

RESEARCH REPORT SERIES
(*Survey Methodology* #2004-02)

**Comparing Pretesting Methods: Cognitive Interviews,
Respondent Debriefing, and Behavior Coding**

Kristen Ann Hughes

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: September 30, 2004

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

**Comparing Pretesting Methods:
Cognitive Interviews, Respondent Debriefing, and Behavior Coding**

by

Kristen Ann Hughes
US Census Bureau

Paper prepared for presentation at the
Annual Meeting of the Federal Committee on Statistical Methodology
November 17-19, 2003
Arlington, VA

This report is released to inform interested parties of research and to encourage discussion . The views expressed the author's and not necessarily those of the U. S. Census Bureau.

I. INTRODUCTION

The quality of data collected in a survey rests on, among other things, the notion that the meaning of the question as written by the questionnaire designer is consistent with the way respondents interpret it. Pretesting attempts to ensure this by subjecting the questionnaire to some type of evaluation to assess its ability to collect the desired data. As the popularity and variety of pretesting methods grow, the question remains as to whether different methods yield consistent and/or complementary results. A commonly used method of pretesting, the cognitive interview, has traditionally been used as a front-end tool for identifying problems in a questionnaire at a relatively early stage of its development. Other “front-end” methods include expert review and forms appraisal system (which is a detailed, formal system of coding questions in questionnaires in order to detect problems). On the other hand, respondent debriefing (which involves incorporating follow-up questions in a standardized interview), and behavior coding (which is the systematic coding of the interactions between an interviewer and a respondent) have typically been used as a final stage question evaluation method during a field test or the actual production implementation of the survey itself. This paper compares three common methods of pretesting – cognitive interviewing, behavior coding, and respondent debriefing – to evaluate a series of computer-crime questions from the National Crime Victimization Survey (NCVS).

The remainder of this paper is organized as follows: Section I briefly reviews the literature to see how other research has assessed the strengths and weaknesses of these three pretesting methods, both separately and in comparison with each other; Section II lays out the objectives of the present research; Section III describes the methods used in this research; Section IV presents the results of cognitive interviews conducted in the lab, the respondent debriefing questions administered in the field, and behavior coding of interviews tape recorded in the field; and Section V offers some conclusions about the effectiveness of using each of these methods as a pretesting tool.

II. BACKGROUND

Currently, the most widely used pretesting tool is the cognitive interview (Tourangeau, Rasinski, & Rips, 2000), in which respondents are asked to report directly on the internal cognitive processes employed to answer survey questions. In this method, respondents are often asked to think aloud and describe their thought processes as they answer the survey questions. In addition, interviewers often probe about the meaning of specific terms or the intent of specific questions throughout the interview. Although there has been much success (Lessler, Tourangeau, & Salter, 1989; Willis, Schechter, & Whitaker, 1999) 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 the results are not generalizable to a larger population. In addition, the small numbers of respondents often leave the questionnaire’s less-traveled paths completely untested (DeMaio, Mathiowetz, Rothgeb, Beach, & Durant, 1993; Hess & Singer, 1995). Finally, cognitive interviewing methods, as actually practiced, vary widely, as do the styles of analyzing the data collected from such interviews (DeMaio & Landreth, 2004).

Researchers have, for several years, used debriefing questions in a standardized interview context to assess respondents' comprehension of survey questions and concepts (Esposito, Campanelli, Rothgeb, & Polivka, 1991; Cannell, Oksenberg, & Kalton, 1991). A method similar to respondent debriefing, called the "question testing" method, was first used by Belson (1981). In this method, specially trained interviewers visited respondents the day after they completed a questionnaire and asked a series of questions focusing on specific queries from the previous questionnaire. More recent methods place the debriefing questions at the end of the same interview. This particular method has gained in popularity over the years, and has also been referred to in the literature as "special probes" or "frame of reference probing" (DeMaio & Rothgeb, 1996).

DeMaio and Rothgeb (1996) argue that cognitive interviews and respondent debriefing questions can fulfill essentially the same purpose – illuminating respondents' understanding of a question's intent – at different stages of the questionnaire development process. Several studies have been conducted to assess this method's ability to evaluate survey questions (Esposito et al., 1991; Martin & Polivka, 1995; Hess and Singer, 1995; Hughes & DeMaio, 2001). These studies have demonstrated 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 questionnaire development.

Behavior coding was first used in survey research to monitor and evaluate interviewer performance (Cannell, Fowler, & Marquis, 1968; Cannell, Lawson, & Hausser, 1975), and only subsequently to investigate the question-answer process more generally (Cannell & Robinson, 1971; Morton-Williams & Sykes, 1982). The exchange between the interviewer and the respondent can provide useful information about the quality of the questionnaire as a standardized measurement instrument. Perhaps the most important reason for expanding the use of behavior coding is that it is a more systematic, objective, quantitative, and representative means of evaluating survey questions relative to other methods such as the cognitive interview or expert review (Esposito et al., 1991).

Only a few recent studies have attempted to compare pretesting methods to determine whether different methods actually produce different results (Presser and Blair, 1994; Willis et al., 1999; Rothgeb, Willis, & Forsyth, 2001). Two studies have compared the results of cognitive interviewing, expert review, and behavior coding when applied to the same questionnaire. Presser and Blair (1994) compared these three methods, along with the conventional pretest¹, to determine whether they varied in terms of reliability, validity, and cost. They found that on average, expert panels found the most problems, conventional pretests and behavior coding were

¹Presser and Blair's conventional pretest involved telephone interviews conducted by four interviewers. After the conclusion of interviewing, interviewers were extensively debriefed, and their debriefing reports were summarized by senior staff members, focusing on questionnaire problems identified.

better for identifying interviewer problems, and expert panels and behavior coding were the only methods to identify problems that affected data analysis. Behavior coding was the most reliable method and expert panels the least expensive. Presser and Blair were unable to offer conclusions concerning varying seriousness of problems.

Willis et al. (1999) conducted similar research, with somewhat different objectives. Specifically, they were interested in comparing the number of problems identified by cognitive interviewing, behavior coding and expert review, evaluating the consistency of the pretesting methods, both externally (across different techniques) and internally (within a technique across different researchers and research organizations), and assessing the types of problems identified. They found that, overall, the different pretesting techniques appeared to exhibit a “reasonable degree of consistency,” and that cognitive interviews conducted by interviewers with varying levels of experience, and using somewhat different methods, still revealed similar results.

Rothgeb et al. (2001) conducted research that also allowed for the separation of the effects of the pretesting methods – expert review, forms appraisal and cognitive interviews – from the effects of the organizations applying these methods. They found that each of the three pretesting methods contributed differently in terms of the types of problems identified, and all appeared to be especially useful in identifying problems related to question comprehension across the different organizations.

III. RESEARCH OBJECTIVES

The current research is the first study that compares the results of respondent debriefing questions with the results of cognitive interviews and behavior coding. The objective of this research is to assess the relative abilities of the three pretesting methods – cognitive interviewing, behavior coding, respondent debriefing – to identify problems in a question series, and to shed light on the types of problems each method is well-suited – or not so well suited – to identify. Building on previous studies of similar design (see in particular Presser & Blair, 1994; and Willis et al., 1999), this research aims to answer two specific questions:

1. Are these three pretesting methods equally useful in identifying problems in the questionnaire?
2. If they produce different results, what are the differences?

IV. METHODOLOGY

The research reported in this paper derives from a pretest evaluation of a series of computer crime questions proposed for inclusion in the National Crime Victimization Survey (NCVS). The NCVS is an interviewer-administered, longitudinal survey sponsored by the Bureau of Justice Statistics, in which respondents are interviewed, either in person or by telephone, every six months over a period of 3 ½ years. The computer crime series consists of seven questions (See Appendix A) asking about personal use of a computer, use of the Internet, computer-related

incidents that the respondent may have experienced, any monetary loss incurred, and whether any of the incidents were reported to an authority of some kind.

Cognitive Interviews

In the cognitive interviews conducted for this research, respondents were asked to provide a concurrent² think aloud as they decided what the question meant, retrieved relevant information from memory, and formulated their response. The interviewers also used scripted probes to elicit certain thoughts or interpretations.

Twenty cognitive interviews were conducted in the Washington, D.C. metropolitan area. Respondents ranged from eighteen to ninety-two years of age; five were more than fifty years of age. Respondents were of various races (primarily white and black), and were equally divided in gender. Respondents included college students, the unemployed, and working and retired people. All had at least some computer experience. Interviews were conducted using a questionnaire that included the computer crime questions as well as some earlier NCVS screening questions. Two trained and experienced Census Bureau methodologists conducted the interviews.

In order to accomplish some degree of quantitative analysis, I adapted a coding scheme developed by Presser and Blair (1994). Using the interviewers' summaries of their individual interviews, two independent coders³ evaluated each survey question against the following questions and coded them accordingly.

CODING SCHEME

Code	Problem
P1	Does the respondent have any difficulty understanding the meaning of the question or the meaning of particular words or concepts?
P2	Does the respondent have any difficulty remembering the question?
P3	Does the respondent have different understandings of what the question refers to?
P4	Does the respondent have any difficulty recalling, formulating or reporting an answer?

² Think-alouds can be concurrent (probe questions are asked after the respondent answers the question) or retrospective (probe questions are asked at the end of the interview). It is important to note that a cognitive interview does not replicate the normal flow of an interview, and therefore typically gives little indication of interviewer problems.

³ There was an overlap in the staff involved in cognitive interviewing and behavior coding. It is possible that this overlap might present contamination effects, which could produce bias in the results.

Coders had to decide whether each statement in an interview summary indicated that an item had posed a problem for either the respondent or the interviewer, by systematically reading each of the four questions of the coding scheme in order, stopping if the problem matched. Each problem was coded in a single category based on the assumption that respondent semantic difficulties (captured in categories P1, P2, and P3) would render moot any issue of respondent task difficulties (captured in category 4).

Behavior Coding

As noted earlier, behavior coding is the systematic assessment of the interactions between an interviewer and a respondent (Cannell et al., 1968; Cannell and Robinson, 1971). The technique is used here for the purpose of question evaluation. The complexity of coding schemes can vary widely according to a researcher's goals.

For this research, only first-level exchanges, that is, the first interviewer verbalization and the first respondent verbalization, were used to identify problematic questions and terminology. Interviewer codes were used to indicate how the interviewer asked the question. That is, did the interviewer read the question exactly as worded or with some modifications? Respondent codes showed what the respondent did when it was his/her turn to answer. Did the respondent ask the interviewer to repeat the question just read to them? If the respondent gave an answer, did he/she provide an answer that was responsive to the question? For a complete description of the codes, see Appendix B.

All telephone interviews in the final contact of the 3 ½ year panel were tape recorded for the behavior coding during the regular data collection for the 2001 NCVS during the months of September through December. Sampling was used to select 111 of the 450 taped interviews which were then coded as described above by Census Bureau research staff and telephone center interviewers, specially trained on the behavior coding technique.

Respondent Debriefing

Respondent debriefing evaluations typically use follow-up questions incorporated into the end of a standardized interview, generally as part of a field test or the production survey implementation, to gain a better understanding of how respondents interpreted the questions asked of them. Debriefing questions are used to accomplish several goals: (1) to identify words, terms or concepts that respondents do not understand, do not interpret consistently, or do not interpret as the researcher intends; (2) to identify questions that respondents cannot answer accurately; (3) to assess closed-ended response choices; (4) to assess question sensitivity; and (5) to obtain suggestions for revising questions and/or questionnaires (Belson, 1981; Hess & Singer, 1995; Esposito et al., 1991).

Unlike the cognitive interview, respondent debriefing is typically utilized at a later stage in the questionnaire design or evaluation process. An advantage over cognitive interviewing is that respondent debriefing questions offer the potential of much larger and more representative data. Respondent debriefing questions can also be used to evaluate the effectiveness of questionnaire

changes based on cognitive interviews, and to assess whether findings from cognitive interview research are replicated when the questions are fielded in a live survey setting (Hess & Singer, 1995; Willis et al., 1999.). Another major strength of respondent debriefing, compared to other pretesting techniques, is that it can provide insights into the nature of any problem encountered and, in many situations, suggestions for dealing with the problem (DeMaio & Rothgeb, 1996).

In this research, respondent debriefing questions were added to the 2001 NCVS CATI instrument containing the computer crime questions. 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. The specific debriefing questions used are presented in the “Results” section below.

V. RESULTS

In the 2001 NCVS interview, respondents were asked whether they had experienced any of a variety of computer-related incidents in the past 6 months. The objective of the computer crime questions is to collect information about computer incidents that occurred while the respondent was using a computer for personal use. In this section, I compare the results for the three research methods for three NCVS questions that were included in the proposed new series (see Appendix B)⁴. These questions ask about access to computers and specific computer-related incidents dealing with “threats of harm” and “unrequested lewd or obscene messages.”

1. “Access to Computers”

The intent of question 45d (see below) was to find out the number of computers respondents usually have access to. The sponsor’s intent is that this question stimulates respondents to think about a wide variety of computer locations, including those at home, work, school, and libraries.

45d. How many computers do you have access to for personal use or for operating a home business?

- None
- 1
- 2
- 3
- 4 or more

The results from the cognitive interviews show that ten of the twenty subjects had difficulty understanding the meaning of the question or the meaning of a particular word or

⁴The respondent debriefing was the limiting factor here; only three of the new NCVS questions were covered in the debriefing. Ideally, the entire question series would have been included in the respondent debriefings. The survey sponsors, however, thought that this would have placed an unacceptable, increased burden on the respondent.

concept (see Table 1). For this item, subjects repeatedly indicated uncertainty about the meaning of “access.” Some felt that “access to a computer” meant that they owned it, while others thought it meant any computer that they could potentially use.

According to the behavior coding results of the production survey (see Table 1), 17 of the 104 NCVS respondents did have some difficulty in understanding this question, which is less than the extent suggested by the cognitive interviews. Eight respondents asked for clarification, two provided an inadequate answer and seven provided a qualified answer. Requesting clarification is consistent with the problems observed in the cognitive interviews, and suggests that there was a problem with question comprehension. Review of the behavior coding tapes suggests that the requests for clarification were primarily due to the concept of “access.”

Table 1. Summary of Cognitive Interview and Behavior Coding Results for “Access to Computers”

Cognitive Interviews N = 20		Behavior Coding (Respondent Behavior) N = 104			
Number of Respondents with Problems	Problem Type	Adequate Answer	Request Clarification	Inadequate Answer	Qualified Answer
10	P1*	87	8	2	7

* Respondent had difficulty understanding the meaning of the question or a particular word or concept – see Section IV for coding scheme details.

One respondent debriefing question (see Table 2) also evaluated the performance of the computer access question; it asked respondents what types of computers they included when asked how many computers they had access to.

Table 2 displays the actual debriefing question and 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 2. Percent of Respondents with Access to Computers Who Included Computers at Various Locations in Their Count of Accessible Computers

Earlier, you said you had access to ... computer(s). In that number, did you include computers that you use -		
	Yes	Total N
At home	96.0 %	1611 ¹
At work	37.6	1611 ¹
At school	14.3	1611 ¹
At the library	12.2	1611 ¹
At a friend’s or relative’s home	12.0	1611 ¹
At some other place	1.7	1611 ¹

¹ Does not include 33 cases for which this information was not obtained

Respondent debriefing data show that respondents in the field reported that they used computers in a wide variety of places. Many respondents only included the computers that they use at home. At the other end of the spectrum were those respondents who included their access to computers outside of the home, such as work, school, libraries, and friends' and relatives' homes.

In terms of identifying problems, both the cognitive interviews and behavior coding suggest that respondents had trouble comprehending the question, specifically the term "access." The debriefing results, on the other hand, suggest that respondents did consider a wide range of computer locations, as intended, and offer no evidence of any problems understanding the meaning of the question.

2. "Computer-Related Incidents"

While earlier questions focused on respondents' access to computers, and where they use them, question 45f attempts to gather information about specific computer-related incidents that respondents may have experienced in the previous six months (see below). The analysis here focuses on two of the sub-questions – "threats of harm" and "lewd or obscene messages."

45f Have you experienced any of the following COMPUTER-RELATED incidents in the last 6 months –
(Read answer categories 1-6) - Mark (X) all that apply.

- 1 **Fraud in purchasing something over the Internet?**
- 2 **Computer virus attack?**
- 3 **Threats of harm or physical attack made while online or through E-mail?**
- 4 **Unrequested lewd or obscene messages, communications, or images while online or through E-mail?**
- 5 *(Only ask if box 4 is marked in Item 45c)* **Software copyright violation in connection with a home business?**
- 6 **Something else that you consider a computer-related crime? - Specify**

"Threats of Harm"

The intent of this question was to gather information about respondents who had been threatened via email or over the Internet. The sponsor was looking specifically for person-to-person threats, such as cyber-stalking.

Table 3 shows the results of both the cognitive interviews and the behavior coding. The results from the cognitive interviews did not indicate that the respondents had any major problems when probed about "threats of harm." The results of the behavior coding also suggest an absence of problems – only one person asked for clarification and one other provided an inadequate answer.

Table 3. Summary of Cognitive Interview and Behavior Coding Results for “Threats of Harm”

Cognitive Interviews N = 20		Behavior Coding (Respondent Behavior) N = 86			
Number of Respondents with Problems	Problem Type	Adequate Answer	Request Clarification	Inadequate Answer	Qualified Answer
0	N/A	84	1	1	0

The respondent debriefing results offer a very different picture. The debriefing questions attempted to gather information about the specific nature of the “threats of harm or physical attack” from respondents who reported this type of incident in order to determine if the respondent’s interpretation of the question was consistent with what the sponsor intended. Only eight respondents reported exposure to this type of incident in response to the survey. These respondents were asked the following debriefing question: *“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.”

The other two respondents did not provide an answer.

Of these answers, only the first respondent’s report (✓) indicates that he/she interpreted this question as intended. The other five (✗) incidents are not consistent with the sponsor’s definition of “threats of harm or physical attack.” Rather than being included in this category, they should have been reported instead in the “something else you consider a computer crime” category. In some cases, for example, receiving an email that says “California is a crappy place to live” should not have been reported at all. Although the number of cases is small in this finding, the results suggest that those respondents who reported “threats of harm or physical attack” did so incorrectly.

Thus, for this survey question, the respondent debriefing method found much greater evidence of problems than either the cognitive interviews or behavior coding. Respondents did not outwardly show a problem either during the cognitive interviews or behavior coding. One possible explanation is that the small number of cases for the cognitive interviews and behavior coding did not yield enough subjects to detect a behavior with such a small occurrence. Consequently, there was no evidence of a problem. However, when respondents did report “yes”

to the survey question, it is clear that they did not interpret this question as the sponsor intended. Although the implications for the quality of the data are inconsequential (only 8 positive responses out of 1644 cases)⁵, it does show that the debriefing method was useful in identifying errors.

“Unrequested Lewd or Obscene Messages”

The intent of this sub-question was to capture incidents of “unrequested lewd or obscene messages, communications, or images while online or through email.” The sponsor is interested in messages personally addressed, and sent to the recipient individually, rather than as spammed E-mails.

The results from the cognitive interviews show that 16 of the 20 respondents had a different understanding of what sorts of incidents belonged in this category than the sponsor intended and two others had difficulty understanding the meaning of the question or the meaning of a particular word or concept (see Table 4).

In the cognitive interviews, respondents generally thought the survey question referred to any kind of pornography, encountered either through spammed E-mail or at an Internet site. One respondent even went as far as to include “racist information, supremacy groups or anti-whatever religion or race. It could also be pornography somebody sent. Pictures or just words.” Respondents’ interpretations of “lewd and obscene” were much broader than intended by the sponsor. In fact, the intended meaning was not mentioned AT ALL by cognitive interview respondents, perhaps because there is virtually nothing in the wording of the question that would suggest that spammed pornography was out of scope.

A very different picture emerges from the results of the behavior coding, which indicate only very minor problems with this question, not even approaching the level of problems identified in the cognitive interviews. As Table 4 shows, 78 of the 87 respondents provided an adequate answer. Of the coded respondent errors, one asked for clarification, one provided an inadequate answer, and five provided qualified answers.

⁵One could argue that these results are, in fact, consequential, since the survey would produce a victimization rate for this type of crime that is a huge over-estimate. However, given that even the highest estimate is less than one percent, this seems to be fairly trivial.

Table 4. Summary of Cognitive Interview and Behavior Coding Results for “Lewd or Obscene Messages”

Cognitive Interviews N = 20		Behavior Coding (Respondent Behavior) N = 87			
Number of Respondents with Problems	Problem Type	Adequate Answer	Request Clarification	Inadequate Answer	Qualified Answer
2	P1*	78	1	3	5
16	P3**				

* Respondent had difficulty understanding the meaning of the question or a particular word or concept – see Section IV for coding scheme details.

** Respondent had different understanding of what the question referred to.

This study used a series of three debriefing questions (see Figures 1 and 2) to collect interpretive information from respondents who reported “unrequested lewd or obscene messages or communications or images while online or through E-mail.” Tables 6 and 7 show the frequencies for these questions.

Figure 1. Wording of Debriefing Questions Asking About “Unrequested Lewd or Obscene Messages”

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 of the 41 respondents who reported that they had experienced an “unrequested lewd or obscene message” answered these debriefing questions. Table 6 shows that the overwhelming majority of the “lewd or obscene” incident reports (86.2 percent of them) represented out-of-scope spam email. This supports the findings of the cognitive interviews that this question would not provide an accurate measure of personally-targeted pornographic email messages. Both the cognitive interview results and respondent debriefing data indicate this survey item would provide false positives and contribute to response error.

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

Yes	(25)	86.2 %
No	(2)	6.9
Don't Know	(2)	6.9
Total N	(29) ¹	100.0

¹Does not include 12 case for which this information is missing

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

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

Yes	(1)	2.5 %
No	(38)	95.0
Don't Know	(1)	2.5
Total N	(40) ¹	100.0

¹Does not include 1 case for which this information is missing

Respondents were then asked an additional debriefing question – “Did you experience any other kind of ‘unrequested lewd or obscene messages’ that were sent to you personally from another person?” If respondents answered “yes,” they were asked to describe that incident. The purpose of this question was to attempt to gather qualitative information about the kind of “lewd or obscene messages” a respondent may have received. The results of this question show that the one respondent who answered “yes” in Table 7 also misinterpreted this concept. His/her open-ended response (“Porn always in the E-mail for respondent without being asked for.”) suggests that he/she was thinking of spam email, although there may be some ambiguity in the interpretation of this response. At any rate, there were no definitive reports of the type of computer incident of interest to the survey sponsor.

Respondents who did not report that they experienced an “unrequested lewd or obscene message,” but reported some other kind of computer-related incident, were asked a variation of the same question, shown in Figure 2.

Figure 2. Wording of Debriefing Questions Asking About “Unrequested Lewd or Obscene Messages”

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 (97 of 198) said that they would include “spam E-mail” (see Table 8). This further underscores the general problems respondents had in trying to derive the sponsor’s intended meaning from the “lewd or obscene” sub-question.

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

Yes	(97)	49.0 %
No	(92)	46.5
Don’t Know	(9)	4.5
Total N	(198) ¹	100.0

¹ Does not include 102 cases for which this information was not obtained due to missing cases.

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 the results of the cognitive interviews. The extent to which this affects the survey results is unknown, since respondents may also have experienced the type of incident the sponsor intended. That is, respondents were interpreting this question to include “lewd and obscene” spam E-mail as in scope. The behavior coding simply failed to detect this problem.

VI. DISCUSSION

This research compared three methods of pretesting - cognitive interviews, respondent debriefing and behavior coding. The results varied widely (see Table 9). In the first of three questions evaluated, both the behavior coding and cognitive interviews clearly indicated that respondents had trouble understanding the meaning of the question. The respondent debriefing questions did not yield any such evidence. In the second question, the respondent debriefing interviews identified questionnaire problems that did not surface in the cognitive interviews or in the behavior coding. And in the third question, the respondent debriefings and cognitive interviews both identified a problem that did not present itself in the behavior coding.

Table 9. Summary of problems identified with each question by method

	Cognitive Interview	Behavior Coding	Respondent Debriefing
Access to Computers	P1*	P1*	
Threats of Harm			P2-A**
Lewd or Obscene Messages	P2-B***		P2-B***

Problem Type:

P1: Uncertain question meaning; Vague/ambiguous wording (“Access”)*

P2: Respondents over-reported; Answered “Yes” based on out-of-scope events

*A** – Question wording seems clear; No ready explanation for respondent behavior*

*B*** – Question wording fails to convey a key dimension of sponsor intent*

Returning to the goals of the research: First I wanted to determine if the three pretesting methods were useful in identifying problems in the questionnaire. The results show that all three methods successfully identified questionnaire problems. However, no one method consistently identified all problems in the series of computer crime questions.

Second, I wanted to see if the methods produced the same results when compared with one another. The results relevant to this objective showed that they did not always present the same results. The behavior coding only identified one question as problematic, while the respondent debriefing questions and the cognitive interviews each found problems in two questions – but not the same two questions.

In the question (“access to computers”) where respondents did not understand the meaning of the question, cognitive interviews and behavior coding identified the question as problematic. In questions (“threats of harm” and “unrequested lewd or obscene messages”) where respondents did not interpret the questions as the sponsor intended, the respondent debriefing method identified both of these as problematic, while the cognitive interviews only identified “unrequested lewd or obscene messages” as problematic. These results are consistent with other research that suggests cognitive interviews and respondent debriefing questions typically identify more comprehension-related problems than behavior coding (Presser & Blair, 1994; Willis et al., 1999).

Cognitive interviews and respondent debriefing are different than behavior coding in that they allow the interviewer to probe for specific problems, which may or may not have been identified prior to pretesting. When using these two techniques, the interviewer or researcher is typically aware of potential problems before the interview. In behavior coding, this probing or follow-up does not occur and so the respondent must independently and outwardly reveal that a problem exists, without interviewer help (U.S. Census Bureau, 2003).

In assessing these results, it is important to note that this was an observational study, not a controlled experiment. Also, it is likely that differences in the size of the respondent debriefing sample and behavior coding sample may have affected the results, as was shown in the question examining “threats of harm.” In addition, these questions were not selected as case studies to demonstrate possible ways that these methods might be related to each other, but were chosen for this research based on problems identified in an earlier round of cognitive interviews. In an ideal situation, all of the questions in this series would have been included. However, it was difficult to formulate useful debriefing questions of the subject matter questions based on hypotheses about potential response problems, since no problems surfaced in prior research.

The results of this research show the importance of using multiple methods. Although each survey pretesting method has weaknesses associated with it, employing several different methods increases the likelihood of identifying problems in a questionnaire and shedding light on the nature of those problems.

ACKNOWLEDGMENTS

I would like to thank Joyce Farmer, Gloria Prout, Susie McFadden-Resper and the staff at the Hagerstown telephone center for their help in various stages of the data coding. I would also like to express my sincerest gratitude to Jeff Moore, Jennifer Rothgeb, and Terry DeMaio for reviewing and commenting on multiple drafts of this paper.

REFERENCES

- Belson, W. (1981). *The design and understanding of survey questions*. Aldershot, England: Gower.
- Campanelli, P.C., Martin, E., & Rothgeb, J. (1991). The use of respondent and interviewer debriefing as a way to study response error in survey data. *The Statistician*, 40, 253-164.
- Cannell, C.F., Fowler, F.J., & Marquis, K. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. *Vital Health and Statistics*, 2(26). Washington, DC: U.S. Government Printing Office.
- Cannell, C.F., Lawson, S.A., & Hausser, D.L. (1975). *A technique for evaluating interviewer performance*. Ann Arbor, MI: Survey Research Center, University of Michigan.
- Cannell, C.F., Oksenberg, L., & Kalton, G. (1991). New techniques for pretesting survey questions. *Journal of Official Statistics*, 7, 349-365.
- Cannell, C.F., & Robinson, S. (1971). Analysis of individual questions. In J. Lansing, S. Withey, & A. Wolf (Eds.), *Working papers on survey research in poverty areas* (Ch. 11). Ann Arbor, MI: Institute for Social Research.
- DeMaio, T., & Landreth, A. (2004). Cognitive interviews: Do different methodologies produce different results?" In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Questionnaire development, evaluation and testing methods*. New York: Wiley Interscience.
- DeMaio, T., Mathiowetz, N., Rothgeb, J., Beach, M.E., & Durant, S. (1993). Protocol for pretesting demographic surveys at the Census Bureau. *Census Bureau Monograph*. Washington, DC: U.S. Bureau of the Census.
- DeMaio, T., & Rothgeb, J. (1996). Cognitive interviewing techniques in the lab and in the field. In N. Schwartz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass Publishers.
- Esposito, J., Campanelli, P.C., Rothgeb, J., & 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.

- Hess, J., & Singer, E. (1995). The role of respondent debriefing questions in questionnaire development. *Proceedings of the American Statistical Association (Survey Research Methods Section)*. Alexandria, VA: American Statistical Association, 1075-1080.
- Hughes, K., & DeMaio, T. (2001). *Does this question work? Evaluating cognitive interview results using respondent debriefing questions*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, Florida.
- Lessler, J., Tourangeau, R., & Salter, W. (1989). Questionnaire design in the cognitive research Laboratory: Results of an experimental prototype. *Vital and Health Statistics*, 6(1). (DHHS Pub. No. PHS 89-1076). Washington, DC: U.S. Government Printing Office.
- Martin, E., & Polivka, A.E. (1995). Diagnostics for redesigning survey questionnaires: Measuring work in the Current Population Survey. *Public Opinion Quarterly*, 59, 547-67.
- Morton-Williams, J., & Sykes, W. (1984). The use of interaction coding and follow-up interviews to investigate comprehension of survey questions. *Journal of the Market Research Society*, 26, 109-127.
- Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? In P.V. Marsden (Ed.), *Sociological Methodology*, 24, 73-104. Beverly Hills, CA: Sage.
- Rothgeb, J., Willis, G., & Forsyth, B. (2001). *Questionnaire pretesting methods: Do different techniques and different organizations produce similar results*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Canada.
- Tourangeau, R., Rasinksi, K., & Rips, L. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.
- U.S. Bureau of the Census. (2003). *Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses*, July 25. Washington, DC: U.S. Bureau of the Census.
- 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 American Statistical Association (Survey Research Methods Section)*. Alexandria, VA: American Statistical Association, 28-37.

APPENDIX A

COMPUTER CRIME QUESTIONS

INTRO: The next series of questions are about YOUR use of a computer. Please include ALL computers, laptops, or access to WebTV used at home, work, or school for PERSONAL USE or for operating a home business.

<p>45c. During the last 6 months, have YOU used a computer, laptop, or WebTV for the following purposes (Read answer categories 1-4) -</p> <p>Mark (X) all that apply.</p>	<p>100 * 1 <input type="checkbox"/> For personal use at home? 2 <input type="checkbox"/> For personal use at work? 3 <input type="checkbox"/> For personal use at school, libraries, etc? 4 <input type="checkbox"/> To operate a home business? 5 <input type="checkbox"/> None of the above - SKIP to Check Item D</p>
<p>45d. How many computers do you have access to for personal use or for operating a home business?</p>	<p>101 0 <input type="checkbox"/> None 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 or more</p>
<p>45e. Do YOU use the Internet for personal use or for operating a home business?</p>	<p>102 1 <input type="checkbox"/> Personal use 2 <input type="checkbox"/> Operating a home business 3 <input type="checkbox"/> Both 4 <input type="checkbox"/> None of the above</p>
<p>45f. Have you experienced any of the following COMPUTER-RELATED incidents in the last 6 months (Read answer categories 1-6) -</p> <p>Mark (X) all that apply.</p>	<p>103 1 <input type="checkbox"/> Fraud in purchasing something over the Internet? 2 <input type="checkbox"/> Computer virus attack? 3 <input type="checkbox"/> Threats of harm or physical attack made while online or through E-mail? 4 <input type="checkbox"/> Unrequested lewd or obscene messages, communications, or images while online or through E-mail? 5 <input type="checkbox"/> (Only ask if box 4 is marked in Item 45c) Software copyright violation in connection with a home business? 6 <input type="checkbox"/> Something else that you consider a computer-related crime? – Specify _____ _____ 7 <input type="checkbox"/> No computer-related incidents-SKIP to Check Item D</p>
<p>45g. Did you suffer any monetary loss as a result of the incident(s) you just mentioned?</p>	<p>104 1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No - SKIP to 45i</p>
<p>45h. How much money did you lose as a result of the incident(s)?</p>	<p>105 \$_____.00 Amount of loss X <input type="checkbox"/> Don't know</p>
<p>45i. Did you report the incident(s) you just mentioned to (Read answer categories 1-5) -</p> <p>Mark (X) all that apply.</p>	<p>106 * 1 <input type="checkbox"/> A law enforcement agency? 2 <input type="checkbox"/> An Internet Service provider? 3 <input type="checkbox"/> A Website administrator? 4 <input type="checkbox"/> A Systems Administrator? 5 <input type="checkbox"/> Someone else? - Specify _____ _____ 6 <input type="checkbox"/> None of the above</p>

APPENDIX B

INTERVIEWER CODES

- E/S Exact Wording or Slight Change** - The interviewer asked the question exactly as written or with only slight modifications that did not change the meaning of the question.
- M Major Change in Question Wording** - The interviewer asked the question making major changes to the question wording that altered the intended meaning of the question (such as omitting key words or phrases, paraphrasing, or combining several questions into one).
- V Verification** - The interviewer verified or repeated relevant information that the respondent had provided earlier, in place of asking a specific question.
- O Omission** - The interviewer entirely omitted (answered without reading) an applicable question.
- II Inaudible Interviewer** - The interviewer is not audible on the tape.

RESPONDENT CODES

- AA Adequate Answer** - The respondent provided an adequate answer that met the objective of the question.
- IA Inadequate Answer** - The respondent provided an answer that did not meet the objective of the question.
- QA Qualified Answer** - The respondent appeared uncertain about the accuracy of the answer he or she provided by qualifying that answer in some way.
- CL Clarification** - The respondent asked the interviewer to clarify the meaning of a particular question or concept, or asked for a repeat of the question.
- RR Respondent Repeat** - The respondent asked the interviewer to repeat the question.
- RE Refusal** - The respondent refused to answer the question or some part of the question.
- DK Don't Know** - The respondent did not know the answer to the question.
- IR Inaudible Respondent** - The respondent was not audible on the tape.