

## DOES THIS QUESTION WORK? EVALUATING COGNITIVE INTERVIEW RESULTS USING RESPONDENT DEBRIEFING QUESTIONS<sup>1</sup>

Kristen A. Hughes and Theresa J. DeMaio  
United States Census Bureau, Washington, DC 20233

### INTRODUCTION

The quality of data collected in a survey rests, among other things, on the notion that the meaning of the question as written by the questionnaire designer is consistent with the way respondents interpret it. The intended meaning of a survey question or term is not always the meaning that respondents recognize. Cognitive interviews are a useful, time- and resource-efficient pretesting tool to identify potential sources of mis-communication between survey designer and respondent before a survey instrument is fielded. Relatively small numbers of interviews are conducted in a research setting, typically with experienced researchers and potential survey respondents.

Respondent debriefing is another effective tool for identifying questionnaire problems, which is utilized at a later stage in the questionnaire design or evaluation process. Respondent debriefing questions are generally fielded on a larger scale than cognitive interviews, in a field test or in the survey itself. These larger data collections are advantageous in that they enable researchers to talk with a more representative sample of respondents, and produce quantitative assessments of questionnaire performance.

We believe that cognitive interviews and respondent debriefing are complementary methods, which provide the same kinds of information in different settings. In this research, we use respondent debriefing questions administered in a field survey to assess the results of cognitive interviews conducted during questionnaire development. This paper compares the results of these two pretesting methodologies to provide insights about the usefulness of cognitive interviews and the level of consistency between cognitive interviews and respondent debriefing in evaluating and improving survey questions.

In the following sections of the paper, we: 1) review the literature to see how respondent debriefing questions have been previously used in survey development; 2) lay out the objectives of the present research; 3) describe the methods used in the research; 4) presents the results of cognitive interviews conducted in the lab compared with respondent debriefing questions administered in the field; and 5) offer some conclusions about the complementary nature of the two methods.

### LITERATURE REVIEW

Currently, the most widely used pretesting tool is

the cognitive interview (Tourangeau, et al., 2000). Although there has been much success in using cognitive interviews to “pretest” questionnaires, they do have their limitations. Mainly, only a very small number of purposively chosen respondents are interviewed and it would be impossible to collect all combinations of possible answers (Hess and Singer, 1995). In addition, cognitive interviewing methods vary widely, as do the styles of analyzing the data collected from such interviews.

Researchers have, for several years, used debriefing questions in a standardized interview context to assess respondents’ comprehension of survey questions and concepts (Esposito, et al., 1992; Cannell, et al., 1989). DeMaio and Rothgeb (1996) argue that cognitive interviews and respondent debriefing questions can fulfill the same purpose, with input at different stages of the questionnaire development process. However, little research has been done to assess this method’s ability of evaluating survey questions. An exception is Hess and Singer (1995), who argue that debriefing questions can be used as a cognitive tool to assess the quality of the interview, and that because they can be administered to a large and representative sample, they are an important “back-end” supplement to the one-on-one interviews used in the early phases of development.

### RESEARCH OBJECTIVES

The research objectives focus on using respondent debriefing questions to evaluate the revisions that resulted from the cognitive interviews. Specifically we were interested in answering three questions:

1. Were the revisions based on the cognitive interview results successful in a systematically-sampled, large scale environment?
2. Did the “unfixable” problems identified in the cognitive interviews surface in the field?
3. Did any “unexpected” problems surface in the field that were not identified in the cognitive interviews?

### RESEARCH METHODS

#### *Questionnaire*

The vehicle for this research is a series of computer crime questions proposed for inclusion in the National Crime Victimization Survey (NCVS) for the first time by its sponsor, the Bureau of Justice Statistics (BJS). The computer crime series consists of seven questions and asks about personal use of a computer,

---

<sup>1</sup>This report is released to inform interested parties of research and to encourage discussion. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications.

use of the Internet, computer-related incidents that the respondent may have experienced, any monetary loss incurred, and the reporting of any incidents mentioned. As part of the developmental work on these questions, an expert review was conducted.

*Cognitive Interviews*

Fifteen cognitive interviews were conducted in the Washington, D.C. metropolitan area and Respondents ranged from seventeen to seventy-two years of age. Three of the fifteen respondents were more than fifty years of age. We interviewed 8 White, 6 black, and 1 multiracial respondents; 5 males and 10 females. Our respondents included high school and college students, working and retired people. All of the respondents had at least some computer experience and two of the respondents used the computer for a home business.

A single round of interviews was conducted using a questionnaire that included the computer crime questions as well as some earlier NCVS screening questions. These interviews were conducted using the “think-aloud” method along with concurrent probing. All interviews were audio-taped and summarized. based on the taped interview. Analysis of the interviews was based on the summaries.

*Respondent Debriefing*

In an attempt to follow up on the results of the cognitive interviews, respondent debriefing questions were added at the end of the 2001 NCVS CATI instrument containing the computer crime questions. The NCVS is a longitudinal survey sponsored by BJS in which respondents are interviewed, either in person or by telephone, every six months over a period of 3 ½ years. Only respondents who were interviewed by telephone and in their final interview received the debriefing questions. This allowed a random sample of respondents to receive the debriefing questions, while not exerting extra burden on the respondents who would be re-contacted in subsequent interviews. Questions were asked in September through December of 2001. Respondents who reported that they had used a computer in the last 6 months were administered these questions. A total of 1644 cases were included in the respondent debriefing study.

**RESULTS**

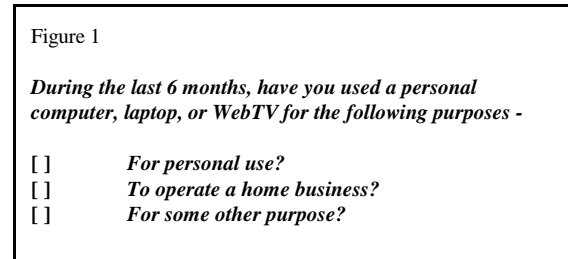
We present the results in three sections, which correspond to our three research objectives. First, we address problems identified during the cognitive interviews that we attempted to remedy through questionnaire revision. Second, we follow up on problems noted in the cognitive interviews, for which a solution was not obvious. Third, we investigate whether new problems surfaced in the respondent debriefing study that were not evident in the cognitive interviews.

**“Fixable” Problems - Were the problems that we deemed “fixable” really fixed?**

The objective of the computer crime questions was to collect information about computer incidents that

occurred while the respondent was using a computer for “personal use.” Many respondents reported in the cognitive interviews that they sometimes use a computer at work for personal use. If an incident did occur at work, we wanted to see if respondents were separating “personal use at work” from “official work use at work.”

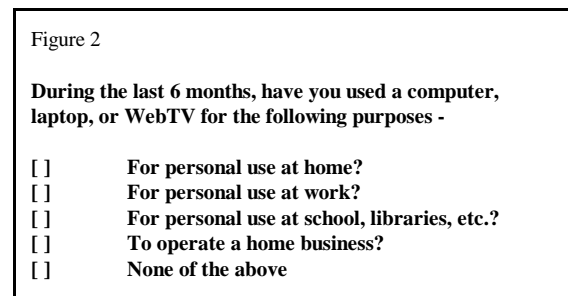
The original question that was intended to elicit reports of personal use of a computer is included in Figure 1. These categories, especially the “for some other purpose” category, captured all possible types of computer use.



In the cognitive interviews, we found two major problems with this question. The first was a domain issue. Many respondents were confused about whether the question was asking them to include only computers in their home or to also include computers that they use in other places. This confusion is due in part to the references to personal computers and “personal use” in the introduction to the question series and to the use of those terms in this question. We asked respondents if they ever used their work computer for personal use. Five of the 15 respondents reported that they did so. However, they did not include that use in their answers because they thought it was out of scope.

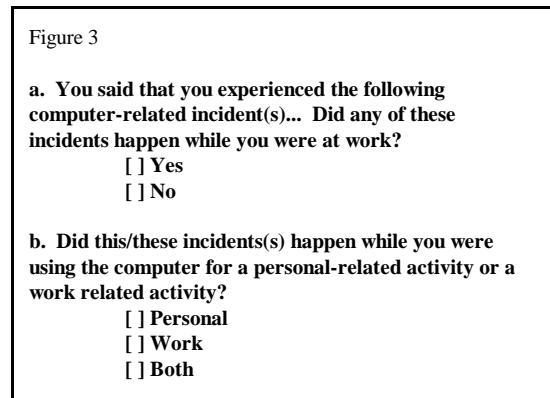
Second, the response categories did not adequately capture the information that the sponsor wanted. Specifically, the "for some other purpose" category captured activities that respondents should have included in the "personal use" category.

Based on the findings from the cognitive interviews, the question and response categories were reworded to focus on personal use regardless of where it occurs and on home businesses, which were also considered in-scope (See Figure 2). These four categories covered all types of personal use intended by the sponsor. We deleted the "for some other purpose" category, because we did not want respondents to think that any other purpose, such as work, was relevant.



We asked two debriefing questions to see if the

changes more successfully conveyed the purpose of this question, and to determine if respondents were only counting those incidents that happened while they were using the computer for personal use (See Figure 3). First, we asked respondents who experienced a computer-related incident whether they experienced it while at work. Respondents who experienced an incident at work were then asked whether they were using the computer for work or personal use at the time of the incident.



The results of these debriefing questions, shown in Tables 1 and 2, illustrate that the changes did not remedy the problem. In fact, they introduced a new problem. Instead of excluding the use of a work computer for personal use, respondents in the field survey included work use of a work computer. Almost 70 percent of the respondents who reported that a computer-related incident occurred at work, said that they were using the computer for a work-related activity. These data suggest that respondents were still making errors in deciding what type of computer use was in scope for the survey. Thus, the questionnaire revisions we made, based on findings from cognitive interviews, did not work in the field interviews.

**“Unfixable” Problems**

Another way in which respondent debriefing can be used is to see if the problems that surfaced in a small scale cognitive interviewing environment were replicated in a field environment with randomly selected respondents.

*Access to computers*

Respondents in the cognitive interviews were asked the following question: *How many computers do you have access to for personal use or for operating a home business?* Although the question was intended to encompass all personal use of computers, it did not. Five of the 15 respondents included only the computers that were in their homes, even though they also used other computers for personal use. However, this was not uniformly the case. Three of the respondents gave an answer that included all the computers that they had access to, including places like a senior center or a graduate student lounge. One respondent said she had access to 8 computers and listed all the places that she

could potentially use the computer, including her father's office, school computer lab, and her boyfriend's and friend's computers.

Table 1. Frequency Distribution of Whether Respondents Who Experienced a Computer-Related Incident Experienced it While at Work

Yes	36.2 %
No	63.4
Don't Know	.4
Total N (473)	100.0

Table 2. Frequency Distribution of Whether Respondents Who Experienced a Computer-Related Incident at Work Were Using the Computer for Work or Personal Use

Personal	11.6 %
Work	69.8
Both	17.4
DK	1.2
Total N (172)	100.0

Although we knew from the cognitive interviews that respondents were not interpreting this question uniformly, we had no immediate solution to this problem. Respondents were not in error when they included a wide array of locations and we did not want to encourage overestimation of computer access by broadening the question domain. We did, however, review the responses to this item in the field interviews to see how broadly respondents interpreted the current question. Table 3 shows that slightly less than half of all respondents in the field interviews reported access to one computer. Almost a quarter reported access to three or more computers.

Table 3. Frequency Distribution of How Many Computers were Accessible by Respondents

<i>Access to:</i>	
No Computers	.9 %
One computer	47.2
Two Computers	29.8
Three Computers	11.1
Four or more Computers	11.0
Total N (1626) <sup>1</sup>	100.0

<sup>1</sup>Does not include 18 cases for which information was not obtained in this question.

We included one debriefing question (see Figure 4) to evaluate the performance of this question and to see if the same problems were recurring in a large-scale survey environment. That is, were respondents including only those computers that were in their homes, even though they used other computers for personal use?

Figure 4

Earlier, you said you had access to ... computer(s). In that number, did you include computers that you use -

- (a) At home?
- (b) At work?
- (c) At school?
- (d) At the library?
- (e) At friends' or relatives' homes?

Table 4 presents the results of this inquiry. Almost all respondents (96.0 percent) included computers at home. More than one-third of respondents (37.6 percent) included computers at work in their response. Smaller percentages of respondents included computers at school, at work, at the library, at friends or relatives homes', or at some other place.

Table 4. Percent of Respondents with Access to Computers Who Included Computers at Various Locations in Their Count of Accessible Computers

Computer Location:	Yes	No	DK	Total N
At home	96.0 %	3.9 %	.. <sup>1</sup>	1611 <sup>2</sup>
At work	37.6	62.2	—	1611 <sup>2</sup>
At school	14.3	85.6	—	1611 <sup>2</sup>
At the library	12.2	87.6	—	1611 <sup>2</sup>
At a friend's or relative's home	12.0	88.0	—	1611 <sup>2</sup>
At some other place	1.7	98.0	—	1611 <sup>2</sup>

<sup>1</sup>.. = less than 1 %

<sup>2</sup> Does not include 33 cases for which this information was not obtained

These findings are consistent with the findings from the cognitive interviews. Respondents in the field also interpreted the term “access” very widely. Many respondents only included the computers that they use at home. At the other end of the spectrum are those respondents who include their access to computers outside of the home, such as work, school, libraries, and friends and relatives. Thus, the results of the respondent debriefing showed that the problem that occurred during the cognitive interviews also occurred in the field.

*Lewd or obscene messages*

Respondents were asked whether they experienced any of a variety of computer-related incidents in the past 6 months. One of the incident types (unrequested lewd or obscene messages, communications, or images while online or through E-mail) elicited more positive reports in the cognitive interviews than any other. Five of our 15 respondents reported receiving lewd or obscene messages within the past six months. Four of these incidents referred to pornographic E-mails received either at home or at work. Respondents could tell the messages were obscene by the titles (e.g., “go here and see hot babes”). The fifth incident occurred while the respondent was online, looking for something on Yahoo, and a pornographic image “just popped up.”

Generally, respondents thought this referred to any

kind of pornography, encountered either through spammed E-mail or at an Internet site. Several respondents mentioned making a mistake and typing whitehouse.com instead of whitehouse.gov and being surprised to find they were at a pornographic website. They included this within the context of the question. E-mails containing obscene jokes, however, were not considered to be within the context of the question.

These interpretations by respondents reflect a much broader understanding of the concept of lewd and obscene messages than that intended by the questionnaire. The interest of BJS is in messages personally addressed and sent to the recipient individually rather than as spammed E-mails. This interpretation was not mentioned AT ALL by respondents, perhaps because they had more experiences of spammed pornographic messages, rather than personally targeted messages. In order to obtain reports of personally targeted lewd or obscene messages as desired by BJS, we felt that a series of questions would be required, including a broad initial question (such as this one) followed by more specific questions that focus on individually sent messages. This was not possible within the time and money available for these questions on the questionnaire. However, given the magnitude of response to this question, spammed lewd or obscene messages are clearly widespread, and more quantifiable data about the extent to which they bother the population was deemed worthy, in its own right.

We probed respondents in the cognitive interviews about whether they thought “messages” and “communications” meant the same thing. While some respondents thought there was no difference between them, many respondents made distinctions. Furthermore, respondents were not consistent in what they thought was a message versus a communication. The implication of these various views is that both terms are necessary in the question, even though the result is slightly wordy.

Based in part on these findings, the question was revised to read “unrequested lewd or obscene messages, communications, or images while online or through E-mail” and this wording was included in the field interviews. We do not think this change will be effective in narrowing the respondents' interpretation, since spammed pornographic messages are unrequested by the recipient. This change also has the potential to introduce more inaccuracy into the data, since some respondents often do not hear prefixes such as “un” and may think the question is asking about requested rather than unrequested lewd or obscene messages.

Table 5 presents the respondents' reports about experiencing computer-related incidents in the field interviews. It is noteworthy that the level of lewd or obscene messages is lower than would have been predicted based on our cognitive interviews. While 17.6 percent of respondents reported experiencing a computer virus attack, only 2.7 percent of respondents reported experiencing lewd or obscene messages.

Table 5. Percent of Respondents Who Reported Various Types of Computer-Related Incidents

Type of Computer- Related Incident	Yes
Fraud in Purchasing Something over the Internet	2.1 %
Computer Virus Attack	17.6
Threat of Harm or Physical Attack Made While Online or through E-mail	.5
Unrequested Lewd or Obscene Messages, Communications, or Images Made While Online or through E-mail	2.7
Software Copyright Violation in connection with a home business	0.0
Something Else You Consider a Computer-Related Crime	1.5
No Computer-Related Incidents	75.6
Total N (1502) <sup>1</sup>	100.0

<sup>1</sup> Does not include 142 cases for which this information was not

We included a series of three debriefing questions (see Figure 5) to collect interpretive information from respondents who reported “unrequested lewd or obscene messages or communications or images while online or through E-mail.”

Figure 5

- a. You told me earlier that you had experienced ‘unrequested lewd or obscene messages, communications, or images while online or through email.’ Did you include messages from x-rated or pornographic websites that were sent to your email address, but not to you personally? Sometimes this is also known as spam email.**  
 Yes  
 No
- b. Did you experience any other kind of “unrequested lewd or obscene messages” that were sent to you personally from another person?**  
 Yes - Can you describe that incident for me?  
 No

Twenty-nine out of the 41 respondents who reported that they had experienced an “unrequested lewd or obscene message” answered these debriefing questions. Table 6 shows the frequencies for this question. The overwhelming majority of these reports (86.2 percent of them) represented spam email. This supports our finding in the cognitive interviews that this question would not provide an accurate measure of personally-targeted pornographic email messages.

Table 6. Frequency Distribution of Whether Reported Lewd or Obscene Messages Included Spam E-mail

Yes	86.2 %
No	6.9
Don’t Know	6.9
Total N (40) <sup>1</sup>	100.0

<sup>1</sup>Does not include 12 case for which this information is missing

As Table 7 shows, only one respondent (2.5 percent) reported receiving an “unrequested lewd or obscene message” that was sent to him/her personally from another person. Almost all (95 percent) said that they did not experience any other type of “lewd or obscene messages.”

Table 7. Frequency Distribution of Whether Reported Lewd or Obscene Messages Included Mail Sent to Respondent Personally

Yes	2.5 %
No	95.0
Don’t Know	2.5
Total N (40) <sup>1</sup>	100.0

<sup>1</sup>Does not include 1 case for which this information is missing

An additional debriefing question attempted to gather qualitative information about the kind of “lewd or obscene messages” a respondent may have received. If respondents answered “yes,” they did experience another kind of “unrequested lewd or obscene message” that was sent to them personally from another person, they were asked to describe that incident. The one respondent who answered yes in Table 7 also misinterpreted this concept. This is evident by the fact that his/her open-ended response (“Porn always in the E-mail for respondent without being asked for.”) indicates that he/she was thinking of spam email. Thus, there were no reports of the type of computer incident of interest to the survey sponsor.

Respondents who reported that they had not experienced this incident, but reported some other kind of incident were asked a variation of the same question (See Figure 6).

Figure 6

Earlier we asked you about 'unrequested lewd or obscene messages, communications, or images while online or through email.' Do you think this would include messages from x-rated or pornographic websites that were sent to your email address, but not to you personally? Sometimes this is also known as spam email.

- Yes
- No

Almost one-half of the respondents (49 percent) said that they would include "spam E-mail" (See Table 8). These respondents were also misinterpreting the meaning of the question as it was originally intended.

Table 8. Frequency Distribution of Respondents Not Reporting Lewd or Obscene Messages Who Thought They Should Include Spam E-mail

Yes	49.0 %
No	46.5
Don't Know	4.5
Total N (198) <sup>1</sup>	100.0

<sup>1</sup> Does not include 102 cases for which this information was not obtained

In summary, respondents were asked debriefing questions regardless of whether they did or did not report receiving lewd or obscene messages. In all cases, the findings reiterate what we found in the cognitive interviews. That is, respondents were interpreting this question to include "lewd and obscene" spam E-mail as in scope.

### “Unexpected” Problems

Another way in which respondent debriefing can provide insight into questionnaire problems is to

identify unexpected problems – that is, those that were not identified during the cognitive interviews.

In the cognitive interviews, respondents easily and correctly interpreted the question asking whether they had been a victim of "Threats of harm or physical attack made while online or through E-mail." However, when analyzing the results of the debriefing questions, we found that respondents were not comprehending this category as intended.

In the cognitive interviews, we did not elicit any reports of incidents of this type. Our conversations with respondents suggested that there was fairly general agreement among respondents about what constituted a threat of harm or physical attack. It included notes or messages through E-mail or in chat rooms with content that was threatening, nasty, harassing, or vulgar. Most respondents did not mention anything about whether the

sender of the messages was known to the recipient. Those who did, generally thought the sender was a stranger, and one respondent made a distinction between incidents with known versus unknown senders.

Since there were no reports of this type of incident in the cognitive interviews, it was impossible to determine whether or not people would report the right kind of incidents. So the respondent debriefing questions in the field interviews attempted to gather specific information about "threats of harm or physical attack" for respondents who reported this type of incident. Eight respondents reported exposure to this type of incident, and these respondents were asked, "Earlier you said you experienced 'threats of harm or physical attack made while online or through E-mail.' Can you briefly describe that incident for me?"

Six of the eight respondents who answered positively to this category in the NCVS answered this debriefing question as follows:

- ✓ "Message said that if they saw you they will beat you up."
- ✗ "E-mails from unknown persons - one from Nigeria."
- ✗ "Guy online said that California is a crappy place to live."
- ✗ "Hate E-mail that was like a virus that was sent throughout the company at multiple locations."
- ✗ "Just in reference to the porno E-mail."
- ✗ "Obtained password and changed profile to lewd information."

Based on these answers only 1 respondent (✓) interpreted this question correctly. The other five (✗) incidents clearly have nothing to do with "Threats of harm or physical attack." Rather than being included in this category, they should have instead been reported in the "something else you consider a computer crime" category. Although the number of cases is small in this analysis, it suggests that respondents incorrectly interpreted the "threats of harm of physical attack" phrase, and responses to this question are of poor quality.

### CONCLUSIONS

Previous literature in the pretesting area has documented the results of cognitive interviews in improving survey questionnaires, and independently documented the results of respondent debriefing questions in understanding survey results. In this research, we have used respondent debriefing questions to assess results based on cognitive interviews.

It is important to restate that the purpose of this study is not to decide whether one method of pretesting is superior to another. Instead, it is important to view these methods as complementary. As we mentioned earlier in the paper, cognitive interviews do not and can not always identify all problems in a questionnaire. It was our intent to find other, time- and resource- efficient methods to provide data that would supplement the information gained from the cognitive interviews.

First, we wanted to determine if one of the revisions we made based on the cognitive interviews was successful. This revision did in fact solve the problem discussed, but ended up introducing a new problem.

Second, we wanted to see if the “unfixable” problems that surfaced in the cognitive interviews would resurface in a systematically-sampled, large scale environment. We presented problems with respondents’ interpretation of “access to computers” as well as “unrequested lewd and obscene messages,” both of which were identified in the cognitive interviews. While the initial impulse would be to remove problematic questions, there might be legitimate reasons for including them in the survey. As predicted, both of these problems surfaced in the field. With the respondent debriefing information obtained, we now have additional data to re-evaluate and improve these questions.

Third, we wanted find out if there were any problems that were not identified in the cognitive interviews. In this case, there was one question that surfaced in the respondent debriefing and not in the previous interviews. Without these data, this problem may have never been identified. With these data, caveats about the quality of the survey data can be provided to the survey sponsor.

As this research shows, cognitive interviews cannot always identify all the problems in a questionnaire. Respondent debriefing interviews identified questionnaire problems that did not surface in the cognitive interviews. In addition, respondent debriefing was able to replicate problems that occurred in the cognitive interviews. And finally, respondent debriefing was able to document the success (or in our case, the lack of success) of changes based on cognitive interviews.

Some of the evaluative findings could have been predicted, while others could not. From a questionnaire design perspective, there is still more to be learned about why some of the specific questionnaire results occurred.

As a final note, we have used a case-study approach in reporting these results, since the number of questions in the series we were testing was small. In addition, time constraints during the interviews restrict the number of debriefing questions that can be asked. More rigorous research on this topic would use a larger questionnaire, which would facilitate quantitative measure of the extent to which the results of the cognitive interviews and respondent debriefing are consistent.

## REFERENCES

Cannell, Charles, et. al., "New Techniques for Pretesting Survey Questions," final report for the National Center for Health Services Research and Health Care Technology Assessment, August, 1989.

DeMaio, Theresa J. and Jennifer M. Rothgeb,

“Cognitive Interviewing Techniques in the Lab and in the Field,” in Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research, Norbert Schwartz and Seymour Sudman, eds., Joey-Bass Publishers, San Francisco, 1996.

Esposito, James, Jennifer Rothgeb, Anne E. Polivka, Jennifer Hess Pamela Campanelli, “Methodologies for Evaluating Survey Questions: Some lessons from the Redesign of the Current Populations Survey,” paper presented at the International Conference on Social Science Methodology, Trento, Italy, June 22-26, 1992.

Hess, Jennifer and Eleanor Singer. “The Role of Respondent Debriefing Questions in Questionnaire Development,” Proceedings of the American Statistical Association (Section on Survey Research Methods), August 1995, pp.1075-1080.

Tourangeau, Roger, Rips, Lance J., Rasinski, K. The Psychology of Survey Response. Cambridge University Press, New York, NY, 2000.