AN OVERVIEW OF NCES SURVEYS REINTERVIEW PROGRAMS

Valerie Conley, Steven Fink, Mehrdad Saba, Synectics for Management Decisions Inc.,
Steven Kaufman, National Center for Education Statistics
Valerie Conley, Synectics for Management Decisions, Inc., 3030 Clarendon Blvd #305, Arlington VA 22201

KEY WORDS: Reinterview Programs, Simple Response Variance, Gross Difference Rate, Index of Inconsistency, Response Bias

1. Introduction

The National Center for Education Statistics (NCES) conducts a variety of programs to assess the quality of the data it collects in its surveys. Traditionally, the emphasis has been to estimate nonsampling error components in a survey model: reinterview programs and validity evaluations are part of the overall survey design for most of its complex sample surveys. This paper highlights the NCES application of the reinterview while serving also as an overview of the techniques and methods that quantify measurement error which are used in NCES data quality assessment.

2. Programs Purpose and Background

A reinterview -- replicated measurement on the same unit -- is a new interview which repeats all or a subset of the original interview questions. At the 1991 American Statistical Association (ASA) meeting, the National Agricultural Statistics Service (NASS) presented a paper which traced the history of reinterview studies at NASS. The authors concluded that "An important product of reinterview surveys has been the identification of reasons for reporting errors. These include definitional problems, misinterpretation of questions and survey concepts, and simple reporting errors. Such cognitive information obtained from reinterviews has been valuable in survey instrument development, training, and the interpretation of survey results" (Hanuschak et al., 1991).

The purpose of NCES reinterview programs is:
1) to determine how good questions are with respect to a measurement of the response error and 2) to assess the quality of the data collected. The extent of the research effort varies across surveys from small sample reinterview programs conducted as part of a survey questionnaire field test to larger samples that range between 1 and 11 percent of a full-scale study. These programs have been used for three major purposes:

- *Identifying specific questions that may be problematic for respondents and result in high variability*
- *Quantifying the magnitude of the measurement error*
- *Providing feedback on the design of questionnaire items for future surveys*

Specifically, the purpose of the reinterview is to gain insight into the adequacy of questions. This gain can be achieved analytically by measuring two components of survey response -- response variance and response bias. These two measures are explained in more detail in section 5.

Another common purpose of reinterview programs is to verify that the original interviews were genuine. NCES often uses a combination of mail and Computer Assisted Telephone Interview (CATI) for its surveys. Most of the NCES reinterview programs were done using Computer Assisted Telephone Interview (CATI) in a centralized setting. Since the CATI interviews were closely monitored, it is highly unlikely that a telephone interviewer could invent or falsify interviews. Therefore, this aspect of reinterview is not typically a focus of NCES reinterview programs.

3. Surveys and Reinterview Design

Several NCES surveys have conducted reinterview programs for more than one round or cycle of the survey, specifically Baccalaureate and Beyond (B&B), Beginning Postsecondary Students (BPS), National Household Education Survey (NHES), Schools and Staffing Survey (SASS), and Teacher Follow-up Survey (TFS). Most of the programs do not include the same items on subsequent rounds of the reinterview, however. The BPS reinterview programs, for example, are designed "to build on previous analyses by targeting revised or new items not previously evaluated" (Burkheimer et al. 1992). Most of the NCES reinterview programs were developed to estimate response variance, but some, such as the Adult Education component of NHES:95, included a response bias study as well. However, the surveys that involved

testing, such as the National Adult Literacy Survey (NALS), the National Assessment of Educational Progress (NAEP), and the National Educational Longitudinal Study (NELS) never considered retesting or reinterviewing because NCES considered such activities too much of a burden on the respondents.

The following issues are considered in NCES reinterview designs:

- *Time Lag*

Time lag between the original and the reinterview for most of NCES surveys is usually stated as a range of days or weeks following the original survey such as "the reinterviews were conducted in October and November, about 4 to 6 weeks after the original interview" (Brick, Cahalan et al., 1994, p. 3-3). Early Childhood Education (ECE) reinterviews for NHES were designed for 14 days after the completion of the original ECE interview, but they were actually completed between 14 and 20 days after the original interview. BPS reinterviews were conducted up to 8 weeks after completing the original interview, and reinterviews in NPSAS were conducted between one and three months after the original interview.

- *Reinterview Instrument*

The reinterview instrument is a subset of the original questionnaire, but the question wording is almost always identical between the original and the reinterview instrument. In some cases, however, adjustments were made to the question wording in an attempt to gain more reliable data such as National Survey of Postsecondary Faculty (NSOPF). This is most often the case if the reinterview is conducted as part of the field test and not as part of the full-scale study. Other case examples are National Postsecondary Student Aid Study (NPSAS) and B&B.

- *Mode*

The mode of reinterviews is usually telephone regardless of the original interview mode. Conducting all the reinterviews by telephone violates survey error model assumptions that require the reinterview to be an independent replication of the original interview in order to estimate response variance accurately. Therefore, SASS included research in its 1991 reinterview program to determine the impact mode change might have on data quality. Most of the mail respondents were reinterviewed by mail and the telephone follow-up cases were reinterviewed by telephone. Generally reinterviews conducted by mail showed relatively lower response variance than the telephone reinterviews (Royce, 1994).

4. Sample Size and Response Rates

The design of NCES reinterview programs typically includes a target number or percentage of completed reinterviews. The reinterview sample size for RCG:91, for example, was 583 with a goal of 500 completed reinterviews. SASS reinterviews 10 percent of the School and Administrator samples and one percent of the Teacher sample to have a reinterview sample of just over 1,000 for each of its components. The reinterview sample sizes are considerably smaller when the reinterview program is conducted as part of the field test. The reinterview sample size for the NPSAS 1992-93 field test, for example, was 237.

NCES reinterview response rates vary from a low of 51 percent to a high of 94 percent with most reinterview programs having response rates in the mid-eighties to low nineties.

5. Measurement Error Estimation

Reinterview programs at NCES tend to measure the response variance and response bias using simple measures of consistency. Response variance is a component of measurement error which examines how consistently respondents answer questions in a survey. Response bias, on the other hand, measures the systematic nonsampling errors. In order to estimate response variance and response bias, it is necessary to define a general measurement error model:

Let $Y_{t1} = 1$ if Yes is recorded for unit t in the original interview and $Y_{t1} = 0$ otherwise.

Let $Y_{t2} = 1$ if Yes is recorded for unit t in the reinterview and $Y_{t2} = 0$ otherwise.

For unit t (t = 1,2,...,n) and the ith measurement (i = 1,2), the assumed model is:

$$Y_{ti} = X_t + B_{ti} + e_{ti}$$

Where Y_{ti} is a Bernoulli random variable, X_t is the "true" value of unit t, assumed unchanged between measurements, B_{ti} is the response bias, and e_{ti} is a random measurement error.

To illustrate the model consider the crossclassification of two measurements of an individual population characteristic (for example, whether a person who participates is a college graduate) obtained from an original interview and a reinterview of the same sample of individuals. Table 1 shows the crossclassification (Brick, Cahalan et al., 1994):

Table1: Two measurement crossclassification

Reinterview	Original Interview		
Response	Yes	No	Total
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

In their simplest form, reinterview results are analyzed using measurements derived from this crossclassification. These measures include the overall deviation between the interview and the reinterview, deviations on individual responses, and the index of crude agreement. The three specific measures most commonly used by NCES are:

- Gross Difference Rate (GDR)
   -Measures the weighted percentage of cases reported differently in the original and the reinterview as
   $$(b + c) / 2n$$

- Index of Inconsistency (IOI)
   -Estimates the proportion of total survey variance due to simple response variance as
   $$n(b + c) / 2(a + c)(b + d)$$
   -Assumes simple random sampling with replacement

- Net Difference Rate (NDR)
   -Computed after reconciliation for each answer category of a question
   -Weighted difference of the false positive and false negative rates calculated as
   $$(c - b) / n$$

Typically, the first two measures, GDR and IOI, are used to estimate simple response variance. The third measure, NDR, is used to estimate response bias, which most measurement models at NCES assume to be constant in the repeated measurement. NDR is also used to test the independence of the reinterview.

6. NCES Reinterview Results

This section will present a summary of selected results from reinterview programs conducted by NCES. Only results from reinterview programs conducted as part of full-scale studies are provided here. These studies include NHES, RCG, and SASS.

National Household Education Survey (NHES)

The National Household Education Survey (NHES) is designed to collect education data from U.S. households through telephone interviews, using random digit dialing (RDD) and computer-assisted telephone interviewing (CATI) procedures. The sample is drawn from the noninstitutionalized civilian population in households having a telephone in the 50 states and the District of Columbia.

NCES has conducted four comprehensive reinterview programs for the full-scale NHES surveys. The reinterview program for NHES:91 was administered only on the early childhood (ECE) component. In NHES:93 both components underwent reinterviews, while only the adult education component was reinterviewed for NHES:95.

All NHES reinterview programs have used gross difference rate (GDR) and net difference rate (NDR), and all except NHES:95 used index of inconsistency (IOI), as measures of response variability and response bias for critical items in the surveys.

The NHES:91 reinterview results suggested that the ECE interview measured some variables with relative success, but it also revealed some items needed to be handled carefully when tabulating findings and for which alternative methods of collection should be considered (Brick et al., 1991).

The early childhood component of NHES:91 reinterview program included questions on current enrollment (whether the child was attending school and, if so, what grade) and home environment (reading and television habits). All of the seven enrollment items had low GDRS and IOIs. Of the four home environment variables there were two worth noting:

- P19/E36: How often do you or other family members read stories to (child)?
- P22/E40: How many hours each day does (child) watch television or videotapes?

Brick et al. (1991) suggested the relatively large IOI (42.0) for the television question might be due to the "general ambiguity in the item, the crude measurement scale (whole hours) relative to the internal variability in the item, and differing circumstances" (p. D-20). The reading question also had a relatively large GDR (23.3) and IOI (33.5).

The two topical components of NHES:93 were: 1) the School Readiness (SR) interview of parents of children (ages 3-7 and 8-10) enrolled in second grade or below and 2) the School Safety and Discipline (SS&D) interview of parents of students enrolled in grades 3-12 and youths enrolled in grades 6 through 12. The subset of the original SR and SS&D questionnaire items chosen were selected because they were substantively important, not highly time dependent, and not examined in the NHES:91 reinterview. The reinterview sample sizes were substantially increased from the 604 of the NHES:91 reinterview program to 977 for SR and 1,131 for SS&D in order to obtain more reliable estimates of the response variance for key questions. The reinterview did not reveal any items with response problems that were severe enough to cause researchers to question analysis based on the item.

The NHES:95 reinterview program examined and estimated measurement errors as components of nonsampling error in the Adult Education (AE) survey. A subset of items from the original interview were selected and the original and reinterview responses were then compared to estimate the consistency of reporting. Interviews were sampled at different rates for participants and nonparticipants (i.e., people who did not participate in adult education activities), with a total of 1,289 cases selected for reinterview.

The GDRs for the NHES:95 reinterview programs were low for the adult education participation and the education background items, indicating that responses to those questions were consistent. The GDRs for barrier to participation items (such as obstacles that prevented respondents from adult education activities) were much higher than for the other subject areas, indicating that responses were not consistent. Only four (out of 15) barrier items had GDRs of less than 10 percent, and the highest GDR approached 50 percent. This inconsistency may have been related to factors like recoding the questions, additional eligibility criteria, and small sample sizes. Nonetheless, barrier items had some response problems and did not appear to be reliable.

NCES also conducted a separate study to measure bias for NHES:95. The methodology used for this survey appeared to have potential for detecting biases; however, this method -- intensive interviews -- was not as successful as the standard NHES reinterviews for estimating consistency of reporting.

Recent College Graduates (RCG)

The 1991 Survey of Recent College Graduates (RCG:91) provides data on the occupational and educational outcomes of bachelor's degree and master's degree recipients one year after graduation. Telephone interviews were conducted between July 1991 and December 1991 using computer-assisted telephone interviewing (CATI).

Two measurement error models were estimated from the reinterview data. The first model (the simple response variance model) assumed the errors were all from random sources. This model was then expanded to allow for systematic errors or biases. Both models assumed that the interviewers were not a source of systematic error in the data collection process, but the first assumed that the measurement errors were the same across sampled graduates and from one trial to the next. Thus, if the reinterview was uncorrelated with the original interview, then the number of original and reinterview errors should be roughly equal.

Of the 16 reinterview items in these categories, only two had an IOI greater than fifty percent. One item was related to employment experience, while the other was a question dealing specifically with teacher certification and employment:

> Q24: Were you looking for work during the week of April 22, 1991?
> Q62: Prior to completing the requirements for your 1989-90 degree, were you at any time employed as a school teacher at any grade level, from prekindergarten through grade 12? Please exclude student or practice teaching and work as a teacher's aide.

Question 24 was asked only of a subset of the sample of graduates -- those who were unemployed. The reduced sample size may have contributed to a larger GDR and IOI. There were also potential recall problems since the question referred to a specific

period of time. For question 62, no explanation was offered for the possibly high random measurement error.

The overall conclusions of Brick, Cahalan et al. were that even though measurement errors were an important source of error in RCG:91, the estimates from the survey were not greatly distorted by these errors. The relatively small GDRs indicated responses were consistent; however, the IOIs being generally moderate implied that improvements in questionnaire wording and construction might help to reduce measurement errors in future surveys.

Schools and Staffing Survey (SASS)

The Schools and Staffing Survey (SASS) is a periodic, integrated system of surveys designed to collect data on characteristics of public and private school teachers, administrators, and their workplaces. The first two rounds of SASS (1987-88, and 1990-91) included the School Survey, the School Administrator Survey, the Teacher Demand and Shortage Survey (TDS), the Teacher Survey, and, one year later each time, the Teacher Follow-up Survey. SASS includes reinterview programs as part of its survey design, although it has used other methodologies for measuring response variance and response bias as well, including validation studies, such as the Teacher Listing Validation Study (TLVS), and follow-up cognitive research.

The following part of this paper summarizes overall results and comparison of the 1987-88 and 1990-91 reinterviews. There was no difference in response variance between public and private administrators, schools, or teachers.

Thirty-nine percent of the 1990-91 SASS reinterview questions showed low response variance. This was significantly better than the 11 percent of reinterview questions for SASS 1987-88 with low response variance (see table 2). Moreover, there was a 23 percentage point difference between 1990-91 and 1987-88 SASS items with a high response variance (26 percent versus 49 percent) (Royce, 1994).

It is important to note that the results across 1987-88 and 1990-91 are not strictly comparable. Different sets of questions were used for the two interviews. Among the 15 factual questions common to both years, 11 showed significant revisions in 1991. Four of these items displayed reduced response variance, which indicates improvement in these questions (Bushery, Royce and Kasprzyk, 1992, p. 459).

Table 2. Summary of 1987-88 and 1990-91 SASS reinterview response variance results*

	Low		Moderate		High	
	Number	Percent	Number	Percent	Number	Percent
All three components:						
1988	4	11%	14	40%	17	49%
1991	43	39%	38	35%	28	26%
Administrator Survey:						
1988	1	11%	4	44%	4	44%
1991	5	20%	10	40%	10	40%
Teacher Survey:						
1988	3	25%	4	33%	5	42%
1991	21	44%	16	33%	11	23%
School Survey:						
1988	0	0%	6	43%	8	57%
1991	17	47%	12	33%	7	19%

*Questions for which index could be reliably estimated.

SOURCE: Royce, D. (1994), *1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report*, (working paper 94-03), U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

7. Conclusion

This study is part of the adjudicated report on "NCES MEASUREMENT ERROR PROGRAMS" which is scheduled for publication in the spring of 1997. This report synthesizes results from a sample of NCES reinterview programs, validity evaluations, and cognitive research studies.

Results indicate that the measurement error programs have helped NCES to improve the quality of its data. Over different rounds of surveys, the reinterview sample sizes have increased and response variance in most of the surveys in different areas have been improved. In some surveys it was found that inconsistencies between responses were attributed to factors like recoding the questions, lack of knowledge about the questions, eligibility criteria and small sample sizes. More recently, NCES applied alternative methods separately from reinterview programs to measure response bias. These methods were effective but costly. It was suggested that these methods be used when there is an indication of reporting errors.

Although, overall studies indicated that questions had low to moderate GDRs and low to moderate IOIs in several different NCES surveys, improvements in questionnaire wording and construction might help to reduce measurement errors in future surveys.

References

Brick, J., Cahalan, M., Gray, L., Severynse, J. and Stowe, P. 1994. *A Study of Selected Nonsampling Errors in the 1991 Survey of Recent College Graduates.* (NCES 95-640). Washington, DC: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Brick, J., Collins, M., Celebuski, C. Nolin, M., Squadere, T., Ha, P., Wernimont, J., West, J., Chandler, K., Hausken, E. and Owings, J. 1991. *NHES:91 1991 National Household Education Survey Methodology Report.* (NCES 92-019). Washington, DC: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Brick, J., Collins, M., Nolin, M., Ha, P., Levinsohn, M. and Chandler, K. 1994a. *National Household Education Survey of 1993: School Safety and Discipline Data File User's Manual.* (NCES 94-218).

Washington, D.C.: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Brick, J., Collins, M., Nolin, M., Ha, P., Levinsohn, M. and Chandler, K. 1994b. *National Household Education Survey of 1993: School Readiness Data File User's Manual.* (NCES 94-193). Washington, D.C.: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Burkheimer, G., Forsyth, B., Wheeless, S., Mowbray, K., Boehnlein, L., Knight, S., Veith, K. and Knepper, P. 1992. *Beginning Postsecondary Students Longitudinal Study Field Test Methodology Report. BPS:90/92.* (NCES 92-160). Washington, DC: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Bushery, J., Royce, D. and Kasprzyk, D. 1992. *The Schools and Staffing Survey: How Reinterview Measures Data Quality.* Proceedings of the Section on Survey Research Methods, American Statistical Association. Alexandria, VA: ASA, 458-463.

Forsman, G. and Schreiner, I. 1991. The design and analysis of reinterview: An overview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiewetz, and S. Sudman (eds.), *Measurement Error in Surveys.* 279-302. New York: John Wiley & Sons, Inc.

Hanuschak, G., Atkinson D., Iwig, W. and Tolemeo, V. 1991. *History of Reinterview Studies at NASS.* Proceedings of the section on Survey Research Methods. American Statistical Association.

Kasprzyk, D. 1992. *The Schools and Staffing Survey: Research Issues.* Proceedings of the section on Survey Research Methods. American Statistical Association.

Royce, D. 1994. *1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report.* (Working Paper 94-03). Washington, D.C.: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.