, particularly those exhibiting comprehension and information storage/retrieval difficulties. Moreover, we gained enough insight from the reinterview and extensive reconciliation to make recommendations for either improving the questions or for further research.

However, there were some methodological shortcomings. We showed that the reinterview and extensive reconciliation produced too few differences and, hence, too few reasons for differences between the original and reinterview responses. We believe this occurred because the reinterview was not independent from the original interview. In the future we strongly suggest employing: (1) an independent reinterview followed by a third visit small-scale unstructured extensive reconciliation, or (2) an independent reinterview followed by a large-scale extensive reconciliation using Computer Assisted Telephone Interview (CATI). We make these suggestions without having evaluated cost or respondent burden. However, given the correct methodology, the reinterview/extensive reconciliation may become an effective questionnaire evaluation technique.

NOTES

1. The SASS is a relatively new set of integrated surveys first launched in the 1987-88, 1990-91, 1993-94 school years, and scheduled every four years hence.

REFERENCES

- Bates, N. and DeMaio, T.J. 1990. "Report on Cognitive Research on the Public and Private School Questionnaire for the Schools and Staffing Survey (SASS-3A and SASS-3B)." Internal Census Bureau memorandum. July 11, 1990.
- Faupel, E., Bobbit, S. and Rohr, C. 1992. "1988-89 Teacher Followup Survey Data File User's Manual." National Center for Education Statistics. U.S. Department of Education. Office of Educational Research and Improvement. June 1992.
- Forsman, G. and Schreiner, I. 1991. "The Design and Analysis of Reinterview: An Overview" in P.B. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds), Measurement Errors in Surveys. New York: John Wiley and Sons, pp. 279-301.
- Forsyth, B.H. and Lessler, J.T. 1991. "Cognitive Laboratory Methods: A Taxonomy" in P.B. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds), Measurement Errors in Surveys. New York: John Wiley and Sons, pp. 393-418.
- Jenkins, C. and Wetzel, A. 1994a. "The Results of the 1992 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation: Part One of a Two-Part Report. A report submitted to the National Center for Education Statistics, Census Bureau Memorandum, March 22, 1994.
- Jenkins, C. and Wetzel, A. In press. "An Evaluation of the Methodology Utilized for the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation: Part Two of a Two-Part Report.
- Jenkins, C., Ciochetto, S., and Davis, W. 1992a. "Results of Cognitive Research on the Public School Questionnaire for the Schools and Staffing Survey (SASS-3A)." Internal Census Bureau Memorandum. June 15, 1992.
- Jenkins, C. 1992b. "Questionnaire Research in the Schools and Staffing Survey: A Cognitive Approach." Proceedings of the Section on Survey Methods Research, American Statistical Association, August 1992.
- Koons, D.A. 1973. Quality Control and Measurement of Non-sampling Error in the Health Interview Survey. Washington, DC: National Center for Health Statistics. Series 2. Number 54 DHEW Publications No. HSM 73-1328. March, 1973.
- Krosnick, J.A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys" Applied Cognitive Psychology. 5:213-236.
- Marquis, K. 1993. Personal Communication. Preliminary Results from a Cognitive Approach to Redesigning Measurement in the Survey of Income and Program Participation. US Bureau of the Census. Washington, DC.
- Schreiner, I. 1980. "Reinterview Results from the CPS Independent Reconciliation Experiment (Second Quarter 1978 through Third Quarter 1979). Internal Census Bureau memorandum. May 7, 1980.

Table 1. Grade Level Question - 629 Responses

Part 1. GDR, Significant NDR's and Confidence Limits (%)			
Category	GDR	Limits	NDR Limits
Ungraded	0.2	(-0.1, 0.4)	
Prekindergarten	0.6	(0.1, 1.2)	
Kindergarten	1.9	(1.0, 2.8)	
1st	2.5	(1.5, 3.6)	
2nd	3.0	(1.9, 4.1)	
3rd	2.5	(1.5, 3.6)	
4th	2.9	(1.8, 4.0)	
5th	3.2	(2.0, 4.3)	1.3 (0.1, 2.4)
6th	1.9	(1.0, 2.8)	1.6 (0.5, 2.7)
7th	2.7	(1.6, 3.8)	1.0 (0.1, 1.8)
8th	2.7	(1.6, 3.8)	1.4 (0.4, 2.5)
9th	2.5	(1.5, 3.6)	1.7 (0.7, 2.8)
10th	2.1	(1.1, 3.0)	
11th	1.7	(0.9, 2.6)	
12th	1.9	(1.0, 2.8)	
Postsecondary	0.5	(0.0, 0.9)	

Part 2. Reasons for Difference between Responses			
Reason	Count	Percent	
Total	49	100.0	
Don't know	16	32.7	
Misunderstood question	6	12.2	
Unsure whether to report level of classes sometimes taught or with few students	5	10.2	
Teaching different students since responding	4	8.2	
Misunderstood what "grade level/class" meant	4	8.2	
Forgot/remembered info	4	8.2	
FR error	3	6.1	
Teach special students - difficulty reporting/unsure whether to report equivalent grade levels	3	6.1	
Other	2	4.1	
Misunderstood reference period	2	4.1	

Table 2. Teaching Assignment Question - 610 Responses

Part 1. Gross Difference Rates and Confidence Limits (%)		
No. of Categories	GDR	Limits
3	2.0	(1.0, 2.9)

Part 2. Significant NDRs and Confidence Limits		
Answer Category	NDR	Limits
Regular full/part-time teacher	1.6	(0.7, 2.6)
Itinerant teacher	-1.5	(-2.4, -0.6)

Part 3. Reasons for Difference between Responses			
Reason	Count	Percent	
Total	13	100.0	
Misunderstood question	3	23.1	
Category problems	3	23.1	
Situation changed since responding	2	15.4	
Don't know	2	15.4	
FR/Manual/general error	2	15.4	
Forgot/remembered info	1	7.7	

Table 3. Base-Year Salary Question - 629 Responses

Part 1. Disagreement Rate and Confidence Limits (%)		
No. of Categories	Rate	Limits
2	14.8	(12.5, 17.1)

Part 2. Reasons for Difference between Responses		
Reason	Count	Percent
Total	109	100.0
Unsure of exact amount	71	65.1
Salary changed since responding	9	8.3
Don't know	9	8.3
Fr/manual/general error	5	4.6
Included other salary earnings	4	3.7
Misunderstood question	3	2.8
Included another source of income	2	1.8
Forgot/remembered info	2	1.8
Misunderstood reference period	2	1.8
Unsure how to report as an itinerant teacher	1	0.9
Gave after-tax earnings	1	0.9

Table 4. Family Income Question - 604 Responses

Part 1. Gross Difference Rate and Confidence Limits (%)		
No. of Categories	GDR	Limits
13	16.2	(13.8, 18.7)

Part 2. Reasons for Difference between Responses		
Reason	Count	Percent
Total	84	100.0
Unsure of exact amount	41	48.8
Don't know	11	13.1
Unsure what to include/exclude	8	9.5
Misunderstood reference period	7	8.3
FR/manual/general error	5	6.0
Wasn't sure whether to include adult children	4	4.8
Misunderstood question	2	2.4
Refused to answer in one interview	2	2.4
Other	1	1.2
Missed skip pattern/question	1	1.2
Forgot/remembered info	1	1.2
Misread question	1	1.2