



Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet [☆]

Don A. Dillman ^{a,*}, Glenn Phelps ^b, Robert Tortora ^b, Karen Swift ^b, Julie Kohrell ^b, Jodi Berck ^b, Benjamin L. Messer ^a

^a Washington State University, Social and Economic Sciences Research Center, Department of Sociology, 133 Wilson Hall, Pullman, WA 99164-4020, USA

^b The Gallup Organization, 1001 Gallup Drive Omaha, NE 68102, USA

ARTICLE INFO

Article history:

Available online 12 May 2008

Keywords:

Response rates
Nonresponse error
Mode effects
Mixed-mode survey

ABSTRACT

The potential for improving response rates by changing from one mode of data collection to another mode and the consequences for measurement and nonresponse errors are examined. Data collection from 8999 households was done in two phases. Phase 1 data collection was conducted by telephone interview, mail, interactive voice response, or the Internet, while Phase 2 focused on nonrespondents to Phase 1, and was conducted by a different mode, either telephone or mail. Results from our study suggest that switching to a second mode is an effective means of improving response. We also find that for the satisfaction–dissatisfaction questions asked in this survey, respondents to the aural modes (telephone and IVR) are significantly more likely than are respondents to the visual modes (mail and web) to give extreme positive responses, a difference that cannot be accounted for by a tendency towards recency effects with telephone. In general, switching to a second mode of data collection was not an effective means of reducing nonresponse error based on demographics.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

One of the major survey trends of the early 21st century is the design and implementation of mixed-mode surveys in which some people prefer to respond by one type of survey mode while others prefer a different type. Several factors have encouraged the emergence of this trend. First, new survey modes such as the Internet and interactive voice response (IVR) give researchers more choices of which mode to use in addition to the traditional telephone, mail, and/or face-to-face surveys. Second, increases in cell phone use, the corresponding decrease in coverage for RDD surveys, and declining telephone response rates force researchers to consider alternative survey modes for reducing nonresponse error. Finally, previous research has shown that higher response rates can be obtained by the use of mixed-modes. For example, de Leeuw (2005) reported that use of a second or even a third mode may improve response rates and may also improve coverage.

However, mixed mode surveys could have potential drawbacks. For example, it has been learned that different survey modes often produce different answers to the same questions, such as more positive responses to scale questions on telephone than on web surveys (Dillman and Christian, 2005; Christian et al., 2008). If switching survey modes produces different measurement, then response rate gains may be offset by undesirable changes in measurement.

[☆] Financial support for this study was provided by The Gallup Organization. Additional support was provided by the Department of Community and Rural Sociology and the Social and Economic Sciences Research Center at Washington State University. The authors wish to acknowledge with thanks the assistance of many Gallup employees who contributed to the data collection and analysis of these data.

* Corresponding author.

E-mail address: dillman@wsu.edu (D.A. Dillman).

Our purpose in this paper is to simultaneously evaluate the use of a second survey (telephone or mail) mode to improve response rates achieved by an initial survey mode (web, IVR, mail or telephone) and the potential measurement differences between the first and second phases as well as across modes. This will allow us to determine the extent to which mixed-mode designs may improve response rates and whether measurement differences result. In addition, we also compare demographic differences among respondents to each mode, and between respondents and nonrespondents to determine whether respondents to a second mode of data collection vary significantly from respondents to the first mode and the population from which the samples were drawn. The issues addressed here are crucial to the design of quality sample surveys in the 21st century.

2. Theoretical background

2.1. Use of a second survey mode to improve response rates

It has long been recognized that some respondents prefer being surveyed by one survey mode, whereas others prefer a different mode. For example, Groves and Kahn (1979) reported that among the respondents to a national telephone interview, 39.4% indicated they would have preferred being surveyed by telephone, 22.7% by face-to-face interview, and 28.1% by mail.

Other studies suggest that giving respondents a choice of which mode to respond to does not necessarily improve response rates. For example, Dillman et al. (1995b) found that offering respondents the choice of whether to send back a mail questionnaire or to call in their answers to an interviewer did not improve response rates, although some respondents did prefer the telephone. Whereas a mail-only control produced a 70.6% response rate, a mail/telephone option achieved nearly the same overall response rate (69.3%), with 5.6% of the total responding by the telephone.

In contrast, it has been shown that a sequential strategy of implementing multiple contacts to ask people to respond to a particular mode and then switching to multiple contacts by another mode will improve response rates. In a national survey of college graduates, Shettle and Mooney (1999) report a 68% response rate after four contacts by mail that included a small cash incentive, 81% after an intensive telephone follow-up, and finally 88% after attempts to complete in-person interviews. The American Community Survey, a large national demographic survey conducted by the US Bureau of the Census, also implemented a sequential strategy of mixed-modes that achieved a 56.2% response rate via a mail survey, 63.5% after a telephone interview follow-up, and finally a 95.4% after face-to-face interviews (Griffin and Obenski, 2002). The use of each mode in both surveys was started after efforts for the previous phase have been concluded rather than being implemented simultaneously. Based upon these designs, the current study evaluates the sequential use of modes rather than offering respondents a choice of modes by which to respond.

The main justification for using a second mode is to increase response rates in hopes of reducing the potential for non-response error. Telephone response rates have declined significantly in recent years and coverage problems are increasing as noted by Singer (2006). Mail coverage remains a concern for general public surveys but response rates seem not to have suffered the large decline experienced for the telephone. Internet access in the US has been increasing with about 67% of American adults (18 and older) having access to the Internet from home in March 2007 (Horriagan and Smith, 2007), but this coverage is not sufficient for general public surveys. In addition, contacting email addresses for people who the survey sponsor does not have a prior established relationship with is considered an unacceptable survey practice. Response rates also tend to be lower for Internet surveys than for other modes (Cook et al., 2000; Couper, 2000). IVR surveys, which often use telephone recruitment followed by a transfer to the recorded interviewing system, remain relatively unstudied with respect to bias and response rate (Steiger and Beverly, 2008). For these reasons it is important to gain an understanding of the potential to follow-up one mode with another to improve response rates and whether doing so contributes to the reduction of nonresponse error and measurement error, both of which we investigate in this study.

The nonresponse objective of this study was to obtain a quasi-general public sample of households that can be contacted by either telephone or mail initially, and to use a normal Gallup implementation procedure for each mode. This entails switching to the other mode in order to examine the extent of response rate improvement and whether different kinds of individuals responded to each mode. A telephone contact was also made for households assigned initially to IVR and web to ask them to respond in that way. Nonrespondents to these modes were then recontacted by telephone as an alternative mode. This aspect of the analysis focuses on response rates achieved by each mode and the demographic characteristics of respondents to each.

2.2. Measurement differences across survey modes

For several decades reports of mode experiments have appeared in the survey literature (de Leeuw, 2005). Together, they suggest that significant differences often occur in the answers that people give to aural and visual surveys. For example, Dillman and Mason (1984), Tarnai and Dillman (1992) and Krysan et al. (1994) have shown in separate studies that aural respondents tend to give more positive extreme answers to opinion questions than do mail respondents. More recently, Christian et al. (2008) have shown that telephone respondents give significantly more positive answers than do web respondents for various kinds of scale questions, including 8 of 9 comparisons for fully labeled five-point scales, 11 of 13 comparisons for polar point labeled five-point scales and 3 of 4 comparisons for polar point labeled 11-point scales. A similar pattern

was noted by Christian (2007) for seven point labeled and unlabeled scales delivered in one and two-step versions, with the latter involving asking direction of attitude first followed by a question on intensity. Together these experiments suggest that respondents to telephone might be expected to express greater satisfaction with the topic being investigated (views on their long distance service) than do respondents using the other modes.

Specific efforts were made in the design of this experiment to avoid possible differences in opinions that stemmed from such things as (1) effects of interviewer presence and its possible ramifications for social desirability and acquiescence, (2) the structure of the survey questions used for each mode in the current experiment, and (3) potential effects of whether communication is visual or aural (Dillman and Christian, 2005).

2.3. Interviewer presence, social desirability, and acquiescence

Experiments have shown that respondents to surveys are more likely to offer socially desirable answers and to demonstrate acquiescence in the presence of an interviewer than in the self-administered situation (de Leeuw, 1992, 2005; Schuman and Presser, 1981). Based on that research it is expected that respondents to telephone interviews are more likely to acquiesce or express social desirability to questions than are respondents to mail questionnaires. The few available studies on IVR show somewhat mixed results. For example, Mu (1999) found that respondents to IVR were much less likely to use “10” and more likely to use “9” than were CATI respondents, perhaps because of the greater effort required when using telephone number pads to enter a “10” response. Tourangeau et al. (2002) found in two comparisons that CATI respondents gave slightly more positive responses for 11- and five-point scales than did IVR respondents. In their third comparison, and in contrast to the other comparisons, they found that IVR respondents gave slightly more positive responses on a five-point scale than did the mail respondents (Tourangeau et al., 2002). All three of these studies concerned satisfaction with a recent experience (i.e., a specific visit to a bank or fast food restaurant) leading to the conclusion by the authors that IVR produces less social desirability than does CATI.

The questions examined in the current study concern satisfaction with long distance telephone service, but were not associated with a specific experience (e.g., a recent visit) to a provider. In addition, questions were asked about whichever provider the respondent happened to have. The questions were also posed at a time when there was a great deal of movement by the public from one subscriber to another. To the extent social desirability may exist, it seems a little less likely to occur than in the study reported by Tourangeau et al. (2002). Nonetheless, their important study establishes the plausibility of different results across these three survey modes.

2.4. Question structure

It is well documented that choice of survey mode often affects how questions are structured, and whether these differences produce mode differences in respondent answers (e.g., Dillman and Christian, 2005; Dillman, *in press*). For example, use of the telephone encourages survey designers to use shorter scales, and/or scales without labeled categories. It becomes quite laborious for interviewers to read fully labeled scale choices for multiple questions to respondents, e.g., “Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?” This has encouraged company’s surveyors to use scales with only the end points labeled, e.g., “. . . where 5 means strongly agree and 1 means strongly disagree, and you may use any number from one to five.” A similar preference exists for IVR.

However, on web and mail no such pressures exist. Research has shown that fully labeled scales often obtain more positive answers than do polar point labeled scales. For example, Christian et al. (2008) found that 6 of 6 telephone and 2 of 6 web comparisons produced significantly more positive answers on fully labeled scales compared to polar point scales. Tourangeau et al. (2007) have shown that polar point labeled scales without numbers are subject to influence from visual qualities (e.g., using different colors for each end point), but not when individual scale points are labeled with numbers. Numerical labeling was used in all four modes included in the current investigation.

To avoid the possibility of differences from question structure, the current experiments also use the same scale formats across all four survey modes for measuring long distance service satisfaction. The polar point labeled format with numerical labels (1–5) favored for telephone, which had become the standard for Gallup telephone surveys, was adopted for all four modes, using identical wording.

2.5. Visual (web and mail) vs. aural (telephone and IVR) communication

Mode differences in respondent answers may also be accounted for by aural vs. visual communication and whether the question stimulus is controlled by the interviewer or the respondent. Research using several nominal categories for response choices has suggested that visual survey modes in which the stimulus is controlled by the respondent sometimes produce a primacy effect in which respondents are more likely to choose items listed first in a list of answer categories (Krosnick and Alwin, 1987). Primacy is thought to occur because in a visual presentation the items listed first are subjected to deeper cognitive processing, thus establishing a standard of comparison that guides interpretation of later items (Krosnick and Alwin, 1987).

In addition, Krosnick and Alwin (1987) argue that when items are delivered aurally to respondents, in which the stimulus is controlled by the interviewer, there is not enough time for the respondent to place each answer choice into long-term

memory before the next one is read (Krosnick and Alwin, 1987). As a consequence, respondents could be more likely to choose the last categories on a list. This tendency is described as a recency effect. However, in later work, Krosnick proposed that scale questions which are read in a sequential order may produce a primacy effect under both aural and visual conditions because people probably consider each response alternative in the order in which they are read (1999, p. 552).

The same conditions as those that could produce a recency effect in telephone (aural communication and control of pacing by interviewer), may also produce similar results in the IVR mode but the situation is still unclear because the touch-tone keypad provides a visual representation of a scale, although not in the linear format that appears in mail surveys. For IVR respondents, in addition to hearing the word labels from the voice recording, numbers associated with those labels are also heard so they may be directed towards the labeled buttons more than the unlabeled ones. This tendency is supported in research by Srinivasan and Hanway (1999), who found for 11 items using five-point scales with labeled endpoints that IVR respondents were significantly more likely (mean difference = six percentage points) than mail respondents to choose strongly agree. They also compared six questions on an IVR survey, labeled only on the end points, with the same six questions on a mail questionnaire that were fully labeled. The differences were in the same direction but they were larger (mean difference = 17 percentage points), with more IVR respondents than mail respondents choosing strongly agree. These data suggest that the visual presence of labeled categories on a paper questionnaire pull respondents even more strongly to the intermediate categories than do unlabeled categories.

Considerable research has reported both recency and primacy effects (Schuman and Presser, 1981; Dillman et al., 1996) but is inconclusive. A series of 82 experiments placed in many different surveys did not reveal a consistent pattern of effects (Dillman et al., 1995a). Similarly, Moore (1998) has reported a mixture of primacy and recency effects for scale questions, although the former were more prevalent in Gallup Poll opinion questions. In addition, the earlier mentioned experiments by Tarnai and Dillman (1992) and Krysan et al. (1994) show similar extremeness in the interview modes although the scales were run in opposite directions. Sudman et al. (1996) after a detailed review of such order effects concluded, "...response order effects may go in different directions... and may cancel one another in heterogeneous samples" (p. 169). In light of the conflicting results and difficulty of knowing which types of questions result in a primacy or recency effect, it was deemed important for the current study to control for such potential effects. Nonetheless, the uncertainty on whether primacy or recency effects might be expected led to the decision to include a partial control into the experimental design, so that scale questions were presented in reversed order to a subsample of telephone respondents.

Another aspect of visual vs. aural communication effects relevant to this study is how alternative visual layouts may impact respondent answers within visual modes. Considerable research has shown that different visual layouts of questions may produce quite different answers from respondents (Christian and Dillman, 2004; Tourangeau et al., 2004). These effects are in general predicted by principles drawn from Gestalt psychology (Jenkins and Dillman, 1997) and the vision sciences (Ware, 2004). Two features of this work are of particular relevance to this study. One is that different visual layouts in mail and web surveys produce similar results for a variety of question structures (see Dillman, 2007, pp. 447–497 for a summary of those effects). In addition, it has been shown in controlled experiments for seven different formats of scale questions, including the polar point labeled structures used here, that telephone respondents provide slightly more positive answers than do web respondents (Christian et al., 2008; Christian, 2007). Consistent with these research findings, the web and mail questions asked in the current experiment used the same visual layouts (see Fig. 1).

In these several ways, the examination of measurement differences was methodologically constrained to avoid confounding the many possible factors—question structure differences, questions subject to social desirability, question formats subject to acquiescence, and visual format differences—that could jointly influence respondent answers, thus making any differences difficult to interpret. Our measurement focus in this paper is therefore limited to primacy or recency considerations for a polar-pointed labeled format.

In sum, it was expected that answers to the critical measurement questions in this survey on long distance survey satisfaction might be more positive among telephone respondents and, to a lesser extent, IVR respondents than for other modes, but that primacy or recency was unlikely to account for those differences.

3. Study procedures

Response rate effects are examined for four different initial implementation strategies: a telephone interview, a mail questionnaire, an attempt by telephone to recruit respondents to answer a self-administered IVR survey, and an attempt by telephone to recruit respondents to complete a web survey. After a pause of one month in the data collection effort, nonrespondents to the telephone survey were asked to complete a mail questionnaire, while nonrespondents to the other modes (mail, web and IVR) were contacted by telephone and asked to complete a telephone interview.

In order to evaluate response rate effects across survey modes, it was necessary to obtain a common sample frame that would allow people to be contacted either by mail or telephone. This required that both telephone numbers and addresses be available. We also wished to have the characteristics of the sample frame approach those of a general public, as opposed to a membership or employee population. Finally, we wanted to obtain demographic characteristics for all members of the population frame so that a nonresponse error determination could be made by comparing respondent attributes to those for nonrespondents.

These objectives were accomplished by purchasing a list of 8999 names from a private company. This list consisted of individuals with a known long distance provider who had sent in cards to register warranties for a wide variety of consumer

Mail

- 2** Overall, how satisfied are you with your long distance company?
- 1 Not at all satisfied
- 2
- 3
- 4
- 5 Extremely satisfied

Telephone, Form A

(Interviewer Read:) Overall, how satisfied are you with your long distance company? Please use a one-to-five scale, where “5” means extremely satisfied, and “1” means not at all satisfied. You may use any number from one to five.

Telephone, Form B

(Interviewer Read:) Overall, how satisfied are you with your long distance company? Please use a one-to-five scale, where “1” means not at all satisfied, and “5” means extremely satisfied. You may use any number from one to five.

Interactive Voice Response

(Recorded voice reads:) Overall, how satisfied are you with your long distance company? Please use a one-to-five scale. If you are extremely satisfied, press “5.” If you are not at all satisfied, press “1.” You may press any number from one to five. To repeat a question, press the star key.

Web

- 2** Overall, how satisfied are you with your long distance company?
- 1 Not at all satisfied
- 2
- 3
- 4
- 5 Extremely satisfied

Fig. 1. Example of question formats for asking overall satisfaction with long distance service.

products or filled out surveys about their shopping behavior or product preferences. We required the name, telephone number, mailing address, and six pieces of demographic information, which included gender, income, whether children were present, age, education, and number in household. After the study began, we learned that the sample had been limited to individuals who reported children in the household and a household income of at least \$60,000. Although it was not our objective to have the sample limited in this way, the availability of a common sample frame that could be used for both mail and telephone access led to our decision to continue the study. Despite these limitations, we concluded that the response rate and mode differences of interest could be reasonably investigated. A questionnaire was developed that consisted of 18 questions, including 12 questions about the household's long distance service and 6 demographic questions.

The topic of long distance service was selected because of our specific interest in that issue on which other recent surveys had been conducted. In addition, the likelihood that all or nearly all of the sampled individuals would have long distance service at the time the survey was conducted meant that virtually every household in the sample frame should be able to

respond to the questions. Also, as mentioned previously, it was a topic for which we did not expect social desirability or acquiescence effects to influence answers. Seven of the questions concerned opinions about that long distance service, five of which were only labeled on the polar points, and two of which were fully labeled. The demographic questions included gender, highest level of education, age, number in household, presence of children under 18, and income.

Names on the sample frame list were randomly divided into four groups for completion of mail, telephone, IVR, and web modes. The telephone sample was then further divided into two subsamples (Treatments 2 and 3), and the five groups were contacted as follows:

3.1. Treatment 1: Phase 1 mail questionnaire, Phase 2 telephone interview

This random subsample of 2000 names received a prenotice in week one, a questionnaire with personalized letter and \$2.00 bill in week two, and a thank-you/reminder postcard in week three. The letter accompanying the questionnaire was personalized with the name and address of the recipient and printed on Gallup stationery. The letter requested "... the person in your household who is either responsible or shares responsibility for making decisions about your long distance service spending just a few minutes to complete and return this brief questionnaire." The questionnaire was printed on an 11 × 17-in. sheet of paper that was folded to a conventional 8 1/2" × 11" size. Questions were printed in black ink on blue background fields with white boxes for marking answers (see Fig. 1). A title was on the outside front page, along with a brief description of the purpose and contact information. All 18 questions were printed on the inside two pages, with two columns of questions on each page. Nothing was printed on the outside back page. These procedures emulated both the questionnaire construction and implementation procedures described by Dillman (2007).

The preletter was mailed on November 15, 1999, and the questionnaire mailing occurred on November 22, 1999. Reminder postcards to respondents who had not turned in their questionnaire were sent out on November 29, 1999. Nonrespondents to the mail questionnaire, including individuals whose addresses turned out to be incorrect, were assigned to the telephone interview process of Phase 2 that began on February 9, 2000. The telephone procedures used are those described under Phase 1 of Treatments 2 and 3.

3.2. Treatments 2 and 3: Phase 1 telephone interview, Phase 2 mail questionnaire

A total of 2999 random names designated for interviewing via telephone were randomly assigned to two different forms of the survey. The direction of the scales was reversed between these two treatments so that it could be determined whether a telephone recency effect existed on the seven opinion questions. For example, in Treatment 2 (Form A), overall satisfaction was measured by interviewers reading from the most positive rating labels to the most negative rating labels, "where '5' means extremely satisfied, and '1' means not at all satisfied. ..." In Treatment 3 (Form B) respondents heard the question with the most negative rating label first, as "where '1' means not at all satisfied and '5' means extremely satisfied. ..." (see Fig. 1).

Form A or Form B was randomly assigned to each respondent at the beginning of the interview in a way that resulted in each interviewer administering both forms. For example, if an interviewer completed four interviews during one session, two of the interviews would have been randomly assigned to Form A, while the other two would have been assigned to Form B. However, if the interviewer did an odd number of surveys during one session, the distribution of Form A and Form B would not be equal. It is for this reason that the number of completions varied slightly (651 vs. 667) in Phase 1 of the data collection.

Attempts to interview by telephone began on November 16, 1999. These calls were made by trained Gallup interviewers. When someone answered, the interviewer identified himself by name as being from The Gallup Organization, and continued, "we are conducting a study of people to find out what they think about the service they receive from their long distance telephone company. The interview is brief and we are not selling anything." The interviewer then asked, "Are you the person in your household who is responsible or shares in the responsibility for making the decisions regarding your long distance telephone service?" If that individual was not available, at least four callbacks were made to the selected respondent at different times of the day and different days of the week to complete the interview. Calls were made from November 16, 1999 to January 9, 2000.

In February, all households that had not completed a telephone interview (including nonworking numbers and some refusals) were sent the mail questionnaire described under Treatment 1. Included in the mail survey packet was a cover letter that acknowledged the attempt to contact them previously, the survey questionnaire, and a \$2.00 bill. A follow-up postcard was sent to potential respondents who did not return the questionnaire.

3.3. Treatment 4: Phase 1 IVR recruited by telephone survey, Phase 2 telephone interview

Another 2000 randomly selected names were contacted by telephone in the same manner as that used for the telephone interviews. After being asked what company provided their long distance telephone service, the first question in the interview, these individuals were then told, "In order to obtain your views in the most confidential and efficient way, the rest of the survey is completed using our automated system, where you enter your answers using the numbers on your phone. It will take approximately five minutes." Respondents were then asked to stay on the line and that it would take about 10 s to transfer them. Those who stayed on the line then heard a prerecorded welcome that began, "During the next few minutes,

we will ask you to rate the overall service quality provided by your long distance telephone company. Your responses will be kept confidential and the remainder of the survey takes about five minutes. Please press '1' on your touch-tone phone to begin." The remainder of the IVR interview was worded in an identical fashion to the other modes, except for modification of instructions to people on pressing the touch-tone numbers, i.e., "If you are extremely satisfied, press '5'. If you are not at all satisfied, press '1'. You may use any number from one to five" (see Fig. 1). If respondents did not press a number, or pressed a number outside the acceptable range, a prerecorded voice gave directions on how to press the right number. During Phase 2 of the study, all nonrespondents to the IVR system were reassigned to a telephone interview where the procedures described for Treatments 2 and 3 were followed.

3.4. Treatment 5: Phase 1 web recruited by telephone survey, Phase 2 telephone interview

The 2000 households assigned to this treatment were approached in the same way as respondents to Treatment 4, except that after getting the person on the line who was responsible for long distance telephone service decisions, he or she was asked whether they had access to the World Wide Web from their home, work, school or any other place, or if they had a subscription to certain online services. A total of 728 households (36%) reported accessing the web within the previous 30 days, or subscription access to American Online (AOL), CompuServe, or Prodigy. These respondents were asked, "Because of your Internet usage, I would like you to participate in an extremely important study about long distance service. The study will be conducted over the Internet. Along with some brief instructions, we will be sending you \$2.00 cash as a token of our appreciation for completing the Internet study." Follow-up letters were mailed to individuals who agreed to participate, along with access information and \$2.00. Towards the end of the field period, an email reminder was sent to everyone who had not yet completed the survey. The web questionnaire was constructed to emulate visual aspects of the paper questionnaire (see Fig. 1). Thus, a similarly colored background, and placement of questions on the page were designed to ensure a very similar stimulus regardless of mail or web data collection. Up to five telephone calls were made to recruit for the web survey. The telephone calls and subsequent mailings to eligible respondents occurred between November 16 and November 29, 1999. In Phase 2 an attempt was made to call all nonrespondents, including those who did not have access to the web, and complete an interview by telephone.

4. Results

4.1. Response rates

Response rates for each phase of the data collection are reported by treatment in Table 1. During Phase 1 it can be seen that rates varied greatly, from lows of 13% for the web, to 28% for IVR, 44% by telephone and 75% for mail. Such wide variations were not unexpected. Many of the potential respondents contacted by phone in the web survey effort did not have computers and/or Internet access, and some who did have access were unwilling to participate. Completion of the IVR questionnaire required respondents to wait for a connection to the automatic voice system (the recruiting telephone interviewer told each respondent the wait was typically about 10 s). In addition, the transfer rate of the current study (i.e., 69%) is consistent with similar surveys conducted by The Gallup Organization, in which about two-thirds of those successfully recruited by telephone complete an IVR interview.

The response for those assigned to the mail treatment group was 31 percentage points higher than the response rate for the telephone interview treatments. Part of this difference is undoubtedly due to the decision to conduct each method as it would normally be implemented alone, thus including a \$2.00 cash incentive with the mail request. Past research has shown that response rates to general public surveys can be increased from 15 to 20 percentage points by the inclusion of such an incentive (Lesser et al., 1999). In addition, research on the use of incentives with telephones has suggested that incentives contingent upon completion of the interview, which would have necessarily been the

Table 1
Response rates^a by treatment groups

Groups	Original sample size	Phase 1		Phase 2		Total %	Improvement in response rate from Phase 1 to Phase 2 %		
			<i>n</i>	%	<i>n</i> ^b			%	
Treatment 1	2000	Mail	1499	75.0	Phone	157	31.7	82.8	7.8
Treatment 2 (Form A) ^c	1500	Phone	651	43.4	Mail	1094	66.3	80.4	37
Treatment 3 (Form B) ^c	1499	Phone	667	44.4					36
Treatment 4	2000	IVR	569	28.5	Phone	438	35.9	50.4	21.9
Treatment 5	2000	Web	253	12.7	Phone	700	44.9	47.7	35

^a Response rate = number of completed/number in treatment.

^b Includes noncontacts and refusals to Phase 1.

^c Treatment groups 2 and 3 cannot meaningfully be separated in Phase 2 because of assignment methods described in text.

case with the telephone procedures used here, may be less effective than unconditional incentives sent ahead of time (Singer, 2002).

Response rates for all five treatments were increased substantially by switching to a second mode of data collection once Phase 1 had been completed (see Table 1). Phase 2 response rates to the attempted telephone interviews ranged from 32% for the Phase 1 mail nonrespondents to 45% for the web nonrespondents. The use of telephone performed similarly in Phase 1 and Phase 2 positions, except for Treatment 1 to which 75% had already responded. The mail response in Phase 2 was 66% (Treatments 2 and 3), only nine percentage points lower than it performed when used as an initial contact, even though some of these letters went to households that had refused the telephone interview.

These data show that the combination of mail and telephone perform similarly, 80–83% overall response, regardless of which method is used first and which comes second, thus confirming strongly the value of using one mode as a means of improving response to the other mode. The overall response rates to the IVR and web treatment groups are much lower, 50% and 48%, respectively, and responses were not much improved over that which could have been obtained by only a telephone interview. However, it is important to remember that 35% of the telephone respondents in Treatment 5 who reported having web access completed the survey on the web. Furthermore, a substantial portion of these individuals had already been contacted by telephone, and being called in Phase 2 with a request to be interviewed by telephone might have been viewed as a recontact by the same mode. Thus, the lower response rates were expected.

4.2. Category order effects for scale questions asked by telephone

The seven opinion questions that asked for scale responses were presented positive to negative in Treatment 2 (Form A) and negative to positive in Treatment 3 (Form B). Our purpose for comparing these two groups was to determine whether any differences occurred, particularly those of recency that might be expected based upon past research.

If a recency effect occurs we would expect that those responding in Treatment 2 “Not at all satisfied” (i.e., last category presented to the respondent) would be significantly different in comparison to the same response in Treatment 3. Likewise, those responding in Treatment 3 “Extremely satisfied” (i.e. last category presented to the respondent) would be larger than in the other treatment. This is not the case. A difference of means test shows that none of the differences are statistically significant and there is no consistent trend in either direction. All χ^2 tests for differences were also insignificant. Because of the lack of differences, or ever slight trends, these results are not shown in the table. Also, because of the lack of differences we have combined Treatments 2 and 3 together for all other analyses reported in this paper.

4.3. Nonresponse error for demographic variables

Two opportunities exist for evaluating nonresponse error on demographic variables. First, we can compare demographics reported by Phase 1 respondents with those obtained for Phase 2 respondents to determine whether the second phase produced respondents with significantly different demographic characteristics. Second, we can compare the demographics of responding households with nonresponding households, using those demographics reported to us by the company that provided the sample list. The latter allows a more direct test to be made for nonresponse error.

As shown in Table 2, Phase 1 and Phase 2 respondents to Treatment 1 did not differ significantly on any of the nine variables for which respondents reported information. Thus, introducing telephone as a follow-up to mail can produce the same types of respondents as did the Phase 1 mail out. However, it should be noted that the initial response rate of 75% was already quite high, and use of the telephone increased it to only 83%.

Comparisons between Phase 1 (telephone) and Phase 2 (mail) respondents assigned to Treatments 2 and 3 reveal differences that are significant for only two variables—education and number in household. Contrary to our expectations, use of the Phase 2 mail strategy tended to bring in respondents with somewhat less education. However, consistent with our expectations, the Phase 2 mail mode was more likely than the Phase 1 telephone mode to bring in respondents from smaller households. In Treatment 4 there were three significant differences between Phase 1 IVR respondents and Phase 2 telephone respondents. As indicated in Table 2 the telephone interview brought in respondents with somewhat less education and smaller households with higher incomes. Only one significant difference existed for respondents to Phase 1 (web) and Phase 2 (telephone) of Treatment 5. Specifically, males were much more likely than females to respond by the web while the opposite was true for phone. With a few exceptions, our overall impression is that introducing a Phase 2 data collection strategy did relatively little to influence the characteristics of people who responded to the survey, and perhaps (we cannot be certain of the direction of the effect) to reduce nonresponse error.

Because of the availability of information for both respondents and nonrespondents, examination of demographic variables provided with the sample frame for all sample units (a criterion for inclusion in the sample frame was their availability) provides a stronger test of whether nonresponse error is reduced through our mixed-mode data collection strategy (Table 3). Four demographic characteristics—gender, education, age and income—were available for all sample units. It is important to understand precisely how these demographics differ from those in Table 2. In Table 3, we are *not* comparing answers given by respondents to our survey; instead, we are comparing the demographics obtained by the provider of the sample list. The characteristics shown in Table 3 (i.e., age and gender) may not correspond to what was given to us by the respondent to our survey, because we asked for the person who made long distance decisions for the household to respond. That person may or may not have been the person who provided the warranty information that led to inclusion of that name in the sample

Table 2

Comparison of self-reported demographics for respondents to Phase 1 and Phase 2 for each treatment

Variable	Treatment 1			Treatments 2 and 3			Treatment 4			Treatment 5		
	Phase 1 Mail (n = 1499)	Phase 2 Phone (n = 157)	Total	Phase 1 Phone (n = 1318)	Phase 2 Mail (n = 1094)	Total	Phase 1 IVR (n = 569)	Phase 2 Phone n = 438	Total	Phase 1 Web (n = 253)	Phase 2 Phone (n = 700)	Total
<i>Gender</i>												
Male	46.6	46.5	46.6	47.1	49.6	48.3	45.7	49.3	47.3	58.5	44.3	48.1
Female	53.4	53.5	53.4	52.9	50.4	51.7	54.3	50.7	52.7	41.5	55.7	51.9
	$\chi^2 (1)$.001			1.515			1.302			15.037**
<i>Education</i>												
High school graduate or less	8.6	7.8	8.5	9.0	9.1	9.0	7.3	10.5	8.7	6.7	9.4	8.7
Some college/trade	25.4	24.0	25.3	20.7	27.5	23.8	25.1	19.2	22.5	20.6	21.9	21.6
Coll. grad.	36.4	42.2	36.9	40.1	37.9	39.1	34.7	41.6	37.7	37.9	42.0	40.9
Post grad. work or more	29.7	26.0	29.3	30.3	25.5	28.1	32.9	28.6	31.0	34.8	26.7	28.9
	$\chi^2 (3)$		2.116			16.971**			11.325**			6.619
<i>Age</i>												
18–34	13.2	16.8	13.5	13.4	13.2	13.3	13.0	13.8	13.3	10.3	12.6	12.0
35–44	43.6	44.5	43.7	45.3	45.1	45.2	43.3	46.2	44.6	48.0	47.4	47.6
45–54	37.5	32.9	37.1	35.5	35.6	35.5	37.9	33.1	35.7	38.9	33.6	35.0
55+	5.8	5.8	5.8	5.8	6.1	5.9	5.8	6.9	6.3	2.8	6.4	5.4
Mean	43.21	42.61	43.15	42.90	43.22	43.05	43.17	42.97	43.08	42.86	42.99	42.96
	$\chi^2 (3)$		2.161			.158			2.539			6.784
<i>Number in household</i>												
1–2 people	5.7	6.5	5.8	5.6	7.8	6.6	4.4	6.4	5.3	6.3	5.1	5.4
3	24.6	19.5	24.1	24.9	28.6	26.6	26.5	23.6	25.3	27.3	25.3	25.8
4	43.3	47.4	43.7	46.1	41.6	44.1	42.2	44.5	43.2	41.1	45.6	44.4
5	19.3	19.5	19.3	15.5	16.7	16.0	14.6	19.5	16.7	19.0	16.6	17.2
6 or more	7.1	7.1	7.1	7.9	5.3	6.7	12.3	6.0	9.5	6.3	7.5	7.2
Mean ^a	3.97	3.99	3.97	3.94	3.82	3.89	4.03	3.94	3.99	3.92	3.96	3.95
	$\chi^2 (4)$		2.213			16.714**			16.733**			2.660
<i>Children in household (% yes)</i>												
	88.1	90.9	88.4	89.0	87.8	88.5	88.5	85.7	87.3	87.0	89.3	88.7
	$\chi^2 (1)$		1.068			.921			1.709			1.004
<i>Monthly long distance bill</i>												
Less than \$10/month	11.5	15.7	11.9	13.8	12.1	13.0	16.4	14.8	15.7	14.2	15.7	15.3
\$10 less than \$20/month	18.3	15.7	18.0	18.3	19.0	18.6	20.0	21.2	20.5	19.4	14.7	16.0
\$20 to less than \$50	37.2	37.9	37.3	36.6	34.8	35.8	35.3	34.9	35.1	34.0	37.2	36.3
\$50 to less than \$75	16.7	15.0	16.5	16.2	17.4	16.7	16.6	15.0	15.9	15.0	15.7	15.5
\$75 to less than \$100	8.5	7.8	8.5	7.2	7.9	7.5	7.1	6.2	6.7	9.5	7.6	8.1
\$100 or more	7.7	7.8	7.7	7.9	8.9	8.4	4.6	7.9	6.0	7.9	9.1	8.8
	$\chi^2 (5)$		2.846			3.564			5.475			4.444
<i>Income</i>												
Under \$60K	8.5	10.8	8.7	11.0	9.5	10.3	13.7	9.9	12.1	7.4	9.1	8.7
\$60K to less than \$75K	22.4	16.5	21.9	23.2	19.6	21.5	26.6	21.0	24.3	13.6	19.1	17.6
\$75K to less than \$100K	36.6	30.9	36.1	36.1	37.5	36.8	32.1	35.7	33.6	45.0	39.1	40.8
\$100K or more	32.5	41.7	33.3	29.7	33.3	31.4	27.7	33.4	30.1	33.9	32.7	33.0
	$\chi^2 (3)$		7.097			7.188			9.092*			5.097

(continued on next page)

Table 2 (continued)

Variable	Treatment 1			Treatments 2 and 3			Treatment 4			Treatment 5		
	Phase 1 Mail (n = 1499)	Phase 2 Phone (n = 157)	Total	Phase 1 Phone (n = 1318)	Phase 2 Mail (n = 1094)	Total	Phase 1 IVR (n = 569)	Phase 2 Phone n = 438	Total	Phase 1 Web (n = 253)	Phase 2 Phone (n = 700)	Total
<i>Interruption of phone service (% yes)</i>												
	3.6	5.2	3.8	4.0	5.0	4.5	7.4	4.8	6.3	3.6	3.9	3.8
$\chi^2 (1)$.97			1.43			2.83			.056
<i>Without phone service^b</i>												
Less than 1 week	70.6	85.7	72.4	84.6	79.6	82.1	86.1	85.7	86.0	100.0	85.2	88.9
One week or more	29.4	14.3	27.6	15.4	20.4	17.9	13.9	14.3	14.0		14.8	11.1
$\chi^2 (1)$.705			.448			.002			1.50

^a Mean value based upon recorded number in household.

^b Based upon respondents with phone interruption (small sample size).

* $p < .05$.

** $p < .01$.

Table 3
Sample frame demographics by response status

Variable	Treatment 1			Treatments 2 and 3			Treatment 4			Treatment 5										
	Sample frame n = 2000 Phase 1 Mail n = 1499 Phase 2 Phone n = 157 Phase 1 and 2 n = 1656 Phase 1 and 2 n = 344	Nonrespondents n = 587 Phase 1 Mail n = 1318 Phase 2 Mail n = 1094 Phase 1 and 2 n = 587	Phase 1 and 2 Phone n = 2999 Phase 1 Phone n = 1318 Phase 2 Mail n = 1094 Phase 1 and 2 n = 587	Sample frame n = 2000 Phase 1 IVR n = 569 Phase 2 Phone n = 438 Phase 1 and 2 n = 1007 Phase 1 and 2 n = 993	Nonrespondents n = 2000 Phase 1 Web n = 253 Phase 2 Phone n = 700 Phase 1 and 2 n = 953 Phase 1 and 2 n = 1047	Phase 1 and 2 Phone n = 2000 Phase 1 Web n = 253 Phase 2 Phone n = 700 Phase 1 and 2 n = 953 Phase 1 and 2 n = 1047														
Gender																				
Male	43.0	44.8	42.0	44.5	35.8	45.0	46.1	47.2	46.6	38.5	43.5	46.2	47.3	46.7	40.3	45.0	52.2	44.3	46.4	43.6
Female	57.0	55.2	58.0	55.5	64.2	55.0	53.9	52.8	53.4	61.5	56.5	53.8	52.7	53.3	59.7	55.1	47.8	55.7	53.6	56.4
Phase 1 vs. Phase 2	$\chi^2 (1)$.427		.258							.107					4.650*				
Respondents vs. nonrespondents	$\chi^2 (1)$	8.895**		12.513**							8.310**					1.504				
Education																				
High school or less	17.7	16.2	15.8	16.2	24.9	18.3	15.1	17.8	16.3	26.1	18.7	17.3	17.3	17.3	20.1	16.7	11.3	17.1	15.6	17.8
Some college/technical	21.8	21.7	18.0	21.4	24.0	20.5	19.5	21.2	20.3	21.1	21.9	21.3	19.0	20.3	23.6	20.4	15.6	19.3	18.3	22.3
Coll. deg.	39.7	40.4	41.0	40.4	36.0	41.7	43.8	42.8	43.3	35.4	39.0	38.8	39.6	39.1	38.8	41.7	45.9	41.7	42.8	40.7
Grad deg./post grad.	20.8	21.7	25.2	22.0	15.1	19.5	21.6	18.3	20.1	17.4	20.4	22.6	24.1	23.2	17.5	21.2	27.3	21.9	23.3	19.2
Phase 1 vs. Phase 2	$\chi^2 (3)$	1.564		.611							.803					7.797*				
Respondents vs. nonrespondents	$\chi^2 (3)$	19.288**		30.712**							11.402**					8.978*				
Age																				
18–34	13.3	12.2	17.8	12.7	16.0	13.7	12.0	14.3	13.0	16.7	12.8	12.5	13.0	12.7	12.9	15.3	10.7	14.3	13.3	17.0
34–44	47.4	46.2	48.4	46.4	52.0	48.4	47.6	47.6	47.6	51.4	48.4	46.6	50.9	48.5	48.2	47.7	50.6	44.7	46.3	48.9
45–54	33.2	35.0	26.8	34.2	27.9	31.7	35.4	31.4	33.5	24.2	32.2	34.1	29.2	32.0	32.4	30.5	33.2	33.1	33.2	28.1
55+	6.2	6.6	7.0	6.6	4.1	6.2	5.0	6.8	5.8	7.7	6.7	6.9	6.8	6.9	6.4	6.6	5.5	7.9	7.2	6.0
Phase 1 vs. Phase 2	$\chi^2 (3)$	6.559		8.405*							2.85					4.595				
Respondents vs. nonrespondents	$\chi^2 (3)$	10.669*		21.731**							.174					10.489*				
Income																				
\$60K < \$75K	25.5	25.5	26.1	25.5	25.3	27.5	29.1	25.9	27.6	27.3	28.4	32.9	26.9	30.3	26.4	25.9	20.9	27.3	25.6	26.1
\$75K < \$100K	42.3	43.2	35.7	42.5	41.0	42.6	40.8	43.9	42.2	44.1	41.6	40.4	45.2	42.5	40.7	43.9	48.6	43.3	44.7	43.1
\$100K+	32.3	31.3	38.2	31.9	33.7	29.9	30.1	30.3	30.2	28.6	30.1	26.7	27.9	27.2	32.9	30.3	30.4	29.4	29.7	30.9
Phase 1 vs. Phase 2	$\chi^2 (2)$	4.069		3.531							4.318					4.160				
Respondents vs. nonrespondents	$\chi^2 (2)$	4.42		.815							8.530*					.563				

* $p < .05$.

** $p < .01$.

frame. Comparisons in this table tell us whether the percent of households in the sample frame for which a “female” reported the warranty information is the same or different among responding and nonresponding households. The sample frame demographics are inherently devoid of any possible survey mode effects on how survey mode may have influenced answers to demographic questions. Although past research has not reported such differences and we could find no theoretical grounds for expecting any mode differences, the check seemed important to have. Two statistical comparisons are shown in Table 3: Phase 1 vs. Phase 2, and those responding to the survey vs. nonresponding households. The first comparison (Phase 1 vs. Phase 2) provides another indicator, in addition to Table 2, of whether switching to a second mode changes the characteristics of responding households. The second comparison provides a measure of whether the households that responded were different from those that did not respond.

Table 3 shows that, in general, the responding household was more likely to have male than female reported for the sample frame characteristics, and the difference was significant for all but Treatment 5. It can also be observed that for Treatment 5, where web was the initial data collection mode, respondents were significantly more likely to come from households with a male sample characteristic (52.2% in Phase 1 web, compared to only 44.3% male for the telephone follow-up). A similar but much smaller compensation occurred in Treatment 1 between the Phase 1 data collection by mail and Phase 2 by telephone (44.8% and 42.0% male respondents, respectively).

With the exception of Treatment 5, where web survey respondents were more highly educated than the telephone follow-up respondents, mode switching did not bring in respondents with significantly different reported education levels. However, for all four treatments shown in Table 3, nonresponding households differed from those that responded. In each case, nonresponding households had sample frame characteristics of significantly lower education than did the responding households. None of the modes appeared to be successful in obtaining responses commensurate with the lower education levels reported for the sample frame.

There were no significant differences between Phase 1 and Phase 2 respondents across all treatment groups for income, the one demographic variable available which was a true household measure rather than a characteristic of the individual. In only one case, those assigned to Treatment 4, did those responding to the survey differ significantly from nonrespondents, with sample frame incomes for the latter group being somewhat higher.

4.4. Mode effects across all treatment groups

Examination for potential mode effects involves two steps. First, we determine whether any differences in means and distributions are more likely the result of mode or type of respondent to Phase 1. This is done by comparing Phase 1 responses from one treatment to Phase 2 responses by the same mode in other treatment groups (Tables 4 and 5). If the mail questionnaire respondents to Phase 2 of Treatments 2 and 3 exhibit response means and distributions to the scale questions that are more similar to the mail questionnaire Phase 1 respondents (Treatment 1) than they are to the telephone respondents for either phase of these two treatments, then we will have increased confidence that any effects observed across modes in Phase 1 stem from the mode, rather than the attributes of early (Phase 1) vs. late (Phase 2) respondents. This in fact turns out to be the case. Table 4 shows that there are no significant differences in means between the phases across treatment groups that use the same mode. Table 5 illustrates consistencies in the frequency distributions for the first and last scale options from each question across phases that use the same mode (Column 1 vs. Column 5 and Column 4 vs. Columns 2, 8, and 11).

Second, we evaluate potential effects between modes for only Phase 1 data collection. This is the manner in which mode comparisons are usually made and Phase 1 utilizes the four modes in this study. To facilitate these comparisons, the means

Table 4

Chi-square tests for mode differences between phases of data collection and across treatments^a

Variable	Mail Treatment 1, Phase 1 vs. Treatments 2 and 3, Phase 2	Phone Treatment 1, Phase 2 vs. Treatments 2 and 3, Phase 1	Phone Treatment 4, Phase 2 vs. Treatments 2 and 3, Phase 1	Phone Treatment 5, Phase 2 vs. Treatments 2 and 3, Phase 1
Q2—overall satisfaction $\chi^2(4)$	1.471	3.198	8.829	0.866
Q3—overall value $\chi^2(4)$	6.748	2.813	7.220	3.023
Q4—retention $\chi^2(4)$	6.570	4.177	2.130	3.552
Q5—recommend $\chi^2(4)$	2.400	0.859	0.847	7.356
Q6—local usage $\chi^2(9)$	4.533	11.240	9.675	6.261
Q7—product/service quality $\chi^2(4)$	4.828	7.553	6.964	2.601

^a Means for statistical tests reported in Table 5.

Table 5

Means of all opinion questions for all modes in all treatments (n's vary slightly for each item)

Variable	Treatment 1			Treatments 2 and 3			Treatment 4			Treatment 5		
	Phase 1 Mail n = 1495 Column 1	Phase 2 Phone n = 157 Column 2	Total Column 3	Phase 1 Phone n = 1308 Column 4	Phase 2 Mail n = 1094 Column 5	Total Column 6	Phase 1 IVR n = 568 Column 7	Phase 2 Phone n = 438 Column 8	Total Column 9	Phase 1 Web N = 249 Column 10	Phase 2 Phone n = 700 Column 11	Total Column 12
Q2—overall satisfaction												
Mean	3.70	4.21	3.74	4.11	3.67	3.91	3.96	4.06	4.00	3.84	4.11	4.04
% choosing first option	2.3	0.6	2.1	2.1	2.9	2.5	3.4	2.1	2.8	2.4	2.0	2.1
% choosing last option	21.1	43.6	23.2	39.1	20.1	30.5	38.6	37.2	38.0	26.5	40.5	36.8
Q3—overall value												
Mean	3.46	3.97	3.51	3.85	3.44	3.66	3.59	3.90	3.73	3.48	3.92	3.80
% choosing first option	2.4	1.9	2.4	2.5	3.1	2.8	3.5	1.8	2.8	2.4	1.9	2.0
% choosing last option	11.4	30.8	13.3	26.3	11.4	19.5	21.1	28.9	24.5	9.3	29.0	23.8
Q4—retention												
Mean	3.74	4.23	3.78	4.12	3.66	3.91	3.98	4.15	4.06	3.79	4.19	4.08
% choosing first option	4.7	1.3	4.4	4.12	4.4	4.2	4.1	3.4	3.8	3.2	3.7	3.6
% choosing last option	28.8	54.5	31.2	50.19	26.1	39.2	43.5	49.4	46.1	29.0	51.8	45.8
Q5—recommend												
Mean	3.31	3.76	3.35	3.68	3.27	3.49	3.59	3.64	3.61	3.35	3.77	3.66
% choosing first option	10.1	9.0	10.0	10.7	11.3	11.0	8.8	10.6	9.6	6.83	7.4	7.2
% choosing last option	18.6	38.1	20.5	35.7	18.2	27.7	32.6	34.3	33.3	18.07	38.2	32.9
Q6—local usage^a												
Mean	6.21	6.27	6.22	6.44	6.15	6.31	6.49	6.53	6.51	5.96	6.61	6.43
% choosing first option	9.8	10.4	9.8	8.8	9.3	9.0	8.0	8.2	8.1	10.0	7.2	8.0
% choosing last option	17.5	22.7	18.0	21.2	17.2	19.3	21.6	23.7	22.5	16.5	22.1	20.6
Q7—product/service quality												
Mean	3.15	3.23	3.16	3.26	3.18	3.22	3.34	3.29	3.32	3.20	3.26	3.24
% choosing first option	0.7	0.6	0.7	1.3	1.2	1.3	4.2	0.5	2.6	0.4	1.3	1.1
% choosing last option	3.9	4.5	3.9	5.4	4.9	5.2	13.7	7.6	11.1	3.6	5.6	5.1

^a Ten-point scale.

and the percent choosing the first and last scale category of six¹ of the scale variables are reported (Table 3) so that any trends may be observed in addition to the specific statistical comparisons that are summarized in Tables 6–8. As shown in Table 5, the means and the frequency of respondents choosing the first and last scale options vary across modes in Phase 1, with the exception of mail vs. web, which have very little variation (Column 1 vs. Columns 4, 7, and 10; Column 4 vs. Columns 7 and 10; and Column 7 vs. Column 10). In regards to the visual questionnaires (e.g., mail and web), web respondents exhibited no statistical differences from the mail respondents (Table 6, Model 4). This result suggests that respondents to the visual questionnaires tended to respond similarly.

In sharp contrast, results from Table 6 show that the mean responses by mail consistently differed from those obtained by telephone. All six χ^2 tests were significant (Table 6, Model 2). Examination of the frequencies in Table 5 reveals that for the first four items, all of which used five-point scales with the ends anchored by labels, telephone respondents were much more likely to use the extreme positive category; differences on the other end of the scale were quite small. This tendency was greatly diminished on the final two items, although the χ^2 tests in Table 6 were significant, suggesting that telephone respondents are more likely to choose a higher positive category. One item, Question 6, used a 10-point scale (with anchored ends) and the other, Question 7, used a five-point scale with all categories labeled.

¹ The seventh scale item, Question 9 (Service Quality Satisfaction with Problem Resolution) has been omitted from remaining discussions because of the large number of respondents to whom it did not apply.

Table 6One-way post hoc Bonferroni and χ^2 tests for Phase 1 across modes and combinations of modes^a

Variable (df)	Bonferroni <i>f</i>	Significance tests						
		Mail vs. phone	Mail vs. IVR	Mail vs. web	Phone vs. IVR	Phone vs. web	IVR vs. web	Mail/web vs. phone/IVR
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 9
Q2—overall satisfaction χ^2 (4)	43.78	168.30**	72.21**	6.58	22.38**	28.84**	15.13**	168.82**
Q3—overall value χ^2 (4)	42.19	160.40**	36.42**	2.39	48.57**	48.22**	20.17**	143.99**
Q4—retention χ^2 (4)	30.28	139.07**	40.69**	2.01	9.67*	42.16**	17.14**	145.74**
Q5—recommend χ^2 (4)	21.82	124.88**	49.29**	3.74	17.09**	50.81**	25.24**	138.23**
Q6—local usage ^b χ^2 (9)	3.59	44.40**	21.76*	6.70	21.96**	25.36**	21.021*	47.40**
Q7—product/service quality χ^2 (4)	11.21	30.55**	114.46**	4.11	68.15**	4.62	38.05**	65.42**

^a Means for statistical tests are reported in Table 3.^b Ten-point scale.* $p < .05$.** $p < .01$.**Table 7**

Odds ratios for choosing above the mean for phone vs. IVR mode and demographic characteristics on answers to six survey questions

Variable	Q2: overall satisfaction?		Q3: overall value?		Q4: retention?		Q5: recommend to friend?		Q6: local usage?		Q7: product/service quality?	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Mode (IVR = 0) (phone = 1)	1.64*** (.19)	1.67*** (.21)	2.03*** (.21)	2.04*** (.23)	1.30*** (.15)	1.37** (.17)	1.36** (.14)	1.43*** (.16)	1.02 (.10)	1.04 (.11)	.68*** (.07)	.65*** (.07)
Education (ordinal)	—	1.06 (.05)	—	.99 (.04)	—	1.00 (.05)	—	.94 (.04)	—	.96 (.04)	—	.96 (.04)
Income (ordinal)	—	.97 (.05)	—	1.10 (.05)	—	1.03 (.05)	—	1.01 (.05)	—	1.03 (.05)	—	.91* (.04)
Age (continuous)	—	1.02* (.009)	—	1.01* (.01)	—	1.02* (.01)	—	1.02* (.01)	—	1.00 (.01)	—	1.01 (.01)
Gender (female = 1)	—	1.09 (.14)	—	1.05 (.12)	—	1.61*** (.19)	—	1.30** (.14)	—	1.10 (.11)	—	.94 (.11)
Children in HH (yes = 1)	—	1.12 (.24)	—	.95 (.18)	—	1.19 (.24)	—	1.04 (.19)	—	.96 (.17)	—	1.01 (.19)
HH size (continuous)	—	1.06 (.04)	—	.99 (.02)	—	1.02 (.03)	—	1.03 (.02)	—	.99 (.02)	—	1.00 (.02)
Mo. telephone bill (ordinal)	—	.92 (.04)	—	.98 (.04)	—	.93 (.04)	—	.99 (.04)	—	1.00 (.04)	—	1.00 (.04)
Model χ^2 (df = 1)	18***	—	47***	—	5**	—	9**	—	.02	—	13***	—
Model χ^2 (df = 8)	—	28***	—	50***	—	27***	—	23**	—	4	—	25***
Pseudo R^2	.01	.02	.02	.02	.002	.01	.004	.01	.00	.002	.006	.01
N	1876	1660	1878	1661	1879	1664	1874	1660	1846	1637	1877	1663

* $p \leq .05$.** $p \leq .01$.*** $p \leq .001$.

The mail vs. IVR comparison (Table 5, Column 1 vs. Column 7) displayed similar results. The frequency distributions for the first and last scale category demonstrate that IVR respondents choose the extreme positive category more often than mail respondents. Furthermore, all six χ^2 tests were significant, again demonstrating that IVR respondents are more likely to choose the higher positive category (Table 6, Model 4). These differences, however, were less pronounced for the 10-point scale, Question 6. Although significant, IVR respondents were only slightly more likely to choose a higher positive category than mail respondents, relative to the five-point scale questions.

Respondents to the web instrument behaved similarly to those to the mail instrument in comparison to telephone and IVR. Telephone respondents significantly differ from web respondents for all mean and distribution comparisons except for the final item, which used the fully labeled scale (Table 5, Column 4 vs. Column 10; Table 6, Model 6). This suggests that telephone respondents are also more likely than web respondents to choose the higher positive category, except when using

Table 8

Odds ratios for choosing above the mean for aural vs. visual modes and demographic characteristics on answers to six survey questions

Variable	Q2: overall satisfaction?		Q3: overall value?		Q4: retention?		Q5: recommend to friend?		Q6: local usage?		Q7: product/service quality?	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Mode												
(mail/web = 0)	2.11 ^{***}	2.11 ^{***}	1.82 ^{***}	1.88 ^{***}	1.64 ^{***}	1.68 ^{***}	1.64 ^{***}	1.66 ^{***}	1.10	1.14	1.57 ^{***}	1.58 ^{***}
(phone/IVR = 1)	(.16)	(.16)	(.12)	(.13)	(.12)	(.13)	(.06)	(.12)	(.07)	(.08)	(.12)	(.13)
Education	–	1.08 ^{**}	–	1.02	–	1.03	–	.98	–	.96	–	.99
(ordinal)		(.03)		(.03)		(.03)		(.03)		(.03)		(.03)
Income	–	.98	–	1.05	–	1.02	–	.98	–	1.02	–	.94
(ordinal)		(.04)		(.04)		(.04)		(.03)		(.04)		(.04)
Age	–	1.03 ^{***}	–	1.02 ^{**}	–	1.02 ^{***}	–	1.01 ^{**}	–	1.00	–	1.01 [†]
(continuous)		(.006)		(.005)		(.006)		(.005)		(.005)		(.006)
Gender	–	1.07	–	1.07	–	1.34 ^{***}	–	.94	–	1.08	–	1.05
(female = 1)		(.09)		(.08)		(.11)		(.09)		(.11)		(.09)
Children in HH	–	1.09	–	.94	–	1.06	–	.94	–	.89	–	.77 [†]
(yes = 1)		(.15)		(.12)		(.14)		(.12)		(.11)		(.10)
HH size	–	1.03	–	.98	–	1.01	–	1.01	–	.99	–	1.01
(continuous)		(.03)		(.02)		(.02)		(.02)		(.02)		(.02)
Mo. telephone bill	–	.96	–	1.00	–	.96	–	1.03	–	1.04	–	1.01
(ordinal)		(.03)		(.03)		(.03)		(.03)		(.03)		(.03)
Model χ^2 (df = 1)	106 ^{***}	–	79 ^{***}	–	47 ^{***}	–	55 ^{***}	–	2	–	35 ^{***}	–
Model χ^2 (df = 8)	–	131 ^{***}	–	93 ^{***}	–	72 ^{***}	–	66 ^{***}	–	13	–	51 ^{***}
Pseudo R^2	.02	.03	.02	.02	.01	.02	.01	.01	.0004	.003	.008	.01
N	3620	3330	3622	3333	3625	3336	3620	3333	3589	3307	3621	3334

† $p \leq .05$.** $p \leq .01$.*** $p \leq .001$.

fully labeled scales. IVR respondents also differ significantly from web respondents (Table 5, Column 7 vs. Column 10) for all six scale questions. The significant differences illustrated in Table 6 (Model 7) suggests that IVR respondents, like telephone respondents, are more likely to choose a higher positive category compared to web respondents, even for the last question with fully labeled scales.

A mixed picture exists with respect to the two aural modes, IVR vs. telephone (Table 5, Column 4 vs. Column 7). Significant differences exist for all mean comparisons with telephone respondents slightly more likely to choose a higher positive category for Questions 2 through 5 (Table 6, Model 5). On Questions 6 and 7, however, IVR respondents are more likely to choose the higher positive category. This seems to be consistent with Tourangeau et al.'s (2002) study and could suggest that telephone respondents are more likely than IVR respondents to choose the middle or lower categories when the scales are greater than 5 categories (Question 6) and when the scale is fully labeled (Question 7). For the fully labeled scale, the difference may be due to telephone respondents hearing only word labels, while IVR respondents hear the word labels and numbers associated with each label, providing them with visual markers on the telephone number pad.

The mixed findings between phone vs. IVR modes might also be accounted for by the fact that people with different characteristics (i.e., different concerns about their long distance service) were more likely to respond differently to each mode. For this reason, we conducted a logistic regression analysis for phone vs. IVR modes, controlling for demographic characteristics (Table 7). Table 7 illustrates the probability of selecting a category above the mean for phone and IVR respondents. It demonstrates that telephone respondents are significantly more likely to choose a higher positive category for Question 2 through 5 up to a factor of 2.03 (Question 3, Model 3). When controlling for demographics, such as education, age, income, gender, household size, children, and monthly telephone bill, telephone respondents are still more likely to choose a higher category, by up to a factor of 2.04 (Question 3, Model 4). Question 6, the 10-point scale question, exhibited no significant difference between telephone and IVR, suggesting that a longer scale could alleviate mode effects between telephone and IVR. For Question 7, the fully labeled scale question, IVR respondents are significantly more likely to choose a higher positive category, supporting the notion that respondents who hear only words, as in the telephone treatment group, are more likely to choose middle or lower categories compared to those respondents hearing both words and numbers, as in the IVR treatment group. Again, this could be due to the visual prominence of numbers on the telephone keypad and to the notion that telephone respondents, hearing only words, may not process the question responses in a scale format.

Finally, it seems plausible that the major comparison is between visual vs. aural modes. As shown above, mail and web respondents behave similarly as do phone and IVR respondents, comparatively (Table 5). To test for visual vs. aural mode effects, we combined the web and mail modes (visual modes) and the telephone and IVR modes (aural modes) for Phase 1 data. Table 6 (Model 9) shows that respondents to aural modes are significantly more likely than respondents to visual modes to choose a higher positive category for all six questions. In fact, the differences in means between visual vs. aural

modes are the greatest among all the mode comparisons. To ensure that these differences are not the result of differences in respondent characteristics, we conducted a logistic regression analysis controlling for demographics (Table 8). Table 8 illustrates the probability of choosing a category above the mean for aural and visual respondents. Even when controlling for demographics, significant differences exist between aural vs. visual modes for all questions except the 10-point scale question (Question 6). On Question 2, for example, respondents to aural modes are significantly more likely than visual mode respondents to choose a higher positive category by a factor of 2.11 (Model 1); the odds ratio also remains unchanged when controlling for demographics (Model 2). Thus, it is reasoned that respondents using aural modes are significantly more likely than respondents using visual modes to choose the higher positive category when the question scale is five points and whether the scale is fully labeled or not.

5. Discussion and conclusions

The use of two or more survey modes in a single data collection effort raises the possibility of improved response rates being achieved. However, those improvements may come at the cost of obtaining different answers to the survey questions asked of respondents. The cost involved is that it is very likely to obtain different answers to each mode. In this study of a national quasi-general public survey of individuals on satisfaction with their long distance service, for whom both telephone and postal mail addresses were available, we have tested whether response rates could be improved by offering a second survey mode after multiple attempts to collect data by the first mode. In addition, we have evaluated whether switching modes caused these critical evaluation questions to be answered differently.

The results show clearly that response rates can be improved by this sequential offering of survey modes. This is particularly apparent when response to the first mode is fairly low, as observed for the web and IVR modes and to some extent the telephone interview. Significant improvements were achieved by reverting from IVR and web back to the completion of telephone interviews. The exception is the mail mode, for which use of the telephone in Phase 2 only increased response rate by eight percentage points. Offering mail to telephone nonrespondents, however, resulted in obtaining results from nearly two-thirds of the outstanding sample members. The high response rate to the mail questionnaire may be attributed to a short survey (18 questions), the presentation of a visually attractive questionnaire, and the inclusion of a \$2.00 incentive. In addition, the topic was one that seemed likely to be salient to the households receiving it. We would expect the response rate to be lower with longer questionnaires on less salient topics, and for which less adequate mailing procedures were used.

The success of the second mode in reducing potential nonresponse error for demographic characteristics, even for the treatment groups with low initial response rates, appears to be small. Although some differences between Phase 1 and Phase 2 respondents were noted for self-reported demographics, they appeared to be fairly minor (with the exception of the higher proportion of male respondents answering the web questionnaire in Treatment 5). Regardless, when we compared the characteristics of respondents and nonrespondents for sample frame demographics, significant differences existed between the two groups. Nonresponding households tended to be less educated, were more likely to be female, and were younger than those responding to the survey. Income differences, the only household variable, was an exception with only one significant difference (for the IVR treatment) between respondents and nonrespondents.

The consequences for measurement of changing modes are significant in our analysis. Much in line with previous research, aural mode respondents tend to give more extreme responses on the positive ends of the scale compared to visual mode respondents, even when controlling for demographics. Mail and web respondents responded similarly while telephone and IVR respondents differed depending on the scale response format. The mode differences between visual and aural modes observed here cannot be accounted for by the recency hypothesis which proposed that respondents to interview surveys are more likely to select the last mentioned response category. We believe that a more plausible explanation is that the mail and web versions have all response choices visually presented to the respondents. Recent experimental research has shown that telephone responses are consistently more extreme towards the positive end of the scale for a variety of scale types, including the scale format used in this study (Christian et al., 2008; Tourangeau et al., 2002). Printed response boxes present more equally available response choices than do the interim categories when a response stimulus such as this is delivered aurally to the respondent: "Overall, how satisfied are you with your long distance company? Please use a one-to-five scale, where '5' means extremely satisfied, and '1' means not at all satisfied. You may use any number from one to five."

The differences observed in the current study have practical consequences. In addition to the obvious difficulty of combining results from two modes in a mixed-mode survey, we live in a time when surveys are conducted with increasing frequency with the goal of measuring changes over time. As more alternatives become available, we expect that in situations such as customer satisfaction surveys and employee work place surveys, surveyors may wish to switch from one mode to another in order to reduce costs or simply as a matter of preference. Results from this analysis leave little doubt that a mixed-mode strategy of following a complete data collection strategy by one mode, with a short pause, followed by an attempt to collect data by another mode, can increase response rates substantially. If one's intent is to measure change, however, switching modes could have major negative consequences for measures of satisfaction. For example, in this survey of long distance satisfaction, the choice of a telephone interview would have produced an estimate of 39% extremely satisfied customers. On the other hand, the choice of mail would have produced an estimate of 21%, about half as large, with somewhat similar differences being obtained for other measures examined here.

Limitations of this study need to be recognized. Our search for a sample frame that would provide equal telephone and mail access to all households as well as provide some sample frame demographics meant that compromises were necessary. The sample frame available to us was from people who had voluntarily sent in product warranty information or completed surveys about shopping behavior or product preferences. The sample frame was further limited by our need for information on sample frame demographics. We were also restricted by the use of diverse warranty question formats, i.e., different categories of income. As a result of discussions with the sample provider, and a mistake we were unaware of until the study was in the field, the sample vendor had limited the list to households with children. The list was skewed heavily towards higher household incomes as well. Although the majority of the sample included requested demographics, missing data was still an issue. Therefore, it is important that further research be done on true general public populations, though we recognize the difficulty of obtaining both postal and telephone contact information for such households.

In addition, three of the four sample frame characteristics available to us were individual ones. For these variables we are limited to conclusions of this nature: “Households in the sample frame in which a female sent in the warranty information responded to a greater or lesser degree than households in which a male sent in that information.” Our specific interest in responses to questions about long distance service and the perceived importance of getting that information from the person in the household who made those decisions did not allow us to seek as a respondent only the person whose name was on the sample list. This remains another limitation of this study.

Moreover, we examined differences for opinion questions that utilized a format favored by many designers of interview and IVR surveys, i.e., a scale format with labeled endpoints that surveyors found easier to administer in these types of surveys as opposed to scales with all points labeled. This format was chosen based upon our observation that anchored endpoints are being used more frequently in mail and web surveys. However, recent research has shown that similar extremeness effects also exist for fully labeled five- and seven-point scales (Christian et al., 2008; Christian, 2007).

Another question for future research is the relative power of following the attempts to obtain web and IVR responses with a mail survey in Phase 2, rather than telephone. In many ways the telephone attempts during Phase 2 were similar to the initial contacts, i.e., both involved interaction by phone. It is reasonable to expect that switching to mail at this stage would have had a much greater impact on improving response to these treatment groups, but remains to be tested experimentally.

In the meantime, there can be little doubt that mixing of survey modes can improve response rates, as shown previously by Shettle and Mooney (1999) and Griffin and Obenski (2002). However, using an alternative mode that depends upon a different channel of communication, i.e., aural vs. visual, to increase response may also introduce measurement differences issues that cannot be ignored. Understanding the basis of these differences should be a high priority for future research.

References

- Christian, L.M., 2007. How Mixed-Mode Surveys are Transforming Social Research: The Influence of Survey Mode on Measurement in Web and Telephone Surveys. Unpublished Ph.D. Dissertation. Washington State University, Pullman.
- Christian, L.M., Dillman, D.A., Smyth, J.D., 2008. The effects of mode and format on answers to scalar questions in telephone and web surveys. In: Lepkowski, J., Tucker, C., Brick, M., de Leeuw, E., Japac, L., Lavrakas, P., Link, M., Sangster, R. (Eds.), *Advances in Telephone Survey Methodology*. Wiley-Interscience, New York, pp. 250–275 (Chapter 12).
- Christian, L.M., Dillman, D.A., 2004. The influence of symbolic and graphical language manipulations on answers to paper self-administered questionnaires. *Public Opinion Quarterly* 68 (1), 57–80.
- Cook, C., Heath, F., Thompson, R.L., 2000. A meta-analysis of response rates in web- or Internet-based surveys. *Educational and Psychological Measurement* 60, 821–826.
- Couper, M.P., 2000. Web surveys: a review of issues and approaches. *Public Opinion Quarterly* 64, 464–494.
- de Leeuw, E.D., 1992. *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. TT Publications, Amsterdam.
- de Leeuw, E.D., 2005. To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21 (2), 233–255.
- Dillman, D.A., in press. Some consequences of survey mode changes in longitudinal surveys. In: Peter, L. (Ed.), *Methodology of Longitudinal Surveys*. John Wiley Co., London.
- Dillman, D.A., 2007. *Mail and Internet Surveys: The Tailored Design Method*, second ed. Hoboken, John Wiley Co., New Jersey (2007 Update).
- Dillman, D.A., Christian, L.M., 2005. Survey mode as a source of instability in responses across surveys. *Field Methods* 17 (1), 30–51.
- Dillman, D.A., Sangster, R.L., Tarnai, J., Rockwood, T., 1996. Understanding differences in people's answers to telephone and mail surveys. In: Braverman, M.T., Slater, J.K. (Eds.), *New Directions for Evaluating Series 70 (Advances in Survey Research)*. Jossey-Bass, San Francisco.
- Dillman, D.A., Brown, T.L., Carpenter, E.H., Carlson, J., Lorenz, F.O., Mason, R., Sangster, R.L., Saltiel, J., 1995a. Effects of category order on answers to mail and telephone surveys. *Rural Sociology* 60, 674–687.
- Dillman, D.A., Clark, J.R., West, K.K., 1995b. Influence of an invitation to answer by telephone on response to census questionnaires. *Public Opinion Quarterly* 58, 557–568.
- Dillman, D.A., Robert, G.M., 1984. The influence of survey method on question response. Paper Presented at the Annual Meeting of the American Association for Public Opinion Research. Delavan, WI.
- Griffin, D.H., Sally, M.O., 2002. Meeting 21st Century Demographic Needs Implementing the American Community Survey: May 2002, Report 2: Demonstrating Survey Quality. US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 45pps.
- Groves, R.M., Kahn, R.L., 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. Academic Press, New York.
- Horrigan, J.B., Smith, A., 2007. Home Broadband Adoption, 2007. Pew Internet & American Life Project. Accessed July 10, 2007. Available from: http://www.pewinternet.org/pdfs/PIP_Broadband%202007.pdf.
- Jenkins, C.R., Dillman, D.A., 1997. Towards a theory of self-administered questionnaire design. In: Lyberg, L., Biemer, P., Collins, M., Decker, L., de Leeuw, E., Dippo, C., Schwarz, N., Trewin, D. (Eds.), *Survey Measurement and Process Quality*. Wiley-Interscience, New York, pp. 165–196 (Chapter 7).
- Krosnick, J.A., 1999. Survey research. *Annual Review of Psychology* 50, 537–567.
- Krosnick, J., Alwin, D.F., 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51, 201–219.
- Krysan, M., Schuman, H., Scott, L.J., Beatty, P., 1994. Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly* 58, 381–399.
- Lesser, V., Dillman, D., Lorenz, F.O., Carlson, J., Brown, T.L., 1999. The influence of financial incentives on mail questionnaire response rates. Paper Presented at the Meeting of the Rural Sociological Society, Portland, Oregon (August).

- Moore, D., 1998. Testing propositions about the occurrence of primacy effects in telephone surveys. Paper Presented at the Annual Meeting of the American Association for Public Opinion Research, May 14–17. St. Louis, MO.
- Mu, X., 1999. IVR and Distribution of Responses: An Evaluation for the Effects of IVR on Collecting and Interpreting Survey Data. Unpublished Paper. The Gallup Organization, Princeton, NJ.
- Schuman, H., Presser, S., 1981. Questions and Answers in Attitude Surveys Experiments on Question form Wording and Context. Academic Press, New York.
- Shettle, C., Mooney, G., 1999. Monetary incentives in US Government surveys. *Journal of Official Statistics* 15 (2), 231–250.
- Singer, E., 2002. The use of incentives to reduce nonresponse in household surveys. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. John Wiley Co., New York, pp. 163–178 (Chapter 11).
- Singer, E., 2006. Introduction: nonresponse bias in household surveys. *Public Opinion Quarterly* 70 (5), 637–645.
- Srinivasan, R., Hanway, S., 1999. A new kind of survey mode difference: experimental results from a test of inbound voice recognition and mail surveys. Paper Presented at the Meeting of the American Association for Public Opinion Research, May. St. Pete Beach, FL.
- Steiger, D.M., Beverly, C., 2008. Interactive voice surveys. In: de Leeuw, E., Dillman, D.A., Joop, J.H. (Eds.), *International Handbook of Survey Methodology*. Taylor and Francis, Boca Raton, FL.
- Sudman, S., Bradburn, N.M., Schwarz, N., 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass Publishers, San Francisco, 304pps.
- Tarnai, J., Dillman, D.A., 1992. Questionnaire context as a source of response differences in mail versus telephone surveys. In: Schwarz, N., Sudman, S. (Eds.), *Context Effects in Social and Psychological Research*. Springer-Verlag, New York.
- Tourangeau, R., Couper, M.P., Conrad, F., 2007. Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly* 71 (1), 91–111.
- Tourangeau, R., Couper, M.P., Conrad, F., 2004. Spacing, position, & order: interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly* 68 (3), 368–393.
- Tourangeau, R., Steiger, D.M., Wilson, D., 2002. Self-administered questions by telephone: evaluating interactive voice response. *Public Opinion Quarterly*, Summer 66(2), 265–278.
- Ware, C., 2004. *Information Visualization: Perception for Design*. Morgan Kaufmann, San Francisco.