

RESEARCH REPORT SERIES
(Survey Methodology #2007-39)

**Respondent Debriefings Conducted by Experts:
A New Qualitative Methodology
for Questionnaire Evaluation**

Elizabeth Nichols
Jennifer Hunter Childs

Statistical Research Division
U.S. Census Bureau
Washington, D.C.

Report Issued: October 3, 2007

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Respondent Debriefings Conducted by Experts: A New Qualitative Methodology for Questionnaire Evaluation

In theory, experts in a domain of interest could perform an in-depth interview using scripted and unscripted questions to obtain information needed to answer research questions. These experts would feel confident that the data they collected was of high quality since they could supplement the scripted probes with unscripted probes to determine the truth (see Conrad and Schober, 2000, and Schober and Conrad, 1997 for studies on conversational interviewing). If they felt the respondent was confused by questions, or if the respondent indicated there was more pertinent information available, then experts could inquire further. This is largely the method employed by experts doing cognitive testing of a questionnaire. However, this method is impractical for a large data collection.

This paper explores a method that uses expert respondent debriefings as a means of evaluating the quality of the production data. This is especially important for surveys for which an assessment of "truth" is not otherwise available. In July of 2006, a small group of experts in Census residence rules observed 169 Census Coverage Measurement Person Interviews in two sites for the 2006 Census Test. Immediately following 50 interviews where a complex living situation was described during the interview, these experts conducted qualitative, ethnographic-type respondent debriefings asking open-ended questions about where the person of interest had been staying over the last year. The goal of the respondent debriefing was to determine the "true" residence for each person. We compare this "true" residence against the residence status obtained by the questionnaire alone. By comparing the two sets of data, we determine if the questionnaire was collecting accurate information, and for what situations the questionnaire needs to be modified. This paper discusses our methodology for these expert respondent debriefings.

Key words: Conversational interviewing; Pretesting; Coverage Improvement

1. Introduction

In theory, experts in a domain of interest could perform in-depth interviews using scripted and unscripted questions to obtain information needed to answer research questions. These experts would feel confident that the data they collected were of high quality since they could supplement the scripted probes with unscripted probes to determine the truth (see Conrad and Schober, 2000; and Schober and Conrad, 1997, for studies on “conversational” interviewing). If the experts felt the respondent was confused by questions, or if the respondent indicated there was more pertinent information available, then experts could inquire further. This is largely the method employed by survey methodologists doing cognitive testing of a questionnaire and the method doctors use to diagnose and treat patients. However, this method is impractical for a large survey data collection due to the expense of training interviewers to become experts.

The survey industry has adopted a form of respondent debriefings as common practice for understanding how questions work (see Campanelli, Martin., and Rothgeb, 1991; DeMaio, 1983; DeMaio and Rothgeb, 1996; Esposito, Campanelli, Rothgeb, and Polivka, 1991). Respondent debriefings are typically scripted questions added to the questionnaire by survey designers and asked by regular survey interviewers either after the question of interest or at the end of the survey questionnaire. Additionally, the survey industry has recognized the value of observational studies conducted by “members of the survey staff who have been involved in planning the survey design, questionnaire, data analysis, or interviewer training” (DeMaio, 1983, pg. 103). These observations allow an assessment of how the survey is working by observing the interaction between the interviewer and respondent and noting particular behaviors or circumstances. Cognitive interviewing is another well-known and accepted form of pretest by which social scientists usually conduct a semi-scripted interview with the goal of understanding how respondents comprehend and answer the survey questions (see Willis, 2005, for a more detailed explanation of the cognitive interview). Finally, retrospective debriefings are often done by researchers in establishment surveys and censuses on reported data on self-administered forms. The goal of retrospective debriefings is to learn about response strategies, data sources, definitions, and to assess whether the answers met the intent of the question (Willimack et al., 2004).

In this paper we propose a combination of these methods (traditional respondent debriefing, observational studies, cognitive testing, and retrospective debriefings) that allows questionnaire design and subject matter experts to be the source for rich observational reports on both regular interview observations as well as brief qualitative interviews they conduct with actual respondents in a field production survey. This is similar to the approach described by Belson (1981), but instead of using specially-trained interviewers, we advocate having experts on the subject matter conduct these interviews. When the survey topic is complex, training a third party in the intricacies of exactly what is being measured may introduce error into the evaluation tool. We argue that this method, which we call “expert respondent debriefings,” can be used as a means of evaluating the quality of production survey data. This is especially important for surveys on complex topics for which an assessment of “truth” is not otherwise available. This paper discusses our experience with this method in a recent test of the decennial census. We assess what went well, and we make recommendations for future implementation of this method.

2. Background

To measure the accuracy of the decennial census, the U.S. Census Bureau has historically conducted an independent survey operation to measure housing unit and person coverage of the census. For 2006, this is called the Census Coverage Measurement (CCM) operation. Concerning person coverage, the operation results in estimates of erroneous enumerations (people counted in the wrong place) and omissions (people missed) in the census (see Hogan, 1993). With recent advances in computer matching, new emphasis is also placed on finding people counted twice (a type of erroneous enumeration) within the census.

A test of the census procedures proposed for the 2010 Census was conducted in two sites during the 2006 Census Test: selected areas within Travis County, TX (near Austin), and the Cheyenne River Sioux Indian Reservation in SD. During the 2006 Census Test, the Census Bureau also tested the CCM operations relating to measurement of person coverage. The 2006 initial CCM survey was called the Person Interview (PI). We used expert respondent debriefings during the CCM PI operation as a method of evaluating its questionnaire to make improvements for the CCM operation in the 2010 Census.

During the 2006 Census Test, the PI was conducted using a laptop Computer Assisted Personal Interview (CAPI) instrument. The PI had undergone cognitive testing and extensive path testing prior to being fielded. In addition to the expert respondent debriefings, more general data analyses and behavior coding were planned to occur once the operation was complete.

Even considering the cognitive testing and path testing described above, the following question remained without a firm answer: *Does the CCM PI collect enough information and the right information to confidently determine whether each person rostered was correctly enumerated at one and only one place?* Doubts were raised about the data quality of the 2000 census coverage operation after additional searches for duplicates were conducted using census data, and many of these duplicates were thought to be misidentified as correct enumerations during the 2000 coverage operation (Feldpausch, 2001; Martin, Fay, and Krejsa, 2002). We needed to evaluate the quality of the data collected by the 2006 PI questionnaire so that in the 2010 Census we would have confidence in the coverage measurement data.

The PI gathers a roster of people living or staying in sampled housing units at the time of the interview, which happens to be three or more months after Census Day (April 1st). A series of questions is asked about where these people lived on April 1st and whether there was any place else where they stayed frequently and might have been counted (e.g., in college, with a relative, in a seasonal home). Because the 2000 coverage instrument failed to identify possible duplicate people in the 2000 Census, the redesigned PI contained additional questions trying to determine if there was any other place these rostered people could have been counted in the 2006 Census Test. We hypothesized that determining where most people should be counted on Census Day would be easy, since most people have a single place to live and sleep. For these people, the questionnaire would flow quickly, and there would be no questions remaining in our minds about where the people should have been counted.

For more complex situations (e.g., people who move often or people who are tenuously attached to several households), we hypothesized that the living arrangements might not be as clear. We wanted to know whether the PI survey questionnaire collected enough information and the right information to code each person's residence status correctly, particularly in complex situations (see residence rules outlined in National Research Council, 2006). We thought, "If only experts in residence rules could interview these people, then we would know where they should be counted." The expert respondent debriefing methodology evolved from this idea.

3. Respondent Debriefing Methodology

In July of 2006, a small group of experts in Census residence rules observed PI interviews at the two sites of the 2006 Census Test. Immediately following interviews during which a complex living situation was described, these experts conducted qualitative, ethnographic-type respondent debriefings asking open-ended questions about where the persons of interest had been staying over the last year. The goal of the respondent debriefing was to determine the "true" residence for each person. The remainder of this section describes the methodology in more detail.

Nine experts went into the field during the PI operation and observed in-person PI interviews. Two experts went to the Sioux Indian Reservation, and the other seven went to Austin¹. Seven of the nine experts were on the team that designed the PI instrument and had intimate knowledge of the data needed. These seven included two survey methodologists (the authors of this paper and the project leaders), three analysts who do actual coding of people as erroneous or correct enumerations during census operations, and two statisticians who were involved in the survey design and will be involved in the data analysis. The two other experts were survey methodologists who had no experience on the PI *per se*, but who are extremely knowledgeable about the types of people missed and duplicated in the census historically, understand good questionnaire design, and are fluent in the census residence rules. Eight of the nine staff were familiar with cognitive interviewing techniques and had been trained in that method.

We held a two-hour training session prior to going into the field. During the training session, we talked about what type of respondents the experts should debrief and when the debriefing should occur. We reviewed a list of possible respondent debriefing probes we created for the project. We held a mock interview and respondent debriefing, with a discussion following it.

During the training, it was critical for us to define who should be debriefed and under what circumstances, but it was difficult to do so with precision. We suggested experts debrief any situation where there seemed to be additional residency information about an individual than was captured in the PI. Some indications of complex living situations where there might be "more to the story" included: (1) the person does not seem to have a regular "pattern of staying" between two places and has difficulty with the residency questions; (2) the respondent starts to list someone for the roster, then changes his or her mind; or (3) the respondent says something about another residence, but it does not get recorded for one reason or another. Even though we gave specific examples, the vagueness of our instruction, "debrief if there is more to the story" was difficult, especially for the two experts who were not on the original design team for the survey.

¹ There were more cases in the sample in Austin than there were in the South Dakota sample, so we assigned most of the experts to Austin.

Although we had a hunch that it would be obvious during the interview when experts needed to ask more questions, we simplified the instruction for them and told them to debrief any situation where another residence was mentioned.

During training, we briefly talked about conducting a respondent debriefing on straightforward cases to see if the PI missed anything. We decided not to do this for two reasons: 1) we did not know how many *real* respondent debriefings the experts would need to do; and 2) we were worried about the increased burden in straightforward cases. We thought debriefing these relatively simple situations might be awkward. Since there were so many probes already in the PI about possible people who could have been missed and possible other residences where someone could have been counted, we were concerned that crafting a unique probe for the debriefing would amount to something like, “Are you really sure?,” which sounds somewhat accusatory and would not be likely to elicit additional information.

The goal was to conduct 60 respondent debriefings. We wanted our group of experts to see as many cases in the field as possible. Based on previous research on mobility (see Bates and Gerber, 1998), we estimated that about 25 percent of cases would have situations “interesting” enough to merit a respondent debriefing. Based on past census observation experience, we estimated that each person would observe three to four completed interviews each day. To conserve resources, we decided to have the experts tape record the interviews (with respondent permission) for behavior coding purposes, rather than having additional staff assigned to do the taping. Though we discussed using payment as an incentive for respondent debriefings, we decided not to do so. Instead, the experts brought census logo magnets and pens to give out as tokens of appreciation.

The group of experts went into the field during the first week of PI interviewing and observed for approximately seven consecutive days. Each day, each expert accompanied a different interviewer into the field and went to different parts of the test site depending on where the interviewer’s case load was located. Since the interviewer assignments were made by the regional office, we had little control over the characteristics of the interviewers that the experts observed, other than requesting to see a variety of interviewers and socio-economic areas. Experts were given a list of the interviewers, their phone numbers, meeting locations, maps to the meeting places, and the phone numbers of the interviewers’ direct supervisors. This worked out extremely well since each expert was able to confirm via telephone with the interviewer the day before the assigned observation, and the expert could call the supervisor if plans fell through and she needed an alternate interviewer. Once each expert met up with an interviewer, she spent a few minutes reviewing the respondent debriefing and taping procedures and the purpose of both. Experts provided a one-page handout that explained the procedures to each interviewer for their records.

Crucial to the success of the overall method was the reassurance each expert gave to the interviewers that the experts were not there to critique them, because many of the interviewers were skeptical at first. The experts encouraged the interviewers not to change their style of interviewing because the experts were observing. Our goal was to see how well the questionnaire worked given “real life” circumstances.

Each expert then accompanied an interviewer into the field for the day and watched the cases he or she attempted. The interviewer introduced the expert to the respondent at the beginning of the interview, and the expert asked the respondent if it was okay to tape record the interview to help us evaluate and possibly improve the questions. The expert listened to the interviews and, if she heard an indication that there might be more information than had been recorded (e.g., about another residence, the time spent at each address, other people associated with the residence) she made a note and waited until the end of the interview.

When the official interview was concluded, the expert asked a few more questions trying to clarify what she had heard so that she would have enough information and the correct information to be able to determine where the person should have been counted on Census Day. It was important that the probes be unbiased and open, so often she would say, "I heard you say THIS, can you tell me more about that situation?" or "What were you thinking about when the interviewer asked you ...?" or "In your own words, can you describe how often and when NAME stayed at LOCATION." After the expert finished asking questions, she thanked the respondent and gave a token gift. In essence, the expert was trying to determine or clarify the "truth" about the residence status of individuals using a conversational interviewing style.

The interviewer was allowed to listen to the respondent debriefing questions (because the expert needed to accompany the interviewer after the debriefing for the remainder of the day), but the experts reminded the interviewer not to change his or her style in later interviews. The experts also reminded interviewers not to write additional notes for the case based on what they heard.

Each expert took brief notes on all the completed cases she saw, including the numeric identifier for the case, how many people were listed in the housing unit, whether the case was taped, whether the case was debriefed, any other short notes about the instrument generally (e.g., questions or skip sequences that did not seem to work as intended even if the case was straightforward), and brief notes about any respondent debriefing that was conducted. Each expert wrote up a more extensive respondent debriefing report for each debriefing during each evening after the interviewing was complete. This debriefing report needed to be specific enough that someone could read the report and know where the person in question should have been counted on Census Day, April 1st, 2006, according to the census residence rules.

The experts in Texas met each morning to review any interesting situations encountered the day before, as well as strategies for taping and asking questions. This allowed the experts to be consistent in their approaches and pass on tips for what worked well, since this was our first experience using this method.

4. Results

The nine experts observed 169 interviews with respondents in occupied housing units. Fifty respondent debriefings were conducted. The group did not meet our expectation of seeing three to four completed interviews a day; instead, each expert saw on average only two to three a day. However, our estimate of 25 percent of the cases being interesting enough to conduct a respondent debriefing was close, as the group debriefed about 30 percent of the cases observed. There was some variability in how many respondent debriefings were conducted by each expert.

One expert conducted only two, while another expert conducted twelve. In several cases, the experts debriefed because another residence was mentioned, even though it was a straightforward case and it was fairly obvious that the PI questionnaire collected all the information necessary for coding. In many of these cases, the experts were simply confirming that the case was as straightforward as it had sounded. Given this, one should not conclude that 30 percent of the units the experts observed had an unusually complex living situation, but rather 30 percent of those cases had some sort of interesting characteristic about them.

Although most of the time it was fairly obvious by what the respondent said which cases should be debriefed, sometimes nonverbal cues were a powerful indicator that there was more information to uncover. For example, a respondent counted on his fingers when the interviewer asked whether anyone else lived at the unit in March and April of 2006. This respondent answered no to that question, but during the debriefing, when probing him about the gesture and what he was thinking, the expert discovered that there was someone else living at the unit during March and April. Because that person had moved out, the respondent had decided not to mention him or her. In this case, the nonverbal cue (counting on fingers) was indicative of a person who should have been rostered, but who had not been captured by the PI.

Experts were trained on picking up verbal cues, such as, “My grandkids stay here sometimes; you don’t want to count them do you?” However, they were not trained on picking up nonverbal cues such as gestures or facial expressions. Because most of our experts have also been trained in cognitive interviewing, they were familiar with detecting nonverbal indicators of problems with understanding. Identifying nonverbal cues is a little bit of an art, similar to knowing when not to probe due to sensitive situations (e.g., a messy divorce situation or a hostile respondent who was hesitant to participate in the PI at all). We conjecture that both verbal and nonverbal cues are important to identifying cases where there might be “more to the story.” This is part of the reason a background in cognitive interviewing is a desirable characteristic for experts conducting respondent debriefings.

We feel fairly confident that our method of using both verbal and nonverbal cues identified most of the interesting situations. In 118 of the cases observed², experts felt the PI instrument adequately captured where the people should have been counted on April 1, 2006 and no debriefing was necessary. After the PI operation, we were able to have an outside source (i.e., analysts from the National Processing Center of the Census Bureau) verify whether or not the PI did collect enough information to code residence or whether a follow-up interview was necessary. Cases containing people whose residence status is still unclear after the PI operation are typically sent back to the field for a followup interview that attempts to obtain additional information in order to code residence status. Out of the 118 cases where experts were confident that the PI instrument had functioned adequately, we were able to look at the data for 116 of the cases³. Out of those 116 households (about whom experts thought enough information had been recorded in the PI), only one was sent back because the PI missed information about other

² In one additional case, the expert requested a debriefing interview; however, the respondent refused. Thus, in fact, 119 cases were observed without a debriefing interview.

³ The case identifier for two cases was recorded incorrectly, so we could not look up those cases.

addresses where they may have been counted⁴. Thus, the respondent debriefings by experts were able to detect almost all of the problematic residence situations.

After their time in the field, each expert wrote a trip report, and the experts held a series of meetings to talk about changes we recommended to the PI questionnaire and operation based on their observations. Although the group recommended numerous changes, these were mainly enhancements rather than changes resulting from major problems. Proposed changes to the PI pertained to question wording, order, and skip problems, usability of field materials, and interviewer procedures. The experts also found a few gaps in the instrument, which slipped through cognitive and/or path testing (e.g., respondents did not understand the question that intended to gather the name of the resident householder; instead they sometimes gave the name of the non-resident landlord). Generally the PI instrument worked well, yet the number and quality of the changes recommended was a surprise and an added bonus from this respondent debriefing operation. Additionally, the list of changes was put together rather quickly, which was an advantage in this situation since preparations for the automated PI for the 2008 Dress Rehearsal were already underway. An argument could be made that the debriefings should be incorporated into other operations simply for the added review of the questionnaire beyond cognitive and path testing. Some things cannot be learned until *real* interviewers ask the questions of *real* respondents.

5. Lessons Learned

Conversational interviewing worked for this project. The method turned out to be very effective, easy to administer, and natural. Most of the time, asking unscripted probes at the end of the interview uncovered or confirmed the true living situation of the people in question. It also was not burdensome on respondents, since typically the experts only had a few more questions taking an additional minute or so of their time. The probes made sense to both the respondent and interviewer since the probes were following up on, or clarifying, something said earlier in the interview.

It is important to reassure interviewers that experts are there to see how the questionnaire works, not to evaluate them. Some interviewers were initially skeptical about the observations and debriefings. Although the experts were prepared with a handout and brief discussion explaining the purpose of the observation, their hesitation only dissipated after hearing some of the debriefings. Some interviewers wanted to change their interviewing techniques based on the debriefings. After hearing some of our probes, some interviewers wanted to do additional probing during the interview to clarify the types of things they heard us clarify in the debriefing. Additionally, some interviewers wanted to make notes in the questionnaire about information uncovered during the debriefing. Experts discouraged this behavior (because we needed to assess the PI as it was, without enhancements, in order to make changes for 2008) but we believe it is not preventable. Good interviewers will try to learn from the example set by the experts. This is part of the reason why we make the next suggestion.

⁴ A few other situations were sent for follow-up, but not for reasons that could have been detected by listening to the interview. For example, some cases were sent back because alternate addresses provided by respondents could not be found on our maps.

Make sure experts observe a variety of interviewers conducting the survey. If only the best interviewers are observed, a biased view of how the survey works will result. Similarly, if only a small number of interviewers are observed, generalization of observational data will be difficult (i.e., it will be impossible to tell what was an individual quirk of an interviewer and what was caused by the questionnaire itself). During the field period, some experts went out with a few of the same interviewers. This overlap also makes it difficult for experts to identify generally applicable usability and cognitive issues. Additionally, the more observations that come from a single interviewer, the more the study will be impacted if interviewers change their behavior based on what they learn from the debriefings.

Experts should have some training in cognitive interviewing methods and have subject matter knowledge. All of our experts (except one) had been trained in cognitive testing, and they all intimately knew the subject matter. Thus, they knew what to look for (subject matter) and could ask unbiased questions (cognitive interviewing) to get at the truth. There was some pressure to include newly-hired people as experts to give them experience in the field, but we believe our results would not have been as useful if we had done that.

While in the field, a daily meeting time with other experts builds synergy and leads to additional improvements in the survey and debriefing procedures. The morning meetings with the experts were beneficial. The group recounted both the previous day's successes and issues, including probes used, complicated situations encountered, problems with the PI itself, or field issues that interfered with obtaining interviews. This time to compare notes with other team members helped in several ways: 1) corroborating evidence of problems; 2) knowing what issues had already been identified with the PI instrument helped the experts reduce their attention on already-known problems and re-focus on issues not yet encountered; 3) hearing about other experts' experiences with taping and observing gave the group some helpful hints on how to gain cooperation from the respondents as well as the interviewers, and 4) learning about cues that others followed up on helped the experts focus on situations that had come up as "potentially problematic." Sometimes it was difficult based on a single observation to know if a particular problem was an instrument error or a respondent or interviewer misunderstanding. Sharing experiences helped the expert team members broaden their understanding of what cues or "symptoms" indicated the need for further probing.

Tape-record the interviews for behavior coding purposes. From a logistical point of view, having observers do the tape-recording frees the interviewer from handling both a tape recorder and a laptop or from having additional staff tape-record. Most of the experts did not listen to the tapes when writing up conclusions from the trip, but the tapes were used for behavior coding of the PI. Most respondents easily consented to give permission to tape record. Of the 169 interviews observed, experts taped 154. Seven respondents refused to be taped for various reasons, often including privacy concerns.⁵ Most of these were reluctant respondents to begin with. We attribute the ease of gaining consent to the fact that people are used to having phone conversations taped for quality assurance. This was a very similar situation. Interestingly, experts found that as long as they treated taping as a "matter-of-fact" part of the process, respondents usually complied.

⁵ Other interviews were not taped due to the interviews being conducted in Spanish, the observer being out of tapes, observing interviews with non-household members and a respondent who was under 18 years of age.

Not providing an incentive for the respondent debriefing was a good decision. Prior to going into the field, operations staff had considerable deliberations about whether or not payment was necessary to get respondents to agree to the respondent debriefings. We decided not to offer payment, but instead to give tokens of appreciation, like Census Bureau pens and magnets. Experts found that asking a few questions immediately after the interview was incredibly natural and easy. It was also a lot quicker than we had anticipated, taking only a few minutes for each debriefing, versus the 10 minutes per interview that we had requested and received in our approval from the Office of Management and Budget. Offering to pay the respondent to answer a few additional questions would have made the whole exchange much more awkward. Although we could have explained why some respondents were paid and some were not, the difference between the two experiences was so minor it would have been hard to justify. The respondents seemed to appreciate the token gifts; and, upon departure there was a feeling of good will between the interviewers, experts and respondents with most of the observed cases.

Put together a list of changes to the survey based on observations. As mentioned previously, the group recommended additional changes to the PI. These issues might not have been identified at all had this study not taken place.

Be prepared to evaluate non-English speaking interviews. In Austin, the experts encountered a number of Spanish speakers but were not able to debrief them since none of our experts were fluent in Spanish. Including fluent Spanish speakers in our group of experts would have been beneficial, though we are limited when it comes to finding experts who fit all three conditions: 1) fluent Spanish speaker; 2) expert in residence rules; and 3) knowledgeable about cognitive interviewing methods.

6. Future Research

Research on this project is still underway. The goal of the respondent debriefing was to determine the “true” residence for each person. Next, we will compare this “true” residence against the residence status obtained by the CCM PI questionnaire alone. In situations where the two match, we assume the questionnaire collected accurate information. Where the two do not match, we assume the questionnaire either collected too little or faulty information. Thus, we can determine whether the questionnaire was adequate, and for what situations, if any, the questionnaire still needs to be modified.

Acknowledgements

We thank our experts, Betsy Martin, Laurie Schwede, Tammy Adams, Jamie Burnham, Julie Bibb, Sandy Norton, and Vicki Smith; Lynn Imel for data preparation; and Terry DeMaio, Elizabeth Murphy, Magda Ramos, Diane Willimack and Kathleen Ashenfelter for their reviews of this paper. We also thank the Decennial Statistical Studies Division for sponsoring this project and being receptive to its findings.

References

- Bates, N. and Gerber, E. R. (1998). Temporary mobility and reporting of usual residence. *Survey Methodology*, 24 (1): 89-98.
- Belson, W. A. (1981). *The Design and Understanding of Survey Questions*. Aldershot, U.K.: Gower.
- Campanelli, P.C., Martin E., and Rothgeb, J. (1991). The use of respondent and interviewer debriefing as a way to study response error in survey data. *The Statistician*, 40, 253-264.
- Conrad, F.G., and Schober, M.F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64, 1-28.
- DeMaio, T. J. (ed.) (1983). *Approaches to Developing Questionnaires*, Statistical Policy Working Paper 10. Washington, DC: Office of Management and Budget.
- DeMaio, T. and Rothgeb, J. (1996). Cognitive interviewing techniques in the lab and in the field. In N. Schwartz and S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey Bass.
- Esposito, J., Campanelli, P.C., Rothgeb, J., and Polivka, A. E., (1991). Determining which questions are best: Methodologies for evaluating survey questions. *Proceedings of the American Statistical Association (Survey Research Methods Section)*. Alexandria, VA American Statistical Association, 46-55.
- Feldpausch, R. (2001). "Census Person Duplication and the Corresponding A.C.E. Enumeration Status." Executive Steering Committee for A.C.E. Policy II, Report 6. U.S. Census Bureau. October 13, 2001. <http://www.census.gov/dmd/www/pdf/Report6.PDF>
- Hogan, H. (1993). The 1990 Post-Enumeration Survey: Operations and Results. *Journal of the American Statistical Association*, Vol. 88, No. 423 Undercount in 1990 Census.
- Martin, E., Fay, R., Krejsa, E., (2002). "Analysis of Questionnaire Errors in Survey Measurement of Census Coverage." Paper Presented at the Annual Meetings of the Section on Survey Research Methods of the American Statistical Association.
- National Research Council (2006). *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census*. Panel on Residence Rules in the Decennial Census. Daniel L. Cork and Paul R. Voss (eds.) Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Schober, M. F., and Conrad, F. G. (1997). "Does conversational interviewing improve survey data quality beyond the laboratory?" Paper Presented at the Annual Meetings of the Section on Survey Research Methods of the American Statistical Association.
- Willimack, D., Lyberg, L, Martin, J., Japac, L, and Whitridge, P. (2004). Evolution and Adaptation of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys. In S. Presser et al., (Eds.) *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
- Willis, G. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.