

## **Response Rate and Measurement Differences in Mixed Mode Surveys**

### **Using Mail, Telephone, Interactive Voice Response and the Internet<sup>1</sup>**

by

Don A. Dillman, Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, and Jodi Berck

#### **Abstract**

The potential for improving response rates by changing from one mode of data collection to another mode is examined in this paper, along with the consequences for measurement and nonresponse errors. Using a data set of 8,999 households, data collection consisted of two phases. Phase 1 data collection was conducted by telephone interview, mail, interactive voice response, and the Internet, while Phase 2 consisted of nonrespondents to Phase 1, and was conducted by telephone or mail. In general, 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 these surveys, 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 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.

---

<sup>1</sup> Financial Support for this study was provided by The Gallup Organization. Additional support was provided by the Department of Rural Sociology under Western Region Project W-183, 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.

## **Response Rate and Measurement Differences in Mixed Mode Surveys**

### **Using Mail, Telephone, Interactive Voice Response and the Internet<sup>2</sup>**

by

Don A. Dillman, Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, and Jodi Berck

An important issue facing survey research is how to obtain satisfactory response rates, which appear to have been declining during the 1990's (Atrostic and Burt, 1999). Changing modes of data collection may be an effective way for improving response rates (e.g., Shettle and Mooney, 1999). For example, nonrespondents to a mail survey may be called and asked to answer by telephone or vice versa.

The use of a mixed-mode data collection strategy raises a second, potentially troubling issue. Of concern is whether people who respond by one mode provide the same answers as would have been the case had they responded by another mode. Traditionally, surveys have been implemented by face-to-face interviews, telephone interviews, or mail questionnaires, for which mode differences in responses have frequently been observed. Two new methods of conducting surveys—touch-tone data entry of answers in response to recorded instructions (also known as Interactive Voice Response or IVR), and Internet surveys—are now being used with greater frequency in mixed-mode survey situations, thus introducing the possibility of new sources of mode differences.

---

<sup>2</sup> Financial Support for this study was provided by The Gallup Organization. Additional support was provided by the Department of Rural Sociology under Western Region Project W-183, 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.

The purpose of this paper is to report results of an experiment in which sample units that did not respond to an initial mode of data collection (mail, telephone, web or IVR) after a series of attempts, were assigned to a second mode of data collection (either telephone or mail) as a means of improving response rates. We examine both the separate and cumulative response rates to the initial and follow-up modes. In addition, we examine whether respondents gave similar or different answers to each survey mode for a series of opinion questionnaires. Finally, we determine whether use of a second mode of data collection brought respondent demographics closer to those of the original sample frame.

## **Theoretical Background**

### Mixing of Modes to Improve Response Rates

It has long been recognized that some respondents prefer being surveyed by one 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 rate. Dillman, Clark, and West (1995) 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 changed their mode of response. 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 using the telephone. In contrast, it has been shown that a sequential strategy of making 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 including an incentive, 81% after an intensive telephone follow-up, and finally 88% after attempts to complete in-person interviews. Each of these phases was started after efforts for the previous phase had been concluded, rather than being intermingled in some way. For these reasons the current study evaluates the sequential use of modes rather than offering respondents a choice of modes by which to respond.

In this study, we examine the effects of 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.

Implementing a second survey mode is costly. The main justification for using a second mode is to reduce nonresponse error. It has been argued that less educated people are more likely to respond to telephone surveys than they are to mail surveys, which require reading and writing skills (e.g., Sudman and Bradburn, 1982). Furthermore, some respondents are difficult to locate by telephone, but may be less difficult to locate by mail, and vice versa. It has been estimated that slightly less than half of the U.S. population has access to the Internet (U.S. Department of Commerce, 2000), therefore web surveys compound the concern of nonresponse error even further. IVR surveys, which often use telephone recruitment followed by a transfer to the recorded interviewing system, remain relatively unstudied with respect to bias. An important aspect of this investigation is to understand the role of a follow-up mode in reducing nonresponse error.

### Mode-Differences in Respondent Answers

For several decades, reports of mode experiments have appeared in the survey literature.

Together, they suggest that significant differences sometimes exist in the answers that people give to face-to-face interviews, telephone and mail surveys (de Leeuw, 1992; Schwarz, Strack, Hippler and Bishop, 1991). Three aspects of these modes appear to contribute to the occurrence of mode differences: 1) presence or absence of an interviewer, 2) the use of communication that is aural versus visual, and 3) whether control of the question stimulus is by the interviewer or the respondent (Dillman, Sangster, Tarnai, and Rockwood, 1996). Because each of these factors may influence responses in different ways, the potential mode effects of each needed to be taken into consideration in the design of this study.

Presence of an interviewer encourages respondents to give answers that put them into a favorable light as defined by the culture that the interviewer is assumed to represent.

Experiments have shown that respondents to interviews are more likely to offer socially desirable answers in the presence of an interviewer than in the self-administered situation (Hochstim, 1967; de Leeuw, 1992). It has also been suggested by many (e.g., Schuman and Presser, 1981) that respondents are more likely to acquiesce, i.e., agree than disagree when interacting with other people. Acquiescence might also be expected from potentially fast pacing when the interviewer controls the answering process, and encourages the respondent towards use of a simplified cognitive representation and answering scheme. It is expected that respondents to interviews are more likely to express agreement in response to questions, than are respondents to mail questionnaires (de Leeuw, 1992). Dillman and Tarnai (1991) found limited support for this hypothesis. Respondents to a telephone sample were more likely than respondents to a mail sample to agree that each of nine proposals would be effective for improving the wearing of seat

belts. To better isolate potential causes of mode differences in the current study, an effort was made to avoid questions with known social desirability attributes and formats such as agree-disagree questions that would encourage an acquiescent response.

When a stimulus is completely visual, as in mail questionnaires, the respondent controls how fast each question is read, or even whether it is read in its entirety. One possible result of these stimulus and control features of questionnaires is whether a primacy or recency effect occurs when selecting response choices. Krosnick and Alwin (1987) found that respondents in a face-to-face interview who were given a written list of desirable qualities that they might wish for their children to possess, were more likely to choose categories early in the list. This tendency is known as a primacy effect. They argued that when the choices are presented visually to respondents that the items listed first are subjected to deeper cognitive processing, thus establishing a standard of comparison that guides interpretation of later items. They also argued that when items are read by an interviewer and received aurally by respondents, there is not enough time for the respondent to place each answer choice into long-term memory before the next one is read. As a consequence, respondents are more likely to choose the last categories on a list. This tendency is described as a recency effect. Although considerable research has reported both recency and primacy effects (e.g., Schuman and Presser, 1981; Dillman, Sangster, Tarnai and Rockwood, 1996), that research is inconclusive. A series of 84 experiments by a group of researchers (Dillman et al., 1995) in many different surveys did not reveal a consistent pattern of effects. In light of the conflicting results and difficulty of knowing for which types of questions an effect might occur, it seemed important in an experiment such as the current study to control for potential primacy-recency effects.

Other evidence on the effect of visual versus aural presentation is somewhat limited. One study that sought to investigate such differences found that responses to a series of community satisfaction questions were more extreme on the telephone, i.e., substantially more people chose “not a problem,” the first offered category, as opposed to small, moderate or serious than when surveyed was self-administered. However, when respondents were surveyed by telephone while they followed along on a paper version, the percent choosing “not a problem” was intermediate between the mail and telephone versions (Tarnai and Dillman, 1992). It appeared that seeing the visual format while answering questions gave somewhat more prominence to the middle categories.

The newness of IVR and web survey methodologies is such that very little experimental research has been done, and the possible effects on survey measurement are mostly unknown. The web survey is a visual stimulus and under the respondent’s complete control with regard to whether and/or how each question is read and comprehended. Therefore, we expect that responses to web questionnaires will more closely resemble those observed for mail questionnaires than those observed for interview surveys (Dillman, 2000). It is less clear what the effect of IVR might be. Like the telephone survey, this survey stimulus is aural, but the respondent knows that the questions are prerecorded, and that when answering she is not actually giving answers to another person whom is recording her answers. Thus, we might expect less social desirability or acquiescence than on the telephone, but the same conditions as those that could produce a recency effect in telephone (aural communication and control of pacing by interviewer), may produce similar results.

Mode effects are also likely to vary by type of question. For example, demographic questions such as age and education seem less prone to mode effects than do attitude or opinion

questions. There are two reasons. First, most respondents know the answer they want to give when demographic questions are asked so that if categories are offered, the respondent only has to wait until she hears the appropriate response category in order to respond. In contrast, the satisfied-dissatisfied questions, types of questions included in this study, seem more subject to mode effects. For these types of questions, the answer choices as well as the questions must be read to respondents. Respondents are less likely to have a preformed answer. Thus, they are required to contemplate their attitude as well as the scale on which they are being asked to respond. People do not possess “somewhat satisfied” in the same way that they possess a level of education.

Choice of survey mode often affects how questions are structured. Interview modes that do not allow respondents to see answer choices tend to encourage survey designers to use shorter scales, and/or scales without labeled categories. For example, it becomes quite laborious for interviewers to read all labeled scale choices to respondents, e.g., “Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?” This has encouraged companies such as The Gallup Organization to frequently 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.” The use of these scales in mode comparisons such as mail versus telephone would seem to decrease the likelihood of primacy or recency effects, because there are fewer categories to be heard and remembered, but it also introduces another potential for mode differences. The fact that the mail questionnaire is visual, and all five categories appear clearly on the questionnaire, raises the likelihood that mail respondents are more likely than telephone respondents to distribute their answers across the scale. Previous



research on community and neighborhood satisfaction, using fully labeled scales, has confirmed this tendency (Tarnai and Dillman, 1992).

With this in mind, we expect that web survey respondents would behave similarly to mail respondents, but the situation is less clear for IVR respondents, where 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 from research by Srinivasan and Hanway (1999), who found for eleven items using five-point scales with labeled endpoints that IVR respondents were significantly more likely (mean difference = six percentage points) to choose strongly agree than were mail survey respondents. 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. Although the differences were in the same direction, 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 labels on a paper questionnaire pull respondents even more strongly to the intermediate categories than do unlabeled categories. In summary, there are good reasons for expecting that the visual versus aural presentation of questions to respondents may influence people's answers to mail, telephone, web and IVR surveys. Our general expectation is that mail and web respondents will make greater use of unlabeled middle categories than will telephone respondents, and that IVR respondents, though more like telephone than mail or web respondents, will be intermediate in their use of extreme categories.

## The Study

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.

Most of these objectives were accomplished by purchasing a list of 8,999 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 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, 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 house-hold, presence of children under 18, and income. Questionnaires were developed through an interactive process designed to achieve a common stimulus for all survey modes for each of the questions, using a unimode construction procedure similar to that described in Dillman (2000).

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:

**Treatment 1. Phase 1 Mail questionnaire, Phase 2 Telephone interview.** This random subsample of 2,000 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 11x17-inch sheet of paper that was folded to a conventional 8 1/2" x 11" size. Questions were printed in black ink on blue background fields with white boxes for marking answers. 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 (2000).

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.

**Treatments 2 and 3. Phase 1 Telephone interview, Phase 2 Mail questionnaire.** A total of 2,999 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....” 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 Figure 1).

Form A or Form B was randomly assigned to each respondent at the beginning of the interview, based on interviewing session. 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 attempting 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.

**Treatment 4. Phase 1 IVR recruited by telephone survey, Phase 2 Telephone interview.**

Another 2,000 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 5 minutes.” Respondents were then asked to stay on the line, and that it would take about ten seconds to transfer them. Those who stayed on the line then heard a prerecorded welcome that began, “During the next few minutes, we’ll 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, e.g., “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.” 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 procedure described for Treatments 2 and 3 were followed.

**Treatment 5. Phase 1 Web recruited by telephone survey, Phase 2 Telephone interview.**

The 2,000 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 their from 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 thirty 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 e-mail 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. 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.

## **Results**

### 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 for Phase 1 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 seconds). 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 in 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 15–20 percentage points by the inclusion of such an incentive (Lesser, et al. 1999). In addition, research on the use of incentives with telephones has shown that incentives sent after completion of the interview, which would have necessarily been the case with the telephone procedures used here, are not an effective way to improve telephone response rates (Singer, In Press). Another factor that undoubtedly contributed to the high response rate for Treatment 1 was the shortness of the questionnaire.

Response rates for all five treatments were increased substantially by switching to a second mode of data collection, once Phase 1 had been completed. 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 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 a substantial portion of these individuals had already been contacted by telephone, and to some degree being called 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.

#### Category Order Effects for Scalar Questions Asked by Telephone

The seven opinion questions that asked for scalar 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 a recency nature that might be expected, based upon certain past research.

Table 2 presents the full category distributions for all seven questions so that potential trends across questions may be fully examined in addition to whether statistically significant differences exist. 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” (presented aurally in last position) would be larger than in the other treatment. This is not the case. Table 2 shows that none of the differences are statistically significant, as measured by both a chi-square and a difference of means test, and there is no consistent trend in either direction. Based on the lack of differences, we have combined Treatments 2 and 3 together for all other analyses reported in this paper.

#### Mode Effects across All Treatment Groups

Examination for potential mode effects involves two steps. First, we evaluate potential effects between treatment groups for only Phase 1 Data collection. This is the manner in which mode comparisons are usually made. However, as shown in Table 1, response rates differed greatly

across modes. Thus, any differences observed here might be accounted for by the fact that people with different characteristics (e.g. different concerns about their long distance service) were more likely to respond to one mode as opposed to another mode. For this reason, any conclusions drawn from comparing Phase 1 responses across treatment groups must be considered tentative.

By comparing Phase 1 responses from one treatment to Phase 2 responses by the same mode in other treatment groups, we can gain insight into whether any differences in distributions are more likely the result of mode or type of respondent to Phase I. If the mail questionnaire respondents to Phase 2 of Treatments 2 and 3 exhibit response distributions to the scalar questions that are more similar to the mail questionnaire Phase 1 respondents (Treatment 1) than they are to the telephone distributions 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 I) and late (Phase 2) respondents. This in fact turns out to be the case.

To facilitate these comparisons, the full distributions of six<sup>3</sup> of the scalar variables are reported (Table 3), so that any trends may be observed in addition to the specific statistical comparisons that are summarized in Table 4. Results show that the responses by mail in Treatment 1 differed consistently from those obtained by telephone for the combined group of Treatment 2 and 3 respondents. All six chi-square tests were highly significant, as were five of the six means comparisons, the one exception being for the 10-point scale (Table 4, Column 2). Examination of the distributions in Table 3 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.

---

<sup>3</sup> The seventh scalar 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.

This tendency was greatly diminished on the final two items. One item used a 10-point scale (with anchored ends) and the other used a five-point scale with all categories labeled. In all cases, the means for the telephone treatment group were higher than the means in the other modes. In sharp contrast, the other visual questionnaire, the web instrument, exhibited no statistical differences from the mail questionnaire, either for means or for distributions (Table 4, Column 4). This result suggests that respondents to the visual questionnaires tended to respond similarly. Telephone respondents, on the other hand, differ from web respondents, exhibiting significant differences for all mean and distribution comparisons except for the final item, which used the fully labeled scale, and the mean on the 10-point scale item (Table 4, Column 6).

However, a mixed picture exists with respect to IVR versus telephone respondents (Table 4, Column 5). For the first two items, significant differences exist for both means and distributions, but do not exist for the means test on the last four items. Visual examination of the distributions suggests that although IVR respondents are more similar to telephone respondents in their use of the extreme positive category, in three of the six listed items, they are also much more likely to use category 3, the middle category, than are telephone respondents. Interestingly, with the last item (product/service quality), the opposite occurs where telephone respondents are much more likely than IVR respondents to use the middle category. This result may be due to telephone respondents hearing only word labels, while IVR respondents hear the word labels and numbers associated with each label. It seems plausible that this use of the middle “three” category may be associated with that category being at the end of the first line of numbers on the touch-tone phone, and therefore having more visual prominence.

The comparison of similar modes across phases is limited to mail and telephone comparisons. None of the comparisons exhibit a significant difference (Table 5). These findings suggest

---

strongly that the mode differences between mail and telephone shown in the preceding tables are the result of direct mode effects than differences in types of people who responded to Phase 1 versus Phase 2 of the data collection.

#### 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 6, Phase 1 and 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 mailout. 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 strategy 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 6,

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, and 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 nonrespondents as well as respondents, 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 7). 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 6. In Table 7, 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 7 (e.g., 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 frame. Comparisons in this table tell us whether, e.g., 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 7: Phase 1 versus Phase 2, and those responding to the survey versus non-responding households. The first comparison (Phase 1 versus Phase 2) provides another indicator, in addition to Table 6, 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 7 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 7, nonresponding households differed from those that responded. In each case, non-responding 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.

### **Discussion and Conclusions**

The decision to undertake this investigation was motivated by concern that relatively little was known about the potential mode impacts of the use of two new modes of data collection—interactive voice response and web—that are increasingly being considered for use in designing surveys. In addition, concern over declining response rates, particularly to telephone surveys, led us to consider the possibility that a mixed-mode strategy could be used to improve response rates substantially. We were also interested in learning whether such increases in response rates might reduce nonresponse error, i.e., the extent to which respondents differ from nonrespondents on characteristics of importance to the survey. Finally, we had observed a tendency among survey researchers to design question formats that seemed most suitable to a particular mode (Dillman, 2000), and which for theoretical reasons we thought might not produce the same results for different modes. Specifically, we had observed a tendency for designers of interview surveys to prefer using scales with only the polar points labeled, which seemed likely to produce a stimulus discrepancy between the aural and visual stimulus communicated to respondents.

Response differences exist as clearly shown in the data for the six scalar items examined in this survey. Respondents to telephone interview surveys are more likely than respondents to mail surveys to select an extreme response, and mail respondents are more likely to select one of the middle categories. This conclusion is based upon a comparison of results from Phase 1 and

Phase 2. There were no statistical differences between the response distributions for either mail or telephone during Phase 1, with the results obtained by their use as a Phase 2 data collection mode.

In addition, the overall mode differences between mail and telephone 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 version has all response choices visually presented to the respondents. 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.” Results for the web mode, which also used a visual form of communication, tracked closely with the results obtained for mail, thus reinforcing this interpretation. Furthermore, the results for the interactive voice response, which is primarily an aural experience, tended to follow the telephone results more closely than they did the mail results. Although cognitive explanations for survey mode effects have sometimes mentioned the possible effects of visual layout and design (e.g., see Chapter 10, Tourangeau, Rips and Rasiniski, 2000), explicit research on the mental processes by which respondents to visually and aurally delivered questions arrive at their answers is lacking, and needs to be conducted. 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. If one's intent is to measure change, switching modes could have major 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.

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. 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. The exception is the mail mode, for which use of the telephone only increased response rate by eight percentage points. The high response rate to the mail questionnaire may be attributed to a short survey (18 questions), the presentation in a visually attractive questionnaire whose brevity would be conveyed to the respondent quickly by the presence of all questions on two facing pages, 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.

However, the success of the second mode in reducing nonresponse error, 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 small (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, were an exception with only one significant difference (for the IVR treatment) between respondents and nonrespondents.

Limitations of this study need to be addressed. 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 limited by the use of diverse warranty question formats, e.g., 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 and the list was skewed heavily towards higher household incomes. Although the majority of the sample included requested demographics, missing data was still an issue. Therefore, the results obtained in the current study may not be the same as those that could be obtained from a sample frame that more closely represents the general public.

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, and thus is another limitation of this study.

Further, we examined differences for opinion questions that utilized a format favored by many designers of interview and IVR surveys, i.e., a scalar format with anchored endpoints that surveyors found easier to administer in these types of surveys than were scales with all points anchored. This format was chosen based upon our observation that anchored endpoints are being used more frequently in mail and web surveys. It's possible that the differences noted in this research could be mitigated by labeling all five categories, but doing so presents difficult challenges for telephone interviews where reading all categories is cumbersome, and in IVR surveys, time-consuming. The one fully labeled scale examined in this paper provides only limited insight into the likely effect, because of being used on a question in which the endpoints were used infrequently. It may be possible to strengthen the stimulus of the unlabeled categories by explicitly mentioning, "You may use any of the categories 1, 2, 3, 4, or 5 to answer," or by the addition of other prose. Whether doing so will tend to decrease differences across modes remains a question for future research.

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's 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. But, 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.

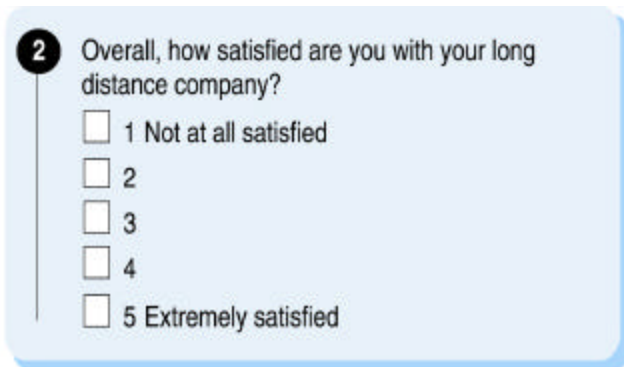
## REFERENCES

- Atrostic, B.K., and Geraldine Burt. 1999. "Household Non-Response: What We have Learned and a Framework for the Future." In Statistical Policy Working Paper 28, Federal Committee on Statistical Methodology, Office of Management and Budget, pp. 153-180.
- de Leeuw, E.D. 1992. *Data quality in mail, telephone, and face-to-face surveys*. Amsterdam: TT Publications.
- Dillman, Don A., Tommy L. Brown, John Carlson, Edwin H. Carpenter, Frederick O. Lorenz, Robert Mason, John Saltiel, and Roberta L. Sangster. 1995. "Effects of category order on answers to mail and telephone surveys." *Rural Sociology*. 60:674-687.
- Dillman, Don A. 2000. *Mail and Internet surveys: The tailored design method*. 2<sup>nd</sup> Edition. John Wiley Co.: New York.
- Dillman, D. A., R. L. Sangster, J. Tarnai, and T. Rockwood. 1996. "Understanding differences in people's answers to telephone and mail surveys." In M. T. Braverman & J. K. Slater (Eds.), *New directions for evaluating series, 70* (Advances in survey research). San Francisco: Jossey-Bass.
- Dillman, D. A., and John Tarnai. 1991. "Mode effects of cognitively-designed recall questions: A comparison of answers to telephone and mail surveys." In P. P. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys*. New York: Wiley.
- Dillman, Don A., John R. Clark, and Kirsten K. West. 1995. "Influence of an invitation to answer by telephone on response to census questionnaires." *Public Opinion Quarterly*. 58:557-568.
- Groves, R. M., and R. L. Kahn. 1979. *Surveys by telephone: A national comparison with personal interviews*. New York: Academic.
- Hochstim, J. R. 1967. "A critical comparison of three strategies of collecting data from households." *Journal of the American Statistical Association*. 62:976-989.
- Krosnick, J., and D. F. Alwin. 1987. "An evaluation of a cognitive theory of response-order effects in survey measurement." *Public Opinion Quarterly*. 51:201-219.
- Lesser, V., D. Dillman, F. O. Lorenz, J. Carlson, and T. L. Brown. August 1999. "The influence of financial incentives on mail questionnaire response rates." Paper presented at the meeting of the Rural Sociological Society, Portland, Oregon.
- Schwarz, N., F. Strack, H. J. Hippler, and G. Bishop. 1991. "The impact of administration mode on response effects in survey measurement." *Applied Cognitive Psychology*. 5:193-212.

- Schuman, H., and S. Presser. 1981. *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic.
- Shettle, C., and G. Mooney. 1999, August. "Evaluation of using monetary incentives in a government survey." *Journal of Official Statistics*.
- Singer, Eleanor. In Press. Chapter 11. "The Use of Incentives to Reduce Nonresponse in Household Surveys." In Groves, R.M., Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, *Survey Nonresponse*. John Wiley Co. New York: NY.
- Srinivasan, R., and S. Hanway. 1999, May. "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, St. Pete Beach, FL.
- Sudman, Seymour, and Norman M. Bradburn. 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Tarnai, John, and Don A. Dillman. 1992. "Questionnaire context as a source of response differences in mail versus telephone surveys." In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research*. New York: Springer-Verlag.
- Tourangeau, Roger, Lance J. Rips and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press: United Kingdom.
- U.S. Department of Commerce Economics and Statistics Administration, National Telecommunications and Information Administration. 2000, October. "Falling through the Net: toward digital inclusion. A report on Americans' access to technology tools." (pp. 33-60).

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

**Mail**

A screenshot of a mail survey question. It features a blue rounded rectangular background. On the left side, there is a vertical line with a black circle containing the number '2' at the top. To the right of this line, the question text reads: 'Overall, how satisfied are you with your long distance company?'. Below the question, there are five radio button options arranged vertically: '1 Not at all satisfied', '2', '3', '4', and '5 Extremely satisfied'.

**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.

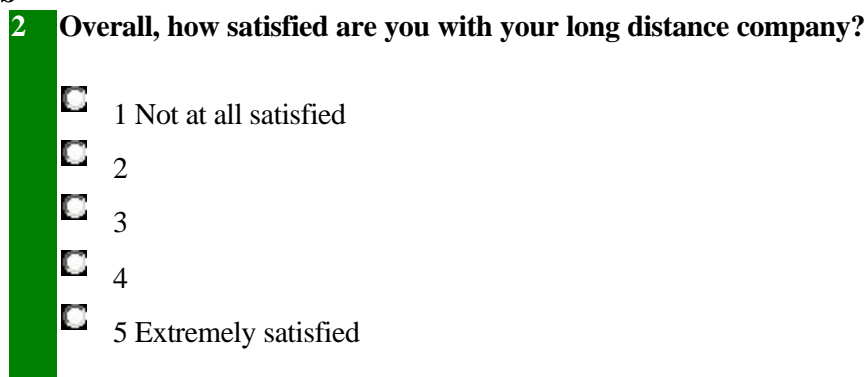
**Telephone, Form B**

(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**

A screenshot of a web survey question. It features a green vertical bar on the left side with the number '2' at the top. To the right of this bar, the question text reads: 'Overall, how satisfied are you with your long distance company?'. Below the question, there are five radio button options arranged vertically: '1 Not at all satisfied', '2', '3', '4', and '5 Extremely satisfied'.

**2** Overall, how satisfied are you with your long distance company?

1 Not at all satisfied

2

3

4

5 Extremely satisfied

Table 1. Response rates\* by treatment groups

Groups	Original sample size	Phase 1		Phase 2		Total		
			n	%	n <sup>a</sup>		%	
Treatment 1	2,000	Mail:	1,499	75.0	Phone:	157	31.7	82.8
Treatment 2 (Form A)**	1,500	Phone:	651	43.4	Mail:	1,094	66.3	80.4
Treatment 3 (Form B)**	1,499	Phone:	667	44.4	Mail:	1,094	66.3	80.4
Treatment 4	2,000	IVR:	569	28.5	Phone:	438	35.9	50.4
Treatment 5	2,000	Web:	253	12.7	Phone:	700	44.9	47.7

\*Response rate=number of completed/number in treatment.

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

<sup>a</sup>Includes non-contacts and refusals to Phase 1.



Table 2. Influence of alternative scale directions on answers to Phase 1 telephone interviews

Variable	Phase 1				One-way analysis of variance		Chi-square		
	Treatment 2		Treatment 3		F ratio	p-value	$\chi^2$	df	p-value
	(Form A)	(Form B)	(Form A)	(Form B)					
<b>Q2 Overall Satisfaction</b>									
1 Not at all satisfied	13	2.0	14	2.1					
2	27	4.2	22	3.3					
3	98	15.2	95	14.4					
4	251	38.9	276	41.7					
5 Extremely satisfied	257	39.8	255	38.5					
<i>Mean</i>		4.10		4.11	0.04	0.85	1.59	4	0.81
<b>Q3 Overall value</b>									
Terrible - 1	18	2.8	15	2.3					
2	36	5.6	34	5.1					
3	140	21.6	164	24.7					
4	277	42.8	281	42.4					
Outstanding - 5	176	27.2	169	25.5					
<i>Mean</i>		3.86		3.84	0.20	0.65	2.20	4	0.70
<b>Q4 Retention</b>		%		%					
1 Not at all likely	32	5.0	22	3.3					
2	31	4.8	35	5.3					
3	98	15.2	104	15.6					
4	161	24.9	170	25.6					
5 Extremely likely	324	50.2	334	50.2					
<i>Mean</i>		4.11		4.14	0.35	0.55	2.39	4	0.66
<b>Q5 Recommend</b>									
1 Not at all likely	79	12.3	61	9.2					
2	50	7.8	47	7.1					
3	123	19.1	151	22.8					
4	159	24.7	170	25.6					
5 Extremely likely	232	36.1	234	35.3					
<i>Mean</i>		3.65		3.71	0.73	0.39	5.3	4	0.25

\*p<.05.

\*\*p<.01.

Table 2. Influence of alternative scale directions on answers to Phase 1 telephone interviews

Variable	Phase 1				One-way analysis of variance		Chi-square		
	Treatment 2		Treatment 3						
	(Form A)	(Form B)	(Form A)	(Form B)	F ratio	p-value	$\chi^2$	df	p-value
<b>Q6 - Local Usage</b>									
1 Not at all likely	57	9.0	56	8.7					
2	12	1.9	18	2.8					
3	36	5.7	31	4.8					
4	18	2.8	26	4.0					
5	138	21.7	124	19.2					
6	45	7.1	47	7.3					
7	71	11.2	76	11.8					
8	92	14.5	99	15.3					
9	32	5.0	31	4.8					
10 Extremely likely	134	21.1	137	21.2					
<i>Mean</i>		6.43		6.46	0.03	0.86	4.27	9	0.90
<b>Q7 Product/Service Quality</b>									
		%		%					
Much Worse Than Expected	9	1.4	8	1.2					
Worse Than Expected	35	5.4	32	4.8					
Just As Expected	433	67.1	423	63.7					
Better Than Expected	135	20.9	164	24.7					
Much Better Than Expected	33	5.1	37	5.6					
<i>Mean</i>		3.23		3.29	2.20	0.14	3.08	4	0.55
<b>Q9 Problem Resolution</b>									
		%		%					
Very Dissatisfied	25	35.2	20	32.8					
Dissatisfied	13	18.3	18	29.5					
Satisfied	21	29.6	14	23.0					
Very Satisfied	6	8.5	7	11.5					
Completely Satisfied	6	8.5	2	3.3					
<i>Mean</i>		2.37		2.23	0.42	0.52	4.11	4	0.39

\*p<.05.

\*\*p<.01.

Table 3. Frequency distribution for all modes in all treatments (n's vary slightly for each item)

Variable	Treatment 1			Treatments 2 & 3			Treatment 4			Treatment 5		
	Phase 1 Mail n=1,499	Phase 2 Phone n=157	Total	Phase 1 Phone n=1,318	Phase 2 Mail n=1,094	Total	Phase 1 IVR n=569	Phase 2 Phone n=438	Total	Phase 1 Web n=253	Phase 2 Phone n=700	Total
	%	%	%	%	%	%	%	%	%	%	%	%
<b>Q2 Overall Satisfaction</b>												
1 Not at all satisfied	2.3	0.6	2.1	2.1	2.9	2.5	3.3	2.1	2.8	2.4	2.0	2.1
2	6.9	2.6	6.5	3.7	7.1	5.3	5.6	2.5	4.3	4.0	4.2	4.1
3	30.9	16.0	29.5	14.8	30.5	21.9	20.8	20.5	20.6	27.3	14.9	18.2
4	38.9	37.2	38.7	40.3	39.4	39.9	31.7	37.7	34.3	39.8	38.4	38.7
5 Extremely satisfied	21.1	43.6	23.2	39.1	20.1	30.5	38.6	37.2	38.0	26.5	40.5	36.8
<i>Mean</i>	3.70	4.21	3.74	4.11	3.67	3.91	3.96	4.06	4.00	3.84	4.11	4.04
<b>Q3 Overall value</b>												
Terrible - 1	2.4	1.9	2.4	2.5	3.1	2.8	3.5	1.8	2.8	2.4	1.9	2.0
2	9.3	4.5	8.8	5.3	11.5	8.1	7.4	3.2	5.6	7.7	4.3	5.2
3	39.3	18.6	37.3	23.2	35.4	28.8	36.8	26.8	32.5	39.1	23.1	27.3
4	37.6	44.2	38.2	42.6	38.6	40.8	31.2	39.2	34.7	41.5	41.7	41.7
Outstanding - 5	11.4	30.8	13.3	26.3	11.4	19.5	21.1	28.9	24.5	9.3	29.0	23.8
<i>Mean</i>	3.46	3.97	3.51	3.85	3.44	3.66	3.59	3.90	3.73	3.48	3.92	3.80
<b>Q4 Retention</b>												
1 Not at all likely	4.7	1.3	4.4	4.1	4.4	4.2	4.0	3.4	3.8	3.2	3.7	3.6
2	8.6	5.8	8.3	5.0	10.1	7.3	7.2	5.1	6.3	9.3	3.3	4.9
3	23.8	16.2	23.1	15.4	27.0	20.7	18.5	13.8	16.5	21.8	15.4	17.0
4	34.1	22.1	33.0	25.2	32.5	28.5	26.8	28.3	27.4	36.7	25.8	28.7
5 Extremely likely	28.8	54.5	31.2	50.2	26.1	39.2	43.5	49.4	46.1	29.0	51.8	45.8
<i>Mean</i>	3.74	4.23	3.78	4.12	3.66	3.91	3.98	4.15	4.06	3.79	4.19	4.08

Table 3. Frequency distribution for all modes in all treatments (n's vary slightly for each item)

Variable	Treatment 1			Treatments 2 & 3			Treatment 4			Treatment 5		
	Phase 1 Mail n=1,499	Phase 2 Phone n=157	Total	Phase 1 Phone n=1,318	Phase 2 Mail n=1,094	Total	Phase 1 IVR n=569	Phase 2 Phone n=438	Total	Phase 1 Web n=253	Phase 2 Phone n=700	Total
	%	%	%	%	%	%	%	%	%	%	%	%
<b>Q5 Recommend</b>												
1 Not at all likely	10.1	9.0	10.0	10.7	11.3	11.0	8.8	10.6	9.6	6.8	7.4	7.2
2	13.5	6.5	12.8	7.4	14.6	10.7	9.7	8.5	9.2	16.5	8.4	10.5
3	30.1	21.9	29.3	21.0	28.0	24.2	28.2	21.9	25.4	30.1	22.2	24.3
4	27.7	24.5	27.4	25.2	28.0	26.4	20.8	24.7	22.5	28.5	23.8	25.1
5 Extremely likely	18.6	38.1	20.5	35.7	18.2	27.7	32.6	34.3	33.3	18.1	38.2	32.9
<i>Mean</i>	3.31	3.76	3.35	3.68	3.27	3.49	3.59	3.64	3.61	3.35	3.77	3.66
<b>Q6 - Local Usage</b>												
1 Not at all likely	9.7	10.4	9.8	8.8	9.3	9.0	8.0	8.2	8.1	10.0	7.2	8.0
2	4.7	5.8	4.8	2.3	4.4	3.3	3.5	4.5	3.9	5.6	2.8	3.6
3	6.2	5.8	6.2	5.2	6.2	5.7	5.1	4.0	4.6	8.0	4.3	5.3
4	5.1	5.2	5.1	3.4	6.1	4.6	4.8	4.2	4.5	6.0	2.9	3.8
5	16.0	14.9	15.9	20.5	17.0	18.9	20.7	19.0	20.0	14.5	22.0	20.0
6	7.4	5.8	7.3	7.2	8.6	7.9	5.8	5.6	5.7	10.4	6.0	7.2
7	12.0	9.1	11.7	11.5	12.1	11.7	7.6	10.6	8.9	10.8	10.3	10.5
8	12.8	15.6	13.1	14.9	11.1	13.2	14.0	14.6	14.2	12.0	17.0	15.6
9	8.5	4.5	8.1	4.9	8.0	6.3	9.0	5.6	7.6	6.0	5.3	5.5
10 Extremely likely	17.5	22.7	18.0	21.2	17.2	19.3	21.6	23.7	22.5	16.5	22.1	20.6
<i>Mean</i>	6.21	6.27	6.22	6.44	6.15	6.31	6.49	6.53	6.51	5.96	6.61	6.43
<b>Q7 Product/Service Quality</b>												
Much Worse Than Expected	0.7	0.6	0.7	1.3	1.2	1.3	4.2	0.5	2.6	0.4	1.3	1.1
Worse Than Expected	9.2	1.9	8.6	5.1	9.2	7.0	7.7	4.6	6.4	5.6	3.9	4.4
Just As Expected	68.1	75.5	68.8	65.4	65.1	65.3	51.4	67.8	58.5	70.7	68.1	68.8
Better Than Expected	18.0	17.4	18.0	22.8	19.6	21.4	22.9	19.5	21.4	19.7	21.1	20.7
Much Better Than Expected	3.9	4.5	3.9	5.3	4.9	5.2	13.7	7.6	11.1	3.6	5.6	5.1
<i>Mean</i>	3.15	3.23	3.16	3.26	3.18	3.22	3.34	3.29	3.32	3.20	3.26	3.24

Table 4. One-way post hoc Bonferroni and Chi-square tests for Phase 1 across modes<sup>a</sup>

Variable	Bonferroni <i>f</i> (1)	Significance Tests					
		Treatment 1 (Mail) vs. Treatments 2 & 3 (Phone) (2)	Treatment 1 (Mail) vs. Treatment 4 (IVR) (3)	Treatment 1 (Mail) vs. Treatment 5 (Web) (4)	Treatments 2 & 3 (Phone) vs. Treatment 4 (IVR) (5)	Treatments 2 & 3 (Phone) vs. Treatment 5 (Web) (6)	Treatment 4 (IVR) vs. Treatment 5 (Web) (7)
		<i>One way p-value</i>	<i>χ<sup>2</sup></i>	<i>f</i>	<i>p-value</i>	<i>χ<sup>2</sup></i>	<i>f</i>
<b>Q2 - Overall Satisfaction</b>	43.78						
<i>One way p-value</i>		.000**	.000**	.178	.020*	.000**	.517
<i>χ<sup>2</sup></i>		167.76	72.02	6.54	22.38	28.84	15.13
<i>f</i>		4	4	4	4	4	4
<i>p-value</i>		.000**	.000**	0.162	.000**	.000**	.004**
<b>Q3 - Overall value</b>	42.193						
<i>One way p-value</i>		.000**	.037*	1.000	.000**	.000**	.658
<i>χ<sup>2</sup></i>		159.98	36.37	2.39	48.57	48.22	20.17
<i>f</i>		4	4	4	4	4	4
<i>p-value</i>		.000**	.000**	0.664	.000**	.000**	.000**
<b>Q4 - Retention</b>	30.28						
<i>One way p-value</i>		.000**	.000**	1.000	.073	.000**	.128
<i>χ<sup>2</sup></i>		138.71	40.56	1.98	9.67	42.16	17.14
<i>f</i>		4	4	4	4	4	4
<i>p-value</i>		.000**	.000**	0.739	.046*	.000**	0.002**
<b>Q5 - Recommend</b>	21.823						
<i>One way p-value</i>		.000**	.000**	1.000	.905	.001**	.069
<i>χ<sup>2</sup></i>		124.42	49.13	3.80	17.09	50.81	25.24
<i>f</i>		4	4	4	4	4	4
<i>p-value</i>		.000**	.000**	0.433	0.002**	.000**	.000**
<b>Q6 - Local Usage<sup>^</sup></b>	3.592						
<i>One way p-value</i>		.195	.265	1.00	1.00	.080	.077
<i>χ<sup>2</sup></i>		44.313	21.655	6.701	21.964	25.358	21.021
<i>f</i>		9	9	9	9	9	9
<i>p-value</i>		.000**	.010*	0.668	.009**	.003**	.013*
<b>Q7 - Product/Service Quality</b>	11.207						
<i>One way p-value</i>		.001**	.000**	1.000	.130	1.000	.076
<i>χ<sup>2</sup></i>		30.51	114.29	4.11	68.15	4.62	38.05
<i>f</i>		4	4	4	4	4	4
<i>p-value</i>		.000**	.000**	0.391	.000**	0.329	.000**

\*p<.05.

\*\*p<.01.

<sup>^</sup>ten-point scale

<sup>a</sup>Means and distributions for statistical tests are reported in Table 3.

Table 5. Statistical test for mode test questions across treatments and phases of data collection

Variable	Mail	Phone	Phone	Phone
	Treatment 1, Phase 1 vs. Treatments 2 & 3, Phase 2	Treatment 1, Phase 2 vs. Treatments 2 & 3, Phase 1	Treatment 4, Phase 2 vs. Treatments 2 & 3, Phase 1	Treatment 5, Phase 2 vs. Treatments 2 & 3, Phase 1
<b>Q2 - Overall satisfaction</b>				
One way <i>p</i> -value	1.000	1.000	1.000	1.000
$\chi^2$	1.471	3.198	8.829	0.866
<i>f</i>	4	4	4	4
<i>p</i> -value	0.832	0.525	0.066	0.929
<b>Q3 - Overall value</b>				
One way <i>p</i> -value	1.000	1.000	1.000	1.000
$\chi^2$	6.748	2.813	7.220	3.023
<i>f</i>	4	4	4	4
<i>p</i> -value	0.15	0.59	0.125	0.554
<b>Q4 - Retention</b>				
One way <i>p</i> -value	0.986	1.000	1.000	1.000
$\chi^2$	6.570	4.177	2.130	3.552
<i>f</i>	4	4	4	4
<i>p</i> -value	0.16	0.383	0.712	0.47
<b>Q5 - Recommend</b>				
One way <i>p</i> -value	1.000	1.000	1.000	1.000
$\chi^2$	2.400	0.859	0.847	7.356
<i>f</i>	4	4	4	4
<i>p</i> -value	0.663	0.93	0.932	0.118
<b>Q6 - Local Usage</b>				
One way <i>p</i> -value	1.000	1.000	1.000	1.000
$\chi^2$	4.533	11.240	9.675	6.261
<i>f</i>	9	9	9	9
<i>p</i> -value	0.871	0.26	0.377	0.714
<b>Q7 - Product/Service Quality</b>				
One way <i>p</i> -value	1.000	1.000	1.000	1.000
$\chi^2$	4.828	7.553	6.964	2.601
<i>f</i>	4	4	4	4
<i>p</i> -value	0.305	0.109	0.138	0.627

\**p*<.05.

\*\**p*<.01.

Table 6. Comparison of self reported demographics for respondents to Phase 1 and Phase 2 for each treatment

Variable	Treatment 1			Treatments 2 & 3			Treatment 4			Treatment 5		
	Phase 1 Mail n=1,499	Phase 2 Phone n=157	Total	Phase 1 Phone n=1,318	Phase 2 Mail n=1,094	Total	Phase 1 IVR n=569	Phase 2 Phone n=438	Total	Phase 1 Web n=253	Phase 2 Phone n=700	Total
<b>Gender</b>												
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
	$\chi^2$	.001 (1)		1.515 (1)			1.302 (1)			15.037**		
<b>Education</b>												
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
	$\chi^2$	2.116 (3)		16.971 (3)**			11.325 (3)**			6.619 (3)		
<b>Age</b>												
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
	$\chi^2$	2.161 (3)		.158 (3)			2.539 (3)			6.784 (3)		
<b>Number in Household</b>												
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 <sup>a</sup>	3.97	3.99	3.97	3.94	3.82	3.89	4.03	3.94	3.99	3.92	3.96	3.95
	$\chi^2$	2.213 (4)		16.714 (4)**			16.733 (4)**			2.660 (4)		
<b>Children in Household (% Yes)</b>												
	88.1	90.9	88.4	89.0	87.8	88.5	88.5	85.7	87.3	87.0	89.3	88.7
	$\chi^2$	1.068 (1)		.921 (1)			1.709 (1)			1.004 (1)		
<b>Monthly Long Distance Bill</b>												
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
	$\chi^2$	2.846 (5)		3.564 (5)			5.475 (5)			4.444 (5)		

<sup>a</sup>Mean value based upon recoded number in household

<sup>b</sup>Based upon respondents with phone interruption (small sample size)

\*p<.05.

\*\*p<.01.

Table 6. Comparison of self reported demographics for respondents to Phase 1 and Phase 2 for each treatment

Variable	Treatment 1			Treatments 2 & 3			Treatment 4			Treatment 5		
	Phase 1 Mail n=1,499	Phase 2 Phone n=157	Total	Phase 1 Phone n=1,318	Phase 2 Mail n=1,094	Total	Phase 1 IVR n=569	Phase 2 Phone n=438	Total	Phase 1 Web n=253	Phase 2 Phone n=700	Total
<b>Income</b>												
Under \$60,000	8.5	10.8	8.7	11.0	9.5	10.3	13.7	9.9	12.1	7.4	9.1	8.7
\$60,000 to less than \$75,000	22.4	16.5	21.9	23.2	19.6	21.5	26.6	21.0	24.3	13.6	19.1	17.6
\$75,000 to less than \$100,000	36.6	30.9	36.1	36.1	37.5	36.8	32.1	35.7	33.6	45.0	39.1	40.8
\$100,000 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
	$\chi^2$											
		7.097 (3)		7.188 (3)			9.092 (3)*			5.097 (3)		
<b>Interruption of phone service (% yes)</b>												
	3.6	5.2	3.8	4.0	5.0	4.5	7.4	4.8	6.3	3.6	3.9	3.8
	$\chi^2$											
		.974 (1)		1.433 (1)			2.830 (1)			.056 (1)		
<b>Without phone service<sup>b</sup></b>												
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
	$\chi^2$											
		.705 (1)		.448 (1)			.002 (1)			1.500 (1)		

<sup>a</sup>Mean value based upon recoded number in household

<sup>b</sup>Based upon respondents with phone interruption (small sample size)

\*p<.05.

\*\*p<.01.



Table 7. Sample frame demographics by response status

Variable	Treatment 1					Treatments 2 & 3					Treatment 4					Treatment 5				
	Sample frame n=2,000	Phase 1 Mail n=1,499	Phase 2 Phone n=157	Phase 1 & 2 n=1,656	Non Resp. n=344	Sample frame n=2,999	Phase 1 Phone n=1,318	Phase 2 Mail n=1,094	Phase 1 & 2 n=2,412	Non Resp. n=587	Sample frame n=2,000	Phase 1 IVR n=569	Phase 2 Phone n=438	Phase 1 & 2 n=1,007	Non Resp. n=993	Sample frame n=2,000	Phase 1 Web n=253	Phase 2 Phone n=700	Phase 1 & 2 n=953	Non Resp. n=1,047
<b>Gender</b>																				
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
<i>Phase 1 vs Phase 2</i> $\chi^2$	.427 (1)*					.258 (1)					.107 (1)					4.650 (1)*				
<i>All Respondents vs. Non Respondents</i> $\chi^2$	8.895 (1)**					12.513 (1)**					8.310 (1)**					1.504 (1)				
<b>Education</b>																				
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 Coll. Deg.	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
Grad Deg/Post Grad.	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.	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
<i>Phase 1 vs Phase 2</i> $\chi^2$	1.564 (3)					.6110 (3)					.803 (3)					7.797 (3)*				
<i>All Respondents vs. Non Respondents</i> $\chi^2$	19,288 (3)**					30.712 (3)**					11.402 (3)**					8.978 (3)*				
<b>Age</b>																				
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
<i>Phase 1 vs Phase 2</i> $\chi^2$	6.559 (3)					8.405 (3)*					2.854 (3)					4.595 (3)				
<i>All Respondents vs. Non Respondents</i> $\chi^2$	10.669 (3)*					21.731 (3)**					.174 (3)					10.489 (3)*				
<b>Income</b>																				
\$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
<i>Phase 1 vs Phase 2</i> $\chi^2$	4.069 (2)					3.531 (2)					4.318 (2)					4.160 (2)				
<i>All Respondents vs. Non Respondents</i> $\chi^2$	.442 (2)					.815 (2)					8.530 (2)*					.563 (2)				

\*p<.05.

\*\*p<.01.