Using Reinterview and Reconciliation Methods to Design and Evaluate Survey Questions

Jeremy E. Morton
RTI International

Paul A. Mullin
IBM Business Consulting Services

Paul P. Biemer
RTI International

Conducting reinterviews is an effective method to estimate and reduce response errors in interview surveys. As part of the School Health Policies and Programs Study 2000 (SHPPS), RTI used reinterview methods to assist in designing and evaluating survey questions. Reinterviews were conducted in the field test with selected respondents to identify discrepancies between the original interviews and reinterviews. Reconciliation interviews were then conducted to determine the reasons for the discrepancies in terms of comprehension, recall, encoding, response options, or other problems. In this paper, we describe the design of the reinterview and reconciliation study and discuss the implications of using these methods for questionnaire design and evaluation, specifically in comparison to cognitive interviewing.
Keywords: reinterview, reconciliation, questionnaire design

Introduction

Threats to survey data quality can arise through various aspects of data collection including respondent selection rules, interview procedures, and questionnaire design. Numerous pretesting methodologies have been used in survey research with the goal of increasing the reliability of questionnaires and reducing measurement error. Some of these methodologies include expert review, focus groups, behavior coding, respondent and/or interviewer debriefing, and cognitive interviewing. Cognitive interviewing has become a common practice in survey research in assisting in the design and evaluation of survey questionnaires. Cognitive interviewing provides valuable information about how respondents view the question-answer process, access short-term and long-term memory, and process terminology.

Another less commonly used technique for pretesting questionnaires is the use of reinterview and reconciliation methods. A reinterview – replicated measurement on the same unit – is a new interview where all or a subset of the questions of the original interview are repeated (Forsman and Schreiner 1991). Reconciliation is the process where inconsistent answers obtained during the original interview and the reinterview are reconciled with the respondent. This may only consist of determining which response was the correct response or it may also include further discussion about the reasons why the inconsistency occurred. Reinterview and reconciliation methods can offer a more direct way to investigate sources of unreliablity in survey questionnaires than

other common pretesting techniques such as cognitive interviewing. In our current research, we used an application of this process to identify sources of error for questionnaire design and evaluation purposes.

The objective of this paper is to present some of the findings of our reinterview and reconciliation procedures conducted as part of the SHPPS 2000 field pretest for the purpose of documenting the effectiveness of these methodologies for questionnaire development. First, we highlight previous research that was conducted using similar reinterview methods and compare reinterview techniques to cognitive testing as pretesting methodologies. We then describe the design of the reinterview and reconciliation methods used in this study and highlight some of the differences between our methods and previous studies using similar methods. Next, we present some of the findings of the study including the sources of error identified by the reconciliation procedures and a few specific examples of survey questions and reconciliation findings from the study. Finally, we discuss the implications and utility of using reinterview methods for the evaluation and refinement of survey questions.

Previous Uses of Reinterview and Reconciliation Methods

Reinterview methods have been widely used to model and measure data quality in a variety of surveys (see Forsman and Shriner 1991 for overview of methods and purposes; and Biemer 2004 for an overview of statistical methodology). However, very few of studies have used reinterview and reconciliation methods in combination for the purpose of questionnaire design and evaluation. In these studies, the reconciliations of discrepant answers were conducted immediately after the reinterview was completed. The reinterviews and reconciliations were conducted by field or telephone in-

Contact information: Jeremy E. Morton, RTI International, 6110 Executive Boulevard, Suite 902, Rockville, Maryland 20852, USA, Email: jmorton@rti.org

terviewers or by interviewer supervisors. In most cases, the reconciliations were conducted using scripted questions provided to the reinterviewers.

The goal of Jenkins and Wetzel's (1994) 1991-92 Teacher Follow-up Survey (TSF) Reinterview and Extensive Reconciliation study was to identify problematic questions, identify the source of the problems, and to offer suggestions for improving the questions. They were able to identify a few problem questions and make recommendations for either improving the questions or for further research. However, the authors note that the reinterview and reconciliation study produced very few differences because the reinterview was not independent from the original interview (original responses were visible to the interviewers on the reinterview questionnaire). They recommended using an independent reinterview followed by an extensive unstructured reconciliation in a third visit.

The objectives of Brick and West's (1992) National Household Education Survey (NHES:91) reinterview program were to identify unreliable survey items, measure the response variance for groups of items, and to provide feedback for improving questions for future surveys. The authors state that all three of these objectives were accomplished using their reinterview methods. Most of the items included in the reinterview had small to moderate measurement errors. For the items with high measurement errors, they were able to identify potential problems associated with those items.

The Medicare Beneficiary Health Status Registry field test reinterview study (Forsyth et al. 1996) was designed to evaluate overall data quality and to collect information on response inconsistencies to evaluate questionnaire features interfering with data quality. The authors reported a generally high level of response consistency, but also were able to identify reasons for discrepancies. The most frequently occurring reasons for inconsistent responses were problems with question wording and response options.

Brick and Morganstein's (1996) NHES:95 study design differed from the reinterview research described above in that the interviewers were specially trained to conduct intensive reinterviews with reconciliation in a conversational mode using probes to trigger recall and comprehension. The authors concluded that the intensive reinterview methodology was effective for detecting biases, and that the ability of the traditional reinterview reconciliation approach to estimate response bias is limited.

The comparison of self and proxy reports may be considered another method of reinterview. However, the focus of most studies in this area (e.g., Mingay et al. 1994; Sudman et al. 1994) has primarily been to evaluate the quality of information from proxy reports, rather than to directly address question development. Nevertheless, it should be clear that some proportion of the incongruence in self and proxy response is attributable to the general unreliability of responses. Therefore, studies in which diverging self and proxy responses are reconciled as a means for identifying the problematic questions (e.g., Edwards et al. 1989) bear considerable similarity to the reinterview reconciliation methods considered in the present study.

Comparison to Cognitive Interview Pretesting

In reinterview reconciliation interviews, question-answering difficulties are identified with respect to variations over time for the same respondents (within subjects). In standard cognitive interviewing, by comparison, question-answering difficulties are identified with respect to deviations from the survey author's intentions, and variations between subjects in performance feed the interpretative process. This difference – the source of question answering variation – is of particular interest in the comparison of the types of errors and question problems that are identified by the two methodologies.

It is widely recognized that different respondents vary in their understanding of questions and in the answering strategies they employ, but it is perhaps less frequently addressed in the survey methods research literature that respondents may understand questions differently and make diverging judgments to questions on separate occasions over time. Despite the common use of reliability measures to evaluate data quality, few studies have attempted to determine the source of this type of error (e.g., comprehension or recall problems) and use this information to direct efforts in questionnaire development and refinement. Yet, the large body of research demonstrating that comprehension and judgment are influenced by what information is most accessible in memory at a given time (Higgins 1996; Schwarz 1995) strongly implies that within respondent variation may hold important clues to weaknesses in question design. To the extent that these variations are related, for example, to ambiguity of concepts, uneven reading of complex questions, or the effort of recall and the use of less than optimal recall strategies, we may discern the same opportunities to improve question design as is possible through cognitive interview pretesting.

One of our goals in the present research was to assess whether the types of errors and their distribution may be comparative to the errors in cognitive interviewing. Previous research has documented that cognitive interviewing identifies primarily comprehension and communication problems with questions in contrasts to other difficulties, such as recall errors and logical problems (Willis et al. 1999). One may question whether this is attributable to the nature of the method or the nature of survey error. It is conceivable that the reinterview may be more sensitive to recall difficulties than standard cognitive interviewing owing to the highly variability in individual ability to remember events and facts. At the same time, there is reason to assume relative consistency in recall from interview to reinterview based in that recall on one occasion is associated with enhanced recall in later attempts (Darley and Murdock 1971; Hogan and Kintsch 1971).

Reinterview and reconciliation procedures also differ from cognitive interviewing in the context in which the interviews are typically conducted. Cognitive interviews are usually conducted in laboratory setting by research staff and the survey is administered in the same sitting as the cognitive interview. In most reinterview studies, including our own, the survey is administered in a field test setting with

trained CAPI interviewers. Thus, reinterview reconciliation procedures have the potential to detect errors resulting from field interviewer administration whereas standard cognitive interviewing methods do not.

Finally, we should point out that our reinterview reconciliation procedures were used to evaluate instruments that had been submitted already to a round of cognitive interview testing. Our reconciliation method was applied to instruments that had undergone significant refinement. This made direct comparisons of the two methods in the context of the SHPPS study somewhat difficult.

Methodology

Background of SHPPS 2000

The School Health Policies and Programs Study 2000 (SHPPS 2000) was a national survey of school health programs at the state, district, school, and classroom levels. The study was designed to address five factors: 1) the prevalence of health risk behaviors among youths, 2) national health objectives and national education goals, 3) the findings and recommendations of the Institute of Medicine report, *Schools and Health: Our Nation's Investment*, 4) the need to monitor the impact of federal funding for Human Immunodeficiency Virus (HIV) prevention efforts in schools, and 5) the need to monitor the impact of federal health education efforts in schools.

Efforts to identify and eliminate threats to data quality in SHPPS 2000 was especially challenging due to the complex mixture of respondent-types (i.e., from different levels within the educational infrastructure and from numerous professional roles within each level) and components of comprehensive school health. Therefore, a Validity and Reliability Pretesting Team (V/R Team) was assembled to evaluate the SHPPS 2000 data collection procedures, and provide recommendations for controlling threats to data quality. Pretesting of the SHPPS 2000 questionnaires and data collection procedures included:

- cognitive interviewing with each questionnaire,
- a field pretest including reinterview data collection procedures at the school and classroom levels,
- reconciliation of discrepancies between the original interview and reinterviews.

Reinterview Study Design

Between April and June, 1999, a field test of data collection procedures was conducted at schools similar to those that were eventually selected for the main study. In addition to a general test of procedures, reinterviews and reconciliation interviews were conducted to examine the reliability of data from four of the study questionnaires that were administered at the school or classroom level: Physical Education – School Level, Health Services – School Level, Mental Health and Social Services – School Level, and Health Education – Classroom Level. The initial school and classroom interviews and the reinterviews were conducted using

computer-assisted personal interviewing (CAPI) with samples of school and classroom respondents.

The field pretest was conducted in three geographical regions that were selected based on the variety of school types (e.g., large public, Catholic, etc.) they contained and their proximity to RTI facilities. The three regions included the Chicago, Illinois metropolitan area, Central North Carolina, including the Raleigh, Durham, Chapel Hill, and Greensboro areas, and Charles County, Maryland. From within these areas, nine public school districts, three Catholic dioceses, and a few private (non-Catholic) schools were selected for the study. A random sample of schools (including an oversampling of elementary and non-public schools) was then drawn from these school districts which yielded 52 participating schools for the field test. Out of the 52 participating schools, a total of 269 eligible respondents were identified and interviews were completed with 256 (95 percent) of them.

All respondents were asked to complete a repeat interview approximately two weeks following the initial interview. The field interviewers completed 242 reinterviews, which represents 90 percent of the eligible respondents and 95 percent of the initial interview respondents. The reinterviews were conducting using the same CAPI system and field interviewers used in the initial interviews. In the vast majority of cases, attempts to have the same interviewer conduct the interview and reinterview with a given respondent were successful. Neither the respondents nor interviewers were explicitly made aware of the answers given by the respondents to the initial interviews.

Reconciliation Interviews

Before field pretest data collection began, a subset of items from the four questionnaires was selected for the follow-up reconciliation interviews. The CDC project officer identified specific items from each of the four questionnaires that merited special evaluation. Some of these items were selected because they focused on critical aspects of the component (e.g., the number of hours of physical education students are required to receive); others were selected because of concerns about potential data quality issues. Since the immediately preceding questions are often most influential context effects (e.g., Schwarz and Bless 1992), we made certain to include all lead-in questions and other items and questions immediately preceding the target items.

During the field pretest data collection period, the V/R Team identified respondents who had provided incongruous answers to the pre-selected questions during the initial interview and the reinterview. Respondents with a greater number of discrepancies were targeted for reconciliation interviews. Within 1 to 2 weeks after the reinterview, the V/R Team telephoned these respondents to reconcile their answers. Reconciliation interviews were conducted with 23 school physical education respondents, 18 classroom health education respondents, 18 mental health and social services respondents, and 25 health services respondents.

Survey methodologists from the V/R Team conducted

the reconciliation interviews. The reconciliation interview protocol included an introductory script explaining why these additional interviews were being conducted (emphasizing that the questions were being evaluated, not the respondents' answers), the survey questions of interest, and the answers to the original interview and reinterview. For the most part, the reconciliation was an unstructured discussion using unscripted and spontaneous probing. We presented the discrepancies to respondents and asked open-ended questions concerning the reasons why they may have provided diverging responses. The discrepancies provided a concrete target of discussion for participants and we considered the need to script additional directed probe questions prior to the interviews to be minimal.

The SHPPS 2000 pretest methods differed from prior reinterview reconciliation studies in several ways. First, the reconciliations were separate from the reinterviews and conducted as a follow-up interview using a different mode (telephone) than the survey interviews/reinterviews. Second, field interviewers conducted the original interviews and reinterviews while the follow-up reconciliation interviews were conducted by survey methodologists from the V/R Team. Third, the reconciliation interviews were conducted using a protocol but for the most part, the V/R Team conducted the interviews using unscripted and spontaneous probing. While the delay in conducting the reconciliation interview posed a potential cost to respondents recall concerning the source of the discrepancies, the use of methodologists experienced in cognitive interviewing techniques offered potentially richer information on question answering difficulties.

Results of Reconciliation Interviews

Sources of Response Error

The reconciliation interviews identified several sources of response error including:

- comprehension problems (e.g., ambiguous terminology such as "regular participation," understanding how health and physical education can be completely separated from one another);
- problems identifying the reference set (e.g., which students should be considered while answering the questions);
- recall problems (e.g., how many hours of training they had received during a two-year period);
- interviewer keying error (e.g., entering numerical responses); and
- real change occurred between the initial interview and the reinterview

We classified the types of the errors identified in the reconciliation interviews according to a scheme that was based on a typology of question difficulties developed by Conrad and Blair (1996) to formalize cognitive interviewing results. Table 1 displays the results of this classification, and the approximate percentage of discrepancies attributable to different sources. We note that for the most part, the reason offered

by respondents for the discrepancy mapped readily into this schema.

As noted earlier the remaining questions were questions about the existence of school policies, facilities, types of staff, and general practices at the school. Our earlier cognitive interviewing, and our experience in the discrepancy interviews suggested that strategies were mixed for answering policy and practices questions. In some cases, respondents responded based on the recall of factual knowledge of policies or practices that seemingly fit the descriptive category; in other cases, respondents responded based on the recall of events that apparently fit the descriptive category. In both cases, respondents frequently deliberated in their judgments as to whether or not the policy or practice was in place. That is, these judgments were not always automatic.

With this general description of the question answering processes in mind, it is perhaps not unexpected that the preponderance of discrepancies were inclusion/exclusion variations in comprehending the question. Likewise, the majority of the recall discrepancies identified were "no" responses in the original interview appearing as "yes" response in the reinterview when respondents recalled additional classes of events or factual knowledge in the reinterview that might fit a policy/practice category. These discrepancies were highly similar to the inclusion/exclusion comprehension discrepancies but were ascribed by respondents as occurring in recall rather than as a variation in what the respondent understood the question to mean.

It is noteworthy that the phenomena of variations in question comprehension, both inclusion/exclusion discrepancies and concept meaning variations, occurred as a "within-subjects" effect. Taken in conjunction with reinterview consistency rates, which indicate the absolute prevalence of the errors, these findings provide evidence of the dynamic and variable manner in which individual respondents bring to mind and weigh information in understanding a question.

One concern expressed by Willis et al. (1999), is that cognitive interviewing may be more sensitive to problems of comprehension than other problems such as recall, and some researchers have asserted that cognitive interviewing is primarily a means for evaluating the communication of question meaning (Gerber and Wellens 1997). In our investigation, the percentage of comprehension and recall errors identified by the method mirrored the percentages documented by Willis et al. (1999) for cognitive interviewing. These findings may support the idea that the preponderance of comprehension problems identified in cognitive interviewing are attributable to the underlying nature of survey measurement. While it is true that our reconciliations employed some cognitive probing techniques, the errors themselves were identified directly in the unreliability of the responses.

About 3% of the errors were attributable to respondents bringing forth conceptions from prior questions that interfered with and led to misconceptions about the meaning of the current question. While these types of errors may be sometimes identified in regular cognitive interviewing (e.g., Conrad and Blair 1996), it seems plausible that the technique

Table 1: Attribution of Error Source from Reconciliation Interviews

Class	Subtype	Number of Errors	Percentage of Errors
Comprehension	Concept Meaning	97	32%
	Reference Period	4	1%
	Inclusion/Exclusion	124	41%
	Syntactical Overload	1	< 1%
Performance	Lack of Knowledge	5	2%
	Recall	25	8%
	Context Carry-over	6	2%
	Calculation Overload	2	< 1%
Logical	Presupposition	5	2%
	Logical Devices	0	0%
	Contradiction	0	0%
Other	Administration	9	3%
	Real Change	13	4%
	Unknown	15	5%
Total		306	100%

of retrospective probing after each question may sometimes disrupt this phenomena, making more difficult to detect such errors. We believe this possibility deserves further study.

The reinterview reconciliation method identified some variations and errors associated with interviewer administration. In most cases these errors were data entry errors associated with marking numerical responses. However, two discrepancies were ascribed by respondents to the use of show cards. In one case, the respondent indicated that the interview employed the show card in one interview, but did not in the other and instead requested an open-ended response. In the other case, the respondent indicated having difficulty reading the show card while listening to the interview read the options at the same time. This latter problem is presumably one that may be identified in standard cognitive interviewing, while the prevalence of interview administration errors is clearly not in the scope of such pretesting.

The occurrence of real change is worth considering further as well. Approximately 5% of inconsistencies reconciled were due to real change. These occurred primarily in the measurement of practices, but a few changes occurred in policy as well. In this particular study, interviews were conducted during the school year rather than at the end of the year because of practical concerns about reaching school staff and teachers during the summer. The reconciliation data provided some suggestion that this may introduce some measurement error if we assume the objective of the research is to assess the practices that occurred at any time during the school year.

Finally, we anecdotally report some occasions in which respondents remarked that low levels of attention or motivation played a role in the discrepancies that occurred. One respondent suggested that all her discrepancies were due to a lack of motivation. These types of errors are not ever likely to be detected in cognitive interviewing using recruited participant samples, but, of course, there is little to recommend for remedying them.

Examples of Using Reconciliation Interviews to Revise Survey Questions

The reinterview and reconciliation study conducted in the field pretest provided important information on why discrepancies were being found between the original and reinterview study. This information was used to either make recommendations for revisions for the main study or suggest further research for many of the survey questions. In this section, we provide the findings for a few specific survey questions from our study.

During the past 2 years, about how many hours of staff development have you received on… (FILL LIST OF VARIOUS TOPICS)?

1. 0 hours
2. less than 4 hours
3. 4 or more hours

This question was asked to most of the respondents in the study. Originally, the response categories were not read to the respondents and only used by the interviewer to code the responses. The results from cognitive testing suggested that respondents would have a difficult time in recalling and estimating the number of hours they received in staff development for the various topics over the last two years. Thus, for the field pretest, the response options were read to the respondents to assist in recall and in answering the question.

The congruence rates (congruence was defined as exact agreement between the original and reinterview responses) obtained for this question from the reinterview study ranged from .60 to .78 depending on the type of questionnaire. The reconciliation interviews suggested that two issues were contributing to the discrepancies: 1) respondents were having difficulty recalling the number of hours in the two year reference period and their answers were inconsistent, even with the addition of the response categories, 2) respondents were having difficulty determining the number of staff development hours for specific topics that were taught in training sessions that covered numerous topics. For example, a respondent might have received staff development in a substance use prevention class which covered multiple topics such as tobacco use, alcohol use, and other drug use prevention and it might be difficult to discern how much time was spent on each topic. Based on these findings, the question was changed to a simple yes/no item for the main study: "During the past 2 years, did you receive any staff development on...?" The goal was to reduce the amount of burden placed on the respondents in having to recall information about staff development in the past two years. This question provides an example where information obtained from the reinterview study was used to modify questions that were previously revised based on cognitive testing results.

Is parent or guardian consent dependent upon...

1. the type of (health service/mental health or social service) or problem,
2. the student's age,
3. both the type of (health service/mental health or social service) or problem and the student's age, or
4. something else?

This question was included in the Health Services questionnaire and in the Mental Health and Social Services questionnaire. This item was the last in a series of three items, which asked about consent issues. The three items were created from a single item that was decomposed after cognitive testing. The results from cognitive testing suggested that the single item was very difficult for the respondents to understand and therefore was split into three separate items.

The congruence rate obtained for this question from the reinterview study was .50 for Health Services and .70 for the Mental Health and Social Services. During the reconciliation interviews, the respondents indicated that they had no idea what this question was asking. They picked response option 3 in the original interview in order to be inclusive, but then selected something else in the reinterview. The question was very unclear and might not have been a logical extension of the preceding two questions. The recommendations for improvements included revising the question because the concept was not being understood properly, adding a lead-in statement to help clarify the question, and revising the previous two items on consent. Because a major overhaul was suggested for these three items, these items were dropped for

the main study in both questionnaires. This question provides an example where revisions based on cognitive testing were evaluated in "real" interview settings. The concepts that were trying to be assessed proved to be too difficult to measure in this particular survey.

Overall, how many hours of physical education are (FILL GRADES) from your school required to take?
_____ ***HOURS***

This question was included in the Physical Education School questionnaire. The congruence rate obtained for this question from the reinterview study was .35. The findings from the reconciliation interviews showed that while a couple of sources of discrepancy were based on comprehension and recall problems, the majority of discrepancies were because of interviewer error. Numerous respondents commented that the interviewer must have keyed in the answer incorrectly as the respondents were confident they had provided the correct answer for both interviews. Perhaps the errors were a result of how the CAPI screen was set-up or something to do with the computer functions for this question. The reasons for these interviewer errors were unknown but this issue could not be ignored because of the frequency of the respondents' comments. The general recommendation was to provide additional interviewer training on this question. This item provides an example where interviewer errors were the cause of the majority of discrepancies in the reinterview study. These types of problems would not be found during cognitive interviewing and probably would not be noticed in a regular field pretest.

Discussion

A substantial number of questions were included in the reinterviews to preserve the continuity of the interview, but were not addressed in the reconciliations. We attempted a heuristic evaluation and recommendations for revision for such items having relatively low reliability. Our experience was that these evaluations were highly speculative, especially without obvious types of question problems inherent in an instrument that has not previously undergone cognitive testing and refinement. The reconciliation interviews offered a clear advantage in determining the types of problems that will occur for questions found to have unacceptable levels of reliability.

The reinterview and reconciliation methods used in this study were effective in assessing overall data quality, identifying problematic questions, and identifying the reasons why the questions were problematic for respondents. The methods were a valuable resource in further evaluating the findings and recommendations based on the cognitive testing results. Our results and our experience with the reinterview reconciliation method suggest that it detects a similar distribution of problems in question design to the errors yielded by standard cognitive interviewing. However, our reinterview methods were successful in discovering problems not found

in cognitive testing (e.g., potential interviewer errors). The reinterview and reconciliation study offers a questionnaire design method that tests survey questions in a real life interview setting. While cognitive laboratory interviewing methods are known to provide important information for questionnaire designers, the artificial context and setting of the cognitive interview may have some drawbacks. For example, during our reconciliation interviews a respondent motivation issue was uncovered as a source of inconsistent answers, especially for items that involved difficult recall tasks. Some of the reconciliation respondents stated they did not try very hard to remember the exact answers and therefore provided very rough estimates which is what led to the different answers in the original and reinterview. During a cognitive interview, one can expect that participants would be motivated to concentrate during the interview and to provide as accurate answers as possible because 1) distractions from real interview settings are eliminated, 2) participants *want* to participate as they most likely responded to an advertisement, and 3) participants are usually provided an incentive to participate.

In addition, we identify two other drawbacks to cognitive interviewing: 1) potential "problem" areas that are found during cognitive testing may not necessarily lead to inconsistent responses, and 2) because of the typical small sample sizes (9-10 participants) used in cognitive testing, problems that are identified during the testing may not occur in a high frequency in the actual survey. We believe reinterview and reconciliation methods offer other advantages over cognitive interviewing because of these drawbacks. Reinterview and reconciliation methods identify only the problems that lead to inconsistent answers and can target those items with the highest frequency of inconsistencies.

The utility of the reinterview and reconciliation methods may be limited when conducted as part of a field test with limited time and resources. It is important that ample time and budget be allotted between the pretest and main study in order to implement the reinterview procedures and incorporate the findings. Another conceivable instantiation of the reinterview and reconciliation method would involve conducting the procedures at an earlier stage of questionnaire development without using professional interviewers and a computerized instrument. For example, methodology research staff might administer a paper version of the survey questions, and then several weeks later conduct a reinterview and then reconcile discrepancies afterwards. There are several obvious practical concerns with this version of the method in comparison to standard cognitive interviewing: a) the additional burden and cost of two laboratory sessions, and b) the need to have large numbers of participants to produce a sufficient yield of inconsistencies in response. The latter is clearly a potential obstacle to the use of the technique, although both concerns might be mitigated to a degree by conducting the procedures on the telephone.

Conducted as part of a field test, reinterview and reconciliation methods are not likely to replace other evaluation techniques (such as cognitive interviewing) in questionnaire development, but it can be effective for final questionnaire refinement. It may also serve as an evaluation of the findings and revisions from earlier cognitive testing. In addition, the results are potentially useful to the analysis and interpretation of the data for reporting. The need to make whatever refinements to a questionnaire are feasibly completed between a field test and main study need to be weighed with the significant costs of the procedures in time and money, and the extra burden on respondents and field staff. In contrast, we believe that the rich information provided by reinterview and reconciliation methods applied in a field pretest or main study may substantially benefit future administrations of continuous or repeated survey studies.

References

Biemer, P. (2004). Modeling Measurement Error to Identify Flawed Questions. In S. Presser et al. (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (p. 225-246). New York: John Wiley & Sons, Inc.

Brick, J. M., & Morganstein, D. (1996). *Estimating Response Bias in an Adult Education Survey.* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 728-733)

Brick, J. M., & West, J. (1992). *Reinterview Program for the 1991 National Household Education Survey.* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 387-392)

Conrad, F., & Blair, J. (1996). *From Impressions to Data: Increasing the Objectivity of Cognitive Interviews.* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 1-9)

Darley, C. F., & Murdock, B. B. (1971). Effects of Prior Free Recall Testing on Final Recall and Recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 701-711.

Edwards, W. S., Levine, R., & Cohany, S. (1989). *Procedures for Validating Reports of Hours Worked and for Classifying Discrepancies Between Questionnaire Reports and Validation Totals.* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 496-501)

Forsman, G., & Schreiner, I. (1991). The Design and Analysis of Reinterview: An Overview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (p. 279-301). New York: John Wiley & Sons, Inc.

Forsyth, B. H., Pate, D. K., Smith, T. K., & Fitterman, L. (1996). *Reinterview Methods for Assessing and Improving the Quality of Data from a Medicare Population.* (Health Survey Research Methods Conference Proceedings, P.H.S. Publication No. 1013, 79-82)

Gerber, E., & Wellens, T. (1997). Perspectives on Pretesting: "Cognition" in the Cognitive Interview? *Bulletin de Methodologie Sociologique*, *55*, 18-39.

Higgins, E. T. (1996). Knowledge Activation: Accessibility, Applicability, and Salience. In E. Higgins & A. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* (p. 133-168). New York: Guildford Press.

Hogan, R. M., & Kintsch, W. (1971). Differential Effects of Study and Test Trials in Long-Term Recognition and Recall. *Journal of Verbal Learning & Verbal Behavior*, *10*, 562-567.

Jenkins, C. R., & Wetzel, A. (1994). *The 1991-92 Teacher Follow-up Survey Reinterview and Extensive Reconciliation.* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 815-820)

Mingay, D. J., Shevell, S. K., Bradburn, N. M., & Ramirez, C. (1994). Self and Proxy Reports of Everyday Events. In N. Schwarz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (p. 235-250). New York: Springer-Verlag.

Schwarz, N. (1995). Social Cognition: Information Accessibility and Use in Social Judgment. In D. Osherson & E. Smith (Eds.), *An Invitation to Cognitive Science (vol. 3, 2ⁿᵈ ed.): Thinking* (p. 345-376). Cambridge, MA: MIT Press.

Schwarz, N., & Bless, H. (1992). Constructing Reality and its Alternatives: Assimilation and Contrast Effects in Social Judgment. In L. Martin & A. Tesser (Eds.), *The Construction of Social Judgments* (p. 217-245). Hillsdale, NJ: Erlbaum.

Sudman, S., Bickart, B., Blair, J., & Menon, B. (1994). The Effect of Participation Level on Reports of Behavior and Attitudes by Proxy Reporters. In N. Schwarz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (p. 235-250). New York: Springer-Verlag.

Willis, G., Schechter, S., & Whitaker, K. (1999). *A Comparison of Cognitive Interviewing, Expert Review, and Behavior Coding: What do they tell us?* (Proceedings of the Section on Survey Research Methods, American Statistical Association, 28-37)