Survey Mode as a Source of Instability in Responses across Surveys[1]

By

Don A. Dillman and Leah Melani Christian

Washington State University

Introduction

Most panel studies require measurement of the same variables at different times. Often, participants are asked questions, several days, weeks, months or years apart in order to measure change in some characteristics of interest to the investigation. These characteristics might include political attitudes, satisfaction with a healthcare provider, frequency of a behavior, ownership of financial resources, or level of educational attainment. Whatever the characteristic of interest, it is important that the question used to ascertain it perform the same across multiple data collections.

Considerable evidence now exists that the choice of survey mode affects respondent answers to survey questions that are *worded* the same (e.g. de Leeuw and Van Der Zowen, 1988; Fowler, Roman and Di, 1998; Dillman, Sangster, Tarnai and Rockwood, 1996). This means that differences in answers to some survey questions between Time 1 and Time 2 may be the result of mode change rather than any actual difference in the behavior or opinion of interest. In addition, Internet and Interactive Voice Response modes of collecting data are now taking their places beside the traditional alternatives of mail, telephone, face-to-face, and group self-administration alternatives. Both of these new methodologies introduce the possibility of mode effects that are not yet well understood.

---

[1] Revision of paper presented at the Workshop on Stability of Methods for Collecting, Analyzing and Managing Panel Data, American Academy of Arts and Sciences, Cambridge, MA, March 26-28, 2003. Don A. Dillman (dillman@wsu.edu) is The Thomas S. Foley Distinguished Professor of Government and Public Policy, Regents' Professor in the Departments of Sociology and Community and Rural Sociology and serves as Deputy Director for Research of the Social and Economic Sciences Research Center (SESRC), all at Washington State University. Leah Melani Christian is Graduate Research Assistant in the SESRC and Department of Sociology, Pullman, Washington 99164-4014.

Our purpose in this paper is to discuss why mode changes are increasingly likely to occur in panel studies and the consequences. The nature of likely mode differences and the reasons for their occurrence are reviewed and suggestions made for how differences in answers between modes might be reduced. In addition evidence is presented that suggests within mode differences may result from variations in the visual layout of questions used in both web and mail surveys. Such changes may have a secondary impact on comparisons between visual and aural modes of surveying.

Why Mode of Data Collection Often Changes

Mode change happens frequently in panel studies after the initial wave of data collection. Decisions to change mode are often made with little consideration of whether such changes might influence answers to the survey questions. Several reasons may be offered as explanations for the desire to switch modes for subsequent data collections.

First, the locations of panel respondents often change between data collections. Some panel surveys extend over many years, or even decades. For example, students who initially responded to a paper survey in a classroom, cannot easily be reassembled several years later for follow up contacts. Thus, the decision may be made to survey those respondents by telephone at their new locations. Whereas group-administration may be the most cost-effective for the initial wave of data collection, the telephone is likely to be far more efficient for the follow-up.

In addition, the introduction of new modes creates new possibilities for conducting surveys. Specifically the emergence in the 1990's of Internet and Interactive Voice Response (voice or touchtone data entry by telephone in response to prerecorded interview questions and instructions), provided additional ways of contacting and surveying respondents. In many cases, these modes offer substantial cost and speed advantages and were not available at the time of initial data collection.

Further, when follow-up studies are done, researchers may have either telephone contact or mailing address information, but not both, so that the method of follow-up is influenced by the kind of contact information available. Traditional contact information, i.e. postal addresses and home telephones, has now expanded to include email addresses and cell phone numbers.

In recent years response rates to traditional modes of surveying, in particular the telephone, appear to be declining. As a result, surveyors have become interested in providing panel members choices so respondents can decide whether to respond by one mode or another. This possibility is encouraged by technological advances, which allow efficient transfer of questionnaire word files from one mode to another. In addition, the particularly low marginal costs of collecting additional questionnaire responses over the Internet, are encouraging surveyors to collect as many responses in this manner as possible, saving telephone interviews for those who do not respond.

The mode choice for follow-up surveys can often be less restrictive. The coverage and/or sampling procedure constraints necessary for initial data collection may determine the mode of surveying. Household personal interview methods are often used for initial data collection because it is necessary to identify truly representative samples of general public populations. For example, the monthly U.S. Current Population Survey, which is relied on to estimate nationwide unemployment rates, depends upon the use of face-to-face interviews to contact households and conduct the initial interviews. This survey then switches to telephone interviews for households with telephones to reduce costs.

In addition, the influence of the personal preferences of researchers on choosing the follow-up survey mode often makes a difference. Researchers may have little experience with some modes resulting in their comfort level being higher for some survey modes than for others. Also, some organizations have a tradition of doing telephone surveys, whereas other firms have a tradition of working with self-administered methods, and choose modes based upon their skills and infrastructure. Firm specialization is increasing as new firms emerge that only do Internet or IVR surveys. Some researchers, who desire

to stay on the cutting-edge methodologically, may also choose a new mode in hopes of making a methodological contribution to panel research methods.

In recent years another kind of mode change has emerged, which may create even more complex mode comparison issues. Instead of only changing modes between data collections, surveyors sometimes design individual data collection efforts that use one mode for some respondents and others for the remaining respondents. An example of this strategy is the National Survey of Earned Graduates for which data collection procedures were tested in 1992 (Shettle and Mooney, 1999). It began with mail, switched to telephone, and finally relied on face-to-face interviews, each of which added a significant amount to response rates. Modes were similarly mixed in later data collections. Thus, not only do some respondents answer by different modes in the initial data collection period than in other periods. The potential also exists for respondents to answer using a variety of modes during later data collections.

Intra-mode differences within data collection periods may be increasing (Dillman, 2002). One reason is the concern that the response rates are lower for some modes than in the past, and offering a second mode may help improve them. Another contributor is the widely different costs associated with modes of data collection. For example, if respondents can now be induced to go directly to the Internet, the marginal costs for each additional interview completed in that way may be far less than for any of the other methods. Whereas the costs of providing the opportunity to respond by more than one mode was once substantial, computer software now makes such efforts much less difficult and costly.

Thus, a variety of reasons, ranging from personal preferences and the changing nature of the available mode choices to cost considerations, may influence decisions by investigators to switch modes either between or within data collections periods. In addition, changing modes has often been viewed as a minor consideration for deciding how to collect data for panel studies. However, this position needs to be reevaluated. In

some instances changing modes may produce results that provide a false indication of change having occurred between waves of data collection.

How Survey Mode Affects the Wording of Questions

Surprisingly, perhaps one of the main reasons for differences in answers to questions when the survey mode is changed, is that each mode encourages surveyors to ask questions in different ways. There are many different ways in which question wording is shaped by the various survey modes.

Face-to-face interviews

The face-to-face interview is the most flexible of survey modes. It can accommodate complexity better than any other method, making full use of both visual and aural communication. Researchers who rely on face-to-face procedures are able to use longer more complex scales, often with word labels for each of the categories or scale points. To accommodate long questions with many answer choices, show cards are typically used where respondents are handed a card with answer choices and asked to choose a category after the interviewer reads the questions.

Face-to-face interviews also encourage the use of longer questionnaires because once begun, it is often difficult for respondents to terminate the interview. The use of open-ended questions for which interviewers can follow-up with additional probes in order to gather more complete and interpretable answers is also encouraged for this survey mode. The presence of an interviewer who can read body language as well as respond to silences and incomplete answers makes it possible to obtain long, complete answers.

Telephone Interviews

The advent of telephone interviewing in the 1970's resulted in significant changes in how researchers composed questions. Scales were shortened, and frequently the wording was

limited to anchored endpoints rather than providing a word description for each category, in order to simplify the interview task. Thus, a question that might have been asked in a face-to-face interview using a fully labeled scale with response choices of not at all satisfied, slightly satisfied, somewhat satisfied, mostly satisfied and completely satisfied, would be changed to a 1 to 5 scale with, "the endpoints of 1 being not at all satisfied and 5 being completed satisfied and you can also use any number in between." (Dillman, Phelps, Tortora, Swift, Kohrell, and Berck, 2001).

Complex questions with longer scales were sometimes divided into steps (Miller, 1984). For example, first asking people if they were satisfied, dissatisfied or neither satisfied nor dissatisfied. Then following-up with a probe that asked satisfied and dissatisfied respondents whether they were completely, mostly or slightly satisfied (or when appropriate, dissatisfied), and asking the neutral respondents whether they would best be described as completely, mostly, or slightly satisfied. Similarly, respondents might be asked whether their income was above or below a given amount, and then asked repeatedly with different amounts to allow the income level to be measured with reasonable precision.

The need to simplify questions for the telephone also led to an increase in the use of screen questions from which the respondent branched different ways depending upon their answer. This procedure allowed less information to be communicated in each interviewer utterance, and the respondent was unaware of the extensive and complex branching patterns that resulted. When telephone questionnaire designers encounter questions with many words and/or answer choices, their frequent response is to break them into smaller parts so respondents can skip past unnecessary information. The compelling reason for question simplification on the telephone is the reliance entirely on aural communication.

The use of telephone interviews also places pressure on researchers to shorten their interviews. Twenty minutes is deemed a "long interview" for most telephone studies, but

not for face-to-face interviews. For example, one national survey organization refuses to conduct telephone surveys that are longer than 18 minutes in length.

Telephone interviews are also quite similar to face-to-face interviews in certain respects. One important similarity is that both interview methods rely on aural communication. Open-ended questions can be asked in telephone interviews, just as in face-to-face interviews, with follow-up probes being used to get interpretable answers. Both the telephone and face-to-face methods have also tended to "hide" the possibility of certain answers, (e.g. "no opinion" or "don't know" categories) from respondents but instead allowing the interviewer to record such answers only if they were offered by the respondent. Refusals to answer may also be recorded as a separate category.

Interactive Voice Response Surveys

IVR surveys may use voice technology to recognize spoken answers or require the respondent to enter data using numbers and/or symbols on the telephone's touchtone keypad. (Dillman, 2000; Toureangeau, Steiger and Wilson, 2002). This mode is used with increasing frequency for very brief routinized surveys where respondents are highly motivated to respond, e.g. certain employee and customer satisfaction surveys. It has even been tested for possible use as an alternative to the 2010 Decennial Census mail-back form (Stapleton, Norris, Brady, 2003).

The early use of IVR surveys has witnessed efforts to make the wording brief, even shorter than that used in regular telephone surveys. Researchers also seem to be attempting to reduce the lengths of scales, and keep questionnaires extremely short. In addition prerecorded instructions on what the respondent should do if he or she wants to have an item reread, and to solve other potential problems, must be developed and a protocol developed for their use.

Mail Surveys

The use of mail surveys has encouraged respondents to write questions in ways that differ from the construction practices for both face-to-face and telephone interviews (Dillman, 1978). Many users try to avoid as many branching questions as possible because respondents are prone to making errors following branching instructions (Redline, Dillman, Dajani and Scaggs, In Press). Thus, instead of asking a series of questions, such as "Do you own or rent a home," followed by, "Do you have a mortgage on this home or is it owned free and clear," a mail survey question is likely to ask the question in this manner: Is the home in which you live: 1) owned free and clear, 2) owned with a mortgage, 3) rented or 4) occupied under some other arrangement."

The self-administered nature of mail surveys also encourage researchers to use check-all-that-apply questions, rather than asking for a yes or no answer to each of many items on a list as done for interview surveys. Check-all formats are likely to produce fewer affirmative answers than are the yes/no formats (Rasinski, Mingay, and Bradburn 1994).

Mail surveys also remove the pressures to shorten questions that are common in telephone surveying. Fully-labeled scales with many categories, similar to those used in face-to-face interviews and displayed as show cards, can be used with equal ease in self-administered mail surveys.

In addition, users of mail surveys are often discouraged from asking open-ended questions because of the inability to encourage more complete answers by probing or asking additional follow-up questions. This concern often leads to breaking individual open-ended questions into multiple parts. For example, an interviewer can ask for one's occupation and then probe for enough details to allow for coding. On mail questionnaires, designers may feel compelled to ask a series of questions (e.g., What is your occupation? What kind of work do you do? Is there a minimal educational

requirement? What is the name of the company for which you work?). Thus, the mail version tends toward explicitly seeking information that may or may not have been offered as responses in the open-ended version.

Similarly, the use of mail questionnaires also requires that researchers decide what to do with the typically "hidden" interview categories, i.e. REFUSED, DON'T KNOW, NO OPINION, or DOES NOT APPLY that are available for use by telephone interview, but are not offered to the respondent as part of the question stimulus. Frequently, such categories are left off the mail questionnaire resulting in respondents skipping the question or leaving it blank in the self-administered survey mode. Alternatively, they may be explicitly listed. Either method changes the stimulus from that is typically provided in interviews.

The reliance of mail surveys on visual communication provides a means by which respondents can interpret context, and review previously read questions while formulating an answer to later questions. Thus, in certain ways self-administered questionnaires make possible the communication of greater complexity than telephone surveys. At the same time, order effects frequently observed in telephone surveys may be reduced (Bishop, Hippler, Schwarz and Strack, 1988).

Internet Surveys

Internet surveys also rely on the use of visual instead of aural communication used in interview surveys. Some similarities exist in how questions are constructed for mail and Internet, but important differences also exist. One difference is that Internet surveys allow the use of audio, video, color and other dynamic features, which are generally not used in other types of surveys.

Internet surveys can be programmed using a page-by-page construction whereby each question is displayed either on a different screen or in one display so that respondents can scroll through all of the questions at once. One of the potential difficulties of using

separate screens for each question is a loss of context that is provided in mail surveys because respondents cannot review previous answers. Although it may be deemed undesirable for respondents to be able to see (or go back and change) answers to previous questions, problems may also be generated from respondents losing their mental place in a questionnaire (e.g. a series of repetitive questions about each job they have held since graduating from college).

Check-all-that-apply questions are used even more frequently on Internet than mail surveys. Because of html programming used to construct Internet surveys, two types of answer category formats are specified: radio buttons when only one choice can be marked, and html boxes that allow more than one answer to be checked. Check-all-that-apply formats are awkward to ask in an interview and as noted earlier are not used in telephone surveys. Further, Internet surveys also allow the use of longer scales with full labeling as normally done in mail and face-to-face surveys.

Branching instructions do not present a significant problem for web surveys inasmuch as the survey can be designed in ways that make branching transparent to respondents, much as it is on interview surveys. Internet surveys are also similar to telephone surveys in that longer surveys are typically discouraged and not likely to be tolerated by many respondents. Another similarity to telephone interviews is that item non-response, which represents a major concern in mail surveys, can be overcome by programming that requires each question to be answered. However, a consequence of this programming may be to produce respondent frustration and fewer completions.

Why Researchers Often Change Question Structures for Follow-up Measurement By A Different Mode

It is desirable to keep question wording the same across modes. Yet, changes often get made. Some of them are quite simple, and even accidental. For example, an open-ended question asked in a telephone survey as "What is

your current marital status?" was changed by adding these choices for the web survey: "single, married, separated, divorced, and widowed." The result was to decrease the number of single and married respondents while increasing the percent that were separated, divorced and widowed. The simple explanation is that some respondents may fit into more than one category, but offer the simpler single or married choice unless the more specific choices are explicitly offered (Tortora, 2002).

Changes from yes/no formats used in interviews to check-all that apply formats used in self-administered formats, have also produced quite different response distributions. Web respondents consistently mark more answers affirmatively when asked to give a response to each item, rather than only checking those that apply to them (Dillman, Smyth, Christian and Stern, 2003).

Survey designers do not as a rule deliberately change questions between rounds of data collection in panel studies. However, implementers often taken it upon themselves to "adjust" question structure believing a particular question format will work better for the new survey mode.

These changes happen because a "preferred" mode of constructing survey questions has evolved for each of the different modes, and tends to be shared across organizations. Thus, when questions are designed, especially in large organizations where there are specialized units for telephone, face-to-face, web, mail and IVR data collection, each unit may have developed a style of question design that they feel is best for their mode. Thus, telephone questionnaire designers are especially sensitive to the problems of length and verbal comprehension. Mail questionnaire designers are especially sensitive to the problems with getting respondents to follow

branching instructions. And, those who implement particular modes tend to want to avoid situations in which their mode does not perform well.

When designing questions for single mode studies, possible differences across mode are rarely discussed. And, surveyors who have learned over the years to design for one mode may find it difficult to accept the need to optimize across modes, rather than to maximize efficacy for their own mode. The argument that, "this" is how telephone (or mail or face-to-face) questions are usually written," often becomes a powerful force when designing survey instruments, and sponsors of panel surveys are often unaware of these consequences.

Why Identically Worded Questions Often Produce Different Answers Across Modes

Even if one succeeds in maintaining the same questionnaire stimulus across survey modes, differences in answers may occur. However, mode differences for identically worded questions are more likely to occur for some types of questions than others. In addition, there are a number of likely causes for these mode differences.

<u>Some questions are more subject to instability across modes than are others.</u>

Schwarz (1996) argues that the survey process is governed by the conduct of conversation that is used in everyday life and as such respondents act as cooperative communicators trying to make sense of the information provided to them. The interview or questionnaire exhibits many of the same characteristics as a conversation in which the meaning assigned to later items reflects earlier information being transmitted between the respondent and questionnaire sponsor (Schwarz, 1996). Thus, respondents actively try to make sense of the survey questions by drawing on all the information the researcher provides, including the context of the survey. Respondents will systematically make use of context information on some questions more so than others.

When respondents are asked an opinion question, “How satisfied are you with your choice of careers?” they are being asked a question for which a ready-made answer may not exist. In addition, surveyors are likely to offer a set of vaguely quantified response categories from which to choose (e.g. “Completely satisfied, Mostly Satisfied, Fairly Satisfied, Slightly satisfied, Not satisfied”). In order to answer these questions, the respondent must think about how to respond to the question including trying to figure out the meaning of the response categories to determine the answer which comes closest to expressing their opinion. Thus, the stem of the question and the response categories are important sources of information for respondents, especially when they are answering questions to which they may have not previously formed an answer.

Additionally, in ordinal scales where respondents must choose where they fit along an implied continuum, the response categories become very important cues in helping respondents decide how to answer the question. In these types of scalar questions, the meaning of the individual categories (e.g. “fairly satisfied) comes in part from the order in which the choices are displayed or read to the respondent. Thus, respondents draw on information provided in the description of the category (verbal or numeric labels) and its position in relation to other categories.

Context and Order Effects

Since respondents are actively making sense of the survey or interview, they may also be affected by the order in which questions are presented. In panel studies, new questions are often added to follow up surveys to ask respondents about change or differences from the last time they were contacted. Whether a question is located after certain other questions and before others, can affect how respondents answer. Depending on the situation, respondents may decide to include or exclude information when answering a question based on what information they have already provided to the researcher in previous questions (Schwarz 1996). If the previous question has asked about satisfaction with something related to the city or state in which one lives, one’s profession, or even one’s marriage, there may be some carryover effects that influence one’s answers

(Schwarz 1996, Schwarz, Strack, and Mai 1991). Thus, questions that use vague quantifiers and require the formulation of an answer on the spot seem more likely to be influenced by context and mode effects.

On the other hand, when respondents are asked questions that they know precisely, such as their age or date of birth, they have already formulated answers to these types of questions and do not need the categories to help formulate their answer. If categories are provided, the respondents need only to select the category that corresponds to their preformulated answer. For the most part we do not expect questions on age, gender, whether a home is rented or owned, or others for which respondents carry-around obvious answers that they are willing to reveal to others, to produce significant changes when asked by a different survey mode.

Other Contributors to Mode Differences

Researchers have identified many examples of mode differences, as well as many reasons for their occurrence. We provide a brief review of some of the most frequently observed differences, i.e. social desirability, acquiescence and primacy/recency effects.

Social desirability is the tendency for people to offer responses that are based upon the perceived cultural expectations of others (DeLeeuw and van der Zowen, 1990). Such expectations are more likely to influence answers in interview surveys than in those that use self-administration because of the presence of an interviewer. One example is the frequently observed tendency for respondents to give more positive answers when asked to rate their health. In the classic study on this topic, Hochstim (1967) found that 43% of respondents to a face-to-face interview ranked their health as excellent, whereas 38% did so over the telephone, and only 30% chose that category when asked using a mail survey. A study by Biemer (1992) found that 15% of a national sample of people over 65 ranked their health as very good (the top choice) on a self-administered questionnaire, but 27% did so in a follow-up face-to-face interview a month later in which an identically worded question was asked.

Since it is a cultural expectation that one offers a casual "fine" or other positive answer when asked how one is doing, it should not come as a surprise that respondents in an interview survey give somewhat more positive answers than they do in a self-administered questionnaire. Similarly, researchers have shown that respondents to interview surveys are less likely to admit to drug use or other socially undesirable behaviors (DeLeeuw, 1991, Fowler, Roman, and Di,1998).

Acquiescence is the tendency to agree when interacting with another person because in general, people find it easier to agree than disagree. A limited amount of research has suggested that people are more likely to agree with questionnaire items, for example, "Well groomed individuals make better leaders" in interview surveys. To avoid possible acquiescence, and thus mode differences, it has been suggested that questions not be formulated using agree/disagree categories (Schuman and Presser, 1981).

Recency is the tendency for respondents to interview surveys to choose from the last offered answer choices, while *primacy* is the tendency for respondents to self-administered surveys to choose from among the first offered categories. The primacy/recency phenomenon has been noted by a number of authors, notably Krosnick and Alwin (1987). Recency effects were found in telephone surveys conducted by Schuman and Presser (1981). Evidence supporting the existence of primacy/recency effects has been presented by other authors as well, but their results have not always been consistent. In 84 experiments reported by Dillman et al. (1995) that used a variety of question structures, evidence for the existence of primacy on mail surveys and recency on telephone surveys was quite inconsistent. Nonetheless differences attributed by authors to primacy or recency have frequently been observed across survey modes.

Although the last word has certainly not been written about mode differences, it is apparent that they occur frequently, and the reasons for their occurrences also differ. Much remains to be learned about how mode influences answers to survey questions.

In Search of Solutions

Just as mode differences do not flow from a single problem, there is no single solution that will facilitate switching modes when conducting panel studies. Rather, there are a set of issues that need to be addressed, some simple, and some of which are much more complex and in need of careful research.

Unimode Construction and it's Limitations

One way of attempting to limit mode effects is to attempt to write survey questions in a manner that will work satisfactorily across different survey modes, and to make sure that questions remain exactly the same in later data collections, regardless of modes. Procedures aimed at accomplishing the same stimuli across modes have been described elsewhere as unimode construction (Dillman, 2000, chapter 6). An example of unimode construction is to avoid using "Don't Know" as an available but unread category in the telephone mode, while preventing it's use altogether in mail or web surveys. Instead one could make it explicitly available in all modes. Unimode construction also means resisting the temptation to change from an open-ended question in an interview to a close- ended question for a mail follow-up, simply because open-ended questions do not perform as well in such surveys. And, if one uses Yes/No formats for interviews one needs to resist changing to a check-all-that-apply format for paper or web self-administered surveys. Further, if one asks occupation through a series of mail questions then one needs to keep the same format for telephone.

Unimode construction requires conductors of surveys to use question styles that are different from those generally accepted for the various modes. For example, if one wants to maintain the same stimulus on telephone and face-to-face, that suggests forgoing the use of "show cards" that are often

considered standard procedure for face-to-face interviews. Accomplishing unimode construction suggests the need to optimize questionnaire design across modes, rather than to maximize for individual modes.

However, unimode construction is not a solution to some of the issues noted earlier, notably social desirability and acquiescence. These effects stem from the presence of an interviewer, and despite common wording telephone and face-to-face interviews are likely to produce different answers to such questions than do self-administered questionnaires. In addition the fact that recency effects have been associated with the verbal presentation of answer categories, while primacy effects are associated with the visual presentation of answers, presents a dilemma to questionnaire designers. It has sometimes been proposed that the answer categories be rotated randomly in order to deal with potential effects, and may be appropriate for some cross-sectional surveys. However, when applied to a panel surveys it raises the need for assuring that individual respondents receive the same category orders across data collections so that individual measures of change are not confounded by such procedures.

The adoption of unimode construction procedures suggest that consideration be given to such practices as reducing the length of scales and perhaps avoiding the use of a word categories for each scale point. Switching to scales having word labels only for the endpoints would seem to facilitate their administration over the telephone, where interviewers often struggle to read scale points consistently. However, recent research on the effects of visual design and layout reveals that such attempts at finding commonality in construction across modes may be more difficult than it seems.

How Visual Design and Layout Affect Respondent Answers

It has long been recognized that question meaning in interview surveys does not come from words alone; voice and interviewer effects act as a paralanguage that give additional meaning to questions. These influences underlie the occurrence of social desirability and acquiescence effects. Evidence is now accumulating that in self-administered questionnaires, which depend upon visual as opposed to aural processing by the respondent, may also be subject to paralanguage affects.

Considerable research has suggested that when surveys are administered visually, respondents draw information from more than words. They may also draw information when determining question meaning from visual appearance and layout. (Smith 1993, Schwarz 1996). It has been proposed that numbers, symbols, and graphical features (such as variations in size, spacing, shapes, and brightness) act as paralanguage that gives additional meaning to words (Redline and Dillman, 2002).

Recent experiments confirm that these paralanguages do influence how respondents interpret the meaning of questions and instructions on questionnaires in a wide variety of situations. Redline, et al. (In Press) found in a large scale national experiment included in the 2000 Decennial Census that respondent errors in following branching instructions, which directed them to skip ahead, could be reduced by one-third, from nearly 20% to 13%. This was accomplished through the joint use of additional symbols (arrows), changes in the graphical appearance of the skip instruction (larger, darker font), and an additional word instruction placed at the beginning of the question to be skipped. The design of that experiment did not allow the individual effects of each change to be measured, but did show that combined changes could produce a large reduction in errors.

Another set of experiments has demonstrated that independent manipulations of symbols and graphics can change respondent answers to paper questionnaires (Christian and Dillman, In press). For example, changing the layout of scales from a vertical linear display of categories to arranging the same categories so that the first two were in one column, the next two in a column to the right, and the fifth to the right of that column, significant affected the response distributions. In addition, changing the distance between

answer spaces, placement of a large arrow (vs. no arrow) that would direct respondents to a subordinate question, and changes in the size of open-ended answer spaces, all significantly influenced people's answers.

An additional experiment in this study revealed the considerable dependence of respondents on graphical layout for providing meaning to questions and raises additional concerns about possible mode differences. Respondents were asked to answer a five point scale with polar point labels that was presented in one of two formats: A) a linear layout with numbers and boxes for each category and B) a layout that conveyed information about the scale in the query, but required respondents to put the number of their choice in a box (Figure 1). The latter version removed the graphical and symbolic languages that may have helped communicate the continuum of the scale. Respondents to the number box version gave consistently more negative answers than did respondents to the linearly displayed check box version. These results are summarized here in Table 1.A. Examination of completed questionnaires revealed numerous erasures and answer changes from 1 for 5 and 2 for 4 or vice versa, suggesting that respondents had become confused when interpreting the direction of the number box scale, because of the removal of support from the other languages (Christian and Dillman, In Press).

A subsequent study repeated this experiment in a web survey using the same three questions (Christian, 2003). The test was also expanded to include a treatment in which the same questions were asked using the same categories with all five scale points labeled. As shown in Table 1B, the polar point vs. number box comparison results are quite similar to those found for the previous paper experiment. For all three questions in both experiments the number box produces more negative answers than did the polar point version. The fully labeled versions produced even more positive answers for two of the three items than did the polar point scale (Table 1C).

These differences suggest clearly that variations in how a five point scale is supported by words, numbers, symbols and graphics can influence answers to scalar satisfaction questions that use visual communication, and these features must be taken into account

when using the same visual mode across data collections. Removing word support (shifting to a polar point scale) or graphical and symbolic support (shifting to a number box) results in different answers for self-administered visual surveys. These results raise the question of which of these formats may translate most effectively across visual and aural survey modes.

An earlier experiment conducted by the Gallup Organization compared answers to similar five-point satisfaction scales using polar point word labels for respondents surveyed by telephone, IVR, web and mail (Dillman, Phelps, Tortora, Swift, Kohrell, and Berck, 2001). In this survey, respondents to the aural modes (telephone and IVR) provided significantly more positive answers, as reflected in greater use of the extreme positive category, than did respondents to the visual modes (web and mail). These results provide evidence that the polar point scale did not translate equivalently between visual and aural presentation. Whether better equivalency might be achieved between fully labeled questions, remains to be seen.

Summary and Conclusions

Mode changes between initial data collection and follow-up data collections for panel studies should be evaluated carefully. There is a considerable likelihood that such changes make it difficult to accurately measure change between survey waves. Nonetheless, mode changes are common in panel studies, and are the result of many considerations from cost and budget concerns to personal and organizational preferences.

Multiple factors may contribute to different measurements being obtained solely as a result of mode change. One of the most frequent causes of change are inadvertent revisions in the structure of questions to follow suggested design procedures for one mode. Each mode favors the construction of questions in ways that differ from those favored by each of the other modes. When, data collection for panel studies is left to technical staff associated with each mode, it is likely that question structures will be

changed. In addition, the insertion of new questions for follow-up studies between questions from previous waves may produce unintended order or context effects.

A third set of factors that appear to contribute to measurement differences across modes—social desirability, acquiescence and primacy/recency—present particularly difficult challenges because it may not be possible to completely eliminate these differences, leaving adjustment or indexing as the only solution.

The construction of questions to provide a more equivalent stimulus across modes, i.e. unimode construction seems desirable. However, to do this will likely require writing questions in ways that optimize their use for all modes rather than allowing what is best (or traditional) for one mode to be forced into other modes. This may be quite difficult to accomplish in our complex survey world in which many different modes exist and are preferred by different organizations that use some modes but not others.

Recent experimental evidence that the visual presentation of questions can influence significantly people's answers raises important questions that remain to be answered. The results reported here, showing how three different versions of five-point satisfaction questions—polar point labels on a linear scale layout, a number box without visual layout support for answer categories, and a scale with word labels for each category—produced significantly different answers for each of the layouts. These results suggest that differences in layout within visual modes must also be considered a possible source of measurement error and that research is needed on how various visual representations compare when used in aural modes of surveying.

The challenges presented to survey researchers because of the variety of modes from which to choose are substantial. However, the theoretical and empirical work that has been performed thus far, provide a rich source of ideas for finding appropriate solutions. The mistake would be to ignore mode, and now visual design effects, as if they did not really matter.

REFERENCES

Biemer, Paul. 1998. Personal Communication.

Bishop, G.H., Hippler, H-J, Schwarz, N., and Strack, F. 1988. A Comparison of Response Effects in Self-Administered and Telephone Surveys. In *Telephone Survey Methodology* ed by R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, II and J. Waksberg, pp. 321-340. Wiley, New York.

Christian, Leah Melani, 2003. *The Influence of Visual Layout on Scalar Questions in Web Surveys.* Unpublished Master of Arts Thesis. Washington State University: Pullman

Christian, Leah Melani and Dillman, D.A. In Press. The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Surveys. *Public Opinion Quarterly* Spring 2004.

De Leeuw, Edith, and Van der Zowen, H. 1988. Data Quality in Telephone and Face-to-Face Surveys: A Comparative Analysis. In *Telephone Survey Methodology* ed by R. M. Groves, P. Biemer, L. Lyberg, J. T. Massey, W. L. Nicholls, II, and J. Waksberg, pp. 283-99. Wiley-Interscience: New York.

Dillman, D. A. 1978. *Mail and Telephone Surveys: The Total Design Method.* John Wiley, New York.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. John Wiley, New York.

Dillman, D. A. 2002. Navigating the Rapids of Change: Some Observations on Survey Methodology in the Early Twenty-First Century. *Public Opinion Quarterly*, 66: 473-494.

Dillman D. A., Brown, T. L., Carlson, J., Carpenter, E. H., Lorenz, F. O., Mason, R., Saltiel, J., and Sangster, R. L. 1995. Effects of Category Order on Answers to Mail and Telephone Surveys. *Rural Sociology* 60: 674-687.

Dillman, D. a. R. Sangster, John Tarnai and T. Rockwood, 1996. Understanding Differenes in People's Answers to Telephone and Mail Surveys." In *Current Issues in Survey Research: New Directions for Program Evaluation Series*, Chapter 4: 45-62. ed by Braverman, M. T. and Slater, J. K. San Francisco: Jossey-Bass.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J. and Berck, J. 2001. Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, interactive voice Response and the Internet. Unpublished paper presented at Annual Meeting of American Association for Public Opinion Research; Montreal. May 18th.

Dillman, D. A., Smyth, J., Christian, L. M. and Stern, M. 2003. Multiple-answer Questions in Self-Administered Surveys: The Use of Check-All-That-Apply and Forced-Choice Question Formats. Unpublished paper presented at 2003 Annual Meeting of the American Statistical Association, San Francisco, CA August 2003.

Fowler, F. J., Jr., Roman, A. M., and Di, Z. X. 1998. Mode Effects in a Survey of Medicare Prostate Surgery Patients. *Public Opinion Quarterly*, 62: 29-46.

Hochstim, J. 1967. A critical comparison of thee strategies of collecting data from households. *Journal of the American Statistical Association*, 62: 976-989

Krosnick, J., and Alwin, D. F. 1987. An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly* 51: 201-219.

Miller, Peter. 1984. Alternative Question Forms for Attitude Scale Questions in Telephone Interviews. *Public Opinion Quarterly* 48: 766-778.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. 1994. "Do Respondents Really 'Mark All That Apply' on Self-Administered Questions?" *Public Opinion Quarterly*. 58: 400-408.

Redline, C. D., and Dillman, D. A. 2002. The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In *Survey Nonresponse* ed by R. Groves, D. Dillman, J. Eltinge, and R. Little. New York: John Wiley.

Redline, C.D. Dillman, D. A., Dajani, A. and Scaggs, M. A. In press. Improving Navigational Performance in Census 2000 by Altering the Visually Administered Languages of Branching Instructions. *Journal of Official Statistics.*

Schwarz, Norbert. 1996. *Cognition and Communication Judgmental Biases, ResearchMethods, and the Logic of Conversation*. Mahwah, NJ: Lawrence Erlbaum.

Schwarz, Norbert, Fritz Strack, and Hans-Peter Mai. 1991. "Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis." *Public Opinion Quarterly*. 55: 3-23.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys Experiments on Question Form, Wording, and Context*. New York, NY: Academic Press.

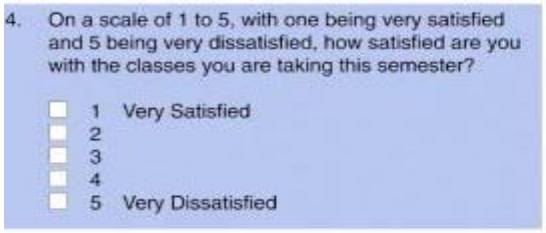
Shettle, Carolyn and Geraldiine Mooney. 1999. Monetary Incentives in U.S. Government Surveys, Journal of Official Statistics, 15(2): 231-250.

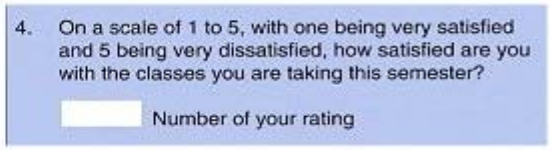
Stapleton, C.N., Norris, S, and Brady, S. 2003. Customer Satisfaction with Internet and IVR as Census Data Collection Tools. Unpublished paper presented at Joint Statistical Meetings, San Francisco, CA, August 6.

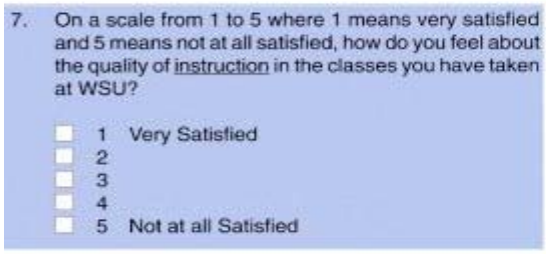
Tortora, Robert. 2002. Personal Communication.

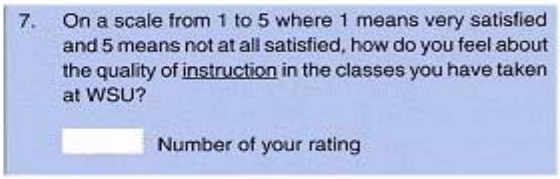
Figure 1: Question wording and formats compared in paper and web experiments

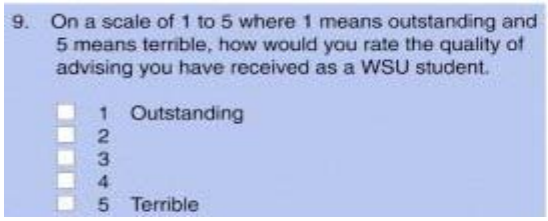
A. Mail Experiment

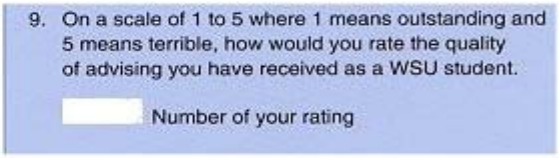




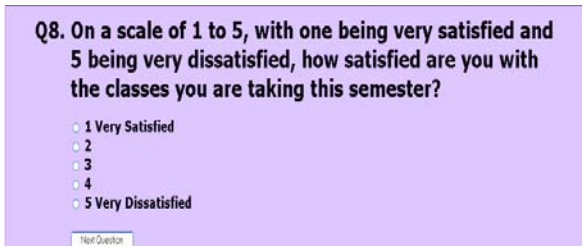


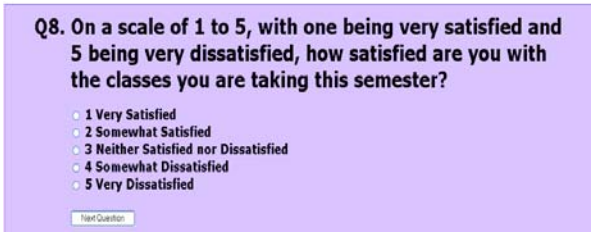


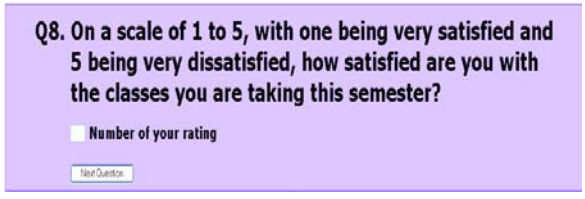




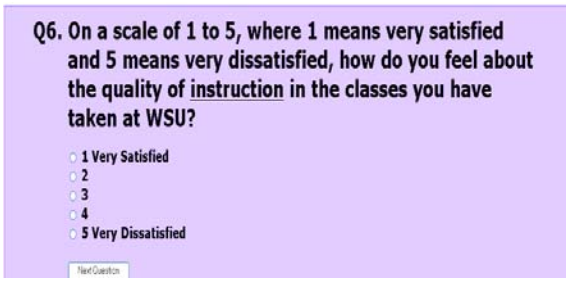
B. Web Experiment

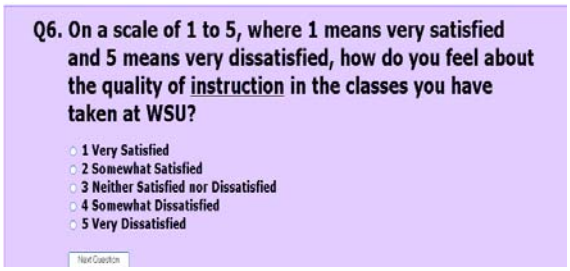


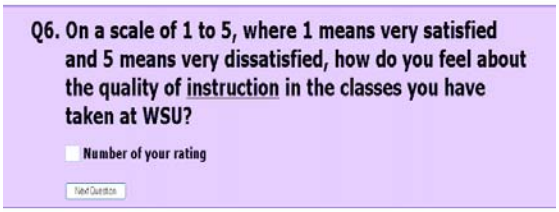


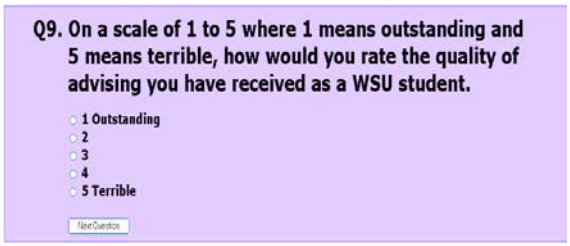


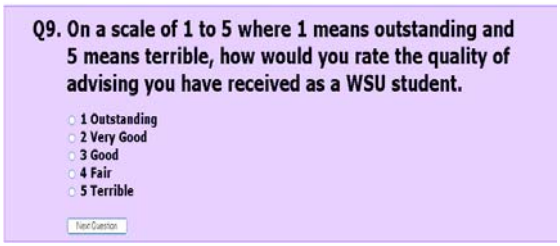
B. Web Experiment (cont.)











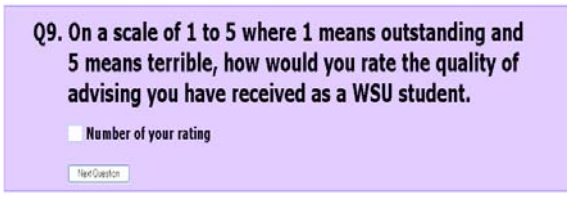


Table 1. Percentage of Respondent choosing each category in paper and web experiments for the same questions.

A. Paper Experiment: Polar-Point vs. Number Box Comparisons

	Question 1		Question 2		Question 3[1]	
Responses	**Polar-Point**	**Number Box**	**Polar-Point**	**Number Box**	**Polar-Point**	**Number Box**
(n)	**517**	**513**	**517**	**512**	**506**	**466**
(1) Very Satisfied	15.9	9.8	10.8	6.3	15.2	14.2
(2)	43.5	29.2	47.6	35.4	31.8	23.8
(3)	31.0	34.5	31.3	33.8	31.6	29.6
(4)	7.7	21.3	9.5	20.9	15.4	21.7
(5) Very Dissatisfied	1.9	5.3	0.8	3.7	5.9	10.7
Total	100%	100%	100%	100%	100%	100%
Mean	2.4	2.8	2.4	2.8	2.7	2.9
Diff. of Means	**t = 7.7**	**p = .000**	**t = 6.9**	**p = .000**	**t = 3.5**	**p = .000**
Chi-square	**χ^2 = 63.4**	**p = .000**	**χ^2 = 48.1**	**p = .000**	**χ^2 = 18.0**	**p = .001**

[1]Response categories to this item were (1) Outstanding and (5) Terrible.

B. Web Experiment Polar Point vs. Answer box Comparisons

	Question 1		Question 2		Question 3[1]	
Responses	**Polar-Point**	**Number Box**	**Polar-Point**	**Number Box**	**Polar-Point**	**Number Box**
(n)	**438**	**302**	**438**	**305**	**437**	**303**
(1) Very Satisfied	19.4	17.2	10.5	6.9	14.0	15.8
(2)	40.0	32.8	43.6	38.0	29.3	21.8
(3)	30.6	27.5	35.2	40.4	30.0	31.7
(4)	8.2	18.5	9.8	12.4	19.0	18.8
(5) Very Dissatisfied	1.8	4.0	0.9	2.3	7.8	11.9
Total	100%	100%	100%	100%	100%	100%
Mean	2.3	2.6	2.5	2.7	2.8	2.9
Diff. of Means	**t = 3.48**	**p = .000**	**t = 2.86**	**p = .002**	**t = 1.34**	**p = .091**
Chi-square	**χ^2 = 21.9**	**p = .000**	**χ^2 = 8.72**	**p = .068**	**χ^2 = 7.63**	**p = .106**

[1]Response categories to this item were (1) Outstanding and (5) Terrible.

C. Web Experiment fully labeled vs. polar point comparisons

Responses	Question 1		Question 2		Question 3[1]	
	Fully Labeled	Polar-Point	Fully Labeled	Polar-Point	Fully Labeled	Polar-Point
(n)	**434**	**438**	**434**	**438**	**433**	**437**
(1) Very Satisfied	29.3	19.4	14.3	10.5	16.2	14.0
(2)	50.2	40.0	63.1	43.6	26.3	29.3
(3)	9.5	30.6	13.4	35.2	26.1	30.0
(4)	9.2	8.2	8.8	9.8	20.3	19.0
(5) Very Dissatisfied	1.8	1.8	0.5	0.9	11.1	7.8
Total	100%	100%	100%	100%	100%	100%
Mean	2.0	2.3	2.2	2.5	2.8	2.8
Diff. of Means	**t = 4.5**	**p = .000**	**t = 5.2**	**p = .000**	**t = .80**	**p = .211**
Chi-square	**$\chi^2 = 62.6$**	**p = .000**	**$\chi^2 = 61.6$**	**p = .000**	**$\chi^2 = 5.27$**	**p = .260**

[1]Response categories to this item were (1) Outstanding and (5) Terrible.