The Web Questionnaire Challenge to Survey Methodologists

Don A. Dillman & Dennis K. Bowker

Abstract

Although the use of web surveys is expanding widely, that growth is not, for the most part, being guided by concerns about survey error reduction. In this paper we discuss the use of web surveys in the context of four traditional sources of survey error–sampling, coverage, measurement and nonresponse. Fourteen proposed principles for the design of web surveys are examined with regard to the attention they give to each of these four sources of error, and the unique error challenges associated with the design of web questionnaires. The needs for these principles to be empirically tested and research to identify the most desirable web surveying practices are also discussed.

Introduction

We are witnessing an explosion in the use of web surveys to collect sample survey information that was previously collected by other modes of surveying. Only a few years ago the use of web questionnaires as a data collection device was not a matter that received research attention from specialists in survey research.

Curiously, survey methodologists have not been the primary advocates for designing and implementing surveys over the Internet. In fact, the use of web surveys seems to have caught the survey methodology community somewhat by surprise. Leadership for the development of web survey procedures has come in large part from computer programmers, many of whom have little or no training in survey methodology. Consequently, as computer programmers rush to out do one another in the creation of complex and dynamic web questionnaires, the traditional designers of surveys find themselves somewhat on the defensive. Rather than being at the forefront of this latest innovation in the conduct of social surveys, survey methodologists are playing catch-up as they learn to master these new survey development tools.

Our purpose in this paper is to discuss the need for survey methodologists to bring their knowledge of the causes and consequences of survey error to the development of web surveys. Specifically, we present and review the four primary sources of survey error and the unique challenges web surveys provide for minimizing such error. Additionally, we consider principles for web survey design in the context of these errors and present the need for research to understand and improve our capability for conducting scientifically valid web surveys.

Sources of Error in Sample Surveys

The remarkable power of a sample survey is its ability to estimate, with precision, the distribution of a characteristic in a defined population (e.g., the proportion of people of various income levels who own their own home). In addition, that estimate can usually be made by surveying only a small portion of the population under study. Sample surveys are subject to four major sources of error, and each must be attended to in order to have confidence in the precision of the sample survey estimates (Groves, 1989). These errors are:

- *Coverage Error*: The result of all units in a defined population not having a known nonzero probability of being included in the sample drawn to represent the population.
- *Sampling Error*: The result of surveying a sample of the population rather than the entire population.
- *Measurement Error*: The result of inaccurate responses that stem from poor question wording, poor interviewing, survey mode effects and/or some aspect of the respondent's behavior.
- *Nonresponse Error*: The result of nonresponse from people in the sample, who, if they had responded, would have provided different answers to the survey questions than those who did respond to the survey.

Any sample survey, for which it is claimed that a relatively precise estimate of the distribution of a characteristic within a population is being made, must attend to the reduction of *all* four sources of error. Unless all members of the population are given a known nonzero chance of being included in the survey, then a sample, no matter how large, cannot be said to represent them. The precision of any survey estimate, or sampling error, typically stated as plus or minus X percent, is based on the number of randomly sampled respondents who are surveyed. Yet, complete coverage, and a very large number of respondents, cannot substitute for poorly worded questions that result in inaccurate answers. Moreover, if respondents to a survey differ from nonrespondents on the variable of interest, then nonresponse error occurs and cannot be compensated for by "doing well" on the other three dimensions of survey error. Although nonresponse error is not conceptually equivalent to response rate, the higher the proportion of sampled respondents who respond to a questionnaire, the greater the likelihood that nonresponse error is small.

All four of these sources of error are as applicable to the design and implementation of web surveys as they are to mail or interview surveys. However, the early implementation of web surveys suggests that some aspects of error, and in particular coverage and nonresponse, have been mostly ignored. Sampling error, though not being neglected, is instead often inferred when it is not appropriate to do so. For example, many web surveys are conducted using samples of convenience or availability and thus depend heavily on the solicitation of volunteer respondents as described by Bandilla (this volume). Each of the four sources of error, and in particular measurement, raises research questions that differ from those raised by other survey modes. It is to these issues that we now turn.

One of the basic assumptions in surveying is the recognition that for simple random samples of a defined population, the precision of results is closely related to completed sample size, or the number of respondents. For example, simple random samples of 100 have a precision of plus or minus ten percentage points, and those of 1,100-1,200 (the size commonly used for election survey prediction) have a precision of plus or minus three percentage points. Sampling error is decreased by about half when sample size is quadrupled. Thus, samples of several thousand are expected to have precision measured in tenths of a percentage point; *assuming no other sources of error*.

The application of this line of reasoning in web surveys leads to the statement of unwarranted but striking claims for survey accuracy. Specifically, the most common indicator of survey validity in a traditional sample survey environment—the number of respondents—has been pushed to the forefront in declarations about the meaning of survey results. For example, when attempting to access a particularly creative and well-advertised web questionnaire, the following message appeared on the screen: "We're sorry, but the survey has been closed. We received more than 50,000 responses, twice the minimum required for scientific validity."

More recently, results were released from a highly visible survey on "Internet Addiction", a topic discussed by Brenner (this volume), that was posted on the ABCNEWS.com website and answered by 17,251 respondents. A newspaper headline that trumpeted a report of the results read, "Internet proves a web of addiction for 11 million worldwide" (Laurence, 1999). An illogical leap was made from knowledge of the percent of these volunteer respondents, for whom the tendency to answer such a volunteer survey seems a likely correlate of any such addiction, to all web users. That leap was the application of the percentage whose answers were defined as revealing an addition to the number of people throughout the world thought to have access to the World Wide Web.

The overlooked problem for both of these surveys is that neither of them gave all members of a defined survey population an "equal or known nonzero chance" of participating in the survey. Furthermore, both of these surveys solicited *volunteer* respondents on the web, the characteristics of which could not have been known to the researchers. Under these conditions, inference from survey respondents to any larger population through inferential statistics is scientifically unjustified.

But, why is the survey logic required for relating population to sample size being ignored as surveyors broadcast response invitations? Perhaps the main reason is a change in how cost considerations influence the design of web surveys. In the past, survey costs made it highly desirable to survey as few people as sample size considerations would allow when using personal interviews, telephone interviews, and/or mail questionnaire surveys. For these traditional methods, each additional respondent who needed to be contacted, and whose answers needed to be processed, represented a significant additional cost to the study.

Historically, cost considerations have provided survey methodologists with support for their appeal to adhere to established survey methods. The argument that one must carefully define the survey population and give each member a known chance of being surveyed and then work to achieve as high a response rate as possible has strong scientific underpinnings. It made little sense to obtain and process responses from tens of thousands when only a thousand responses would provide sufficient precision to meet the survey goals. The waste in travel time, inter-viewer wages, long distance charges, keypunching costs and/or mailing and postage costs for producing survey data from a geographically dispersed population would have been substantial. And, as the importance and visibility of a particular survey increased, so did the likelihood of careful scientific scrutiny of the results.

The advent of web surveying means it is now possible to bring a survey instrument to hundreds of thousands of people and process the results for a cost that would previously only have allowed a few hundred responses to be tabulated. Should several thousand people respond to the posted questionnaire, the answers of all can be processed for little more than the cost of processing the first hundred. Is there little wonder that in universities throughout the world, creative students and faculty are posting questionnaires onto the web and making mass appeals for people to respond? Neither should it be surprising that large corporations are attracted to the prospect of making questionnaires available to all of their employees and customers, and treating returns as if they were taken from a carefully drawn sample. Sponsors appear to be enamored with the prospect that for far less time and cost than would be required for surveying a hundred or two local residents, it is possible to obtain a thousand responses from a much wider, albeit undefined, population. Sophisticated statistical analyses are sometimes applied to such data to test hypotheses about relationships among variables, which the sponsors wish to generalize to undefined larger populations.

Conductors of such surveys have in effect been seduced by the hope that large numbers, a traditional indicator of a high quality survey (because of low sampling error), will compensate in some undefined way for whatever coverage and nonresponse problems that might exist. Large numbers of volunteer respondents, by themselves, have no meaning. Ignoring the need to define survey populations, select probability samples, and obtain high response rates, together provide a major threat to the validity of web surveys.

Coverage of the Web Remains Quite Limited

Coverage error, the result of not giving everyone a known nonzero chance of being included in a survey, is likely to be high for most web surveys of the general public. In no country of the world do most people have access to the Internet. Nua Internet Surveys estimates that 179 million people, or about 3% of the world's population, has been "on-line" at least once in the three months preceding the surveys on which their estimate is based (1999). The United States and Canada account for 102 million, or 60% of the total. In the United States, fairly precise estimates are regularly provided by the National Telecommunications Information Administration through a periodic survey of nearly 60,000 U.S. households. They estimate for December 1998, over 40% of U.S. households now own computers, but only one-quarter of all households have Internet access (NTIA, 1999). This differential access manifests itself more poignantly when one considers that households with incomes of $75,000 and higher are more than 20 times as likely to have access to the Internet than those at the lowest income levels. Moreover, Black and Hispanic households are about two-fifths as likely to have Internet access as are white households, and rural Americans about half as likely to have it as are urban Americans with comparable incomes.

These data suggest some important implications. First, the coverage problem associated with doing web surveys of the general public is huge and cannot be ignored. This problem is likely to persist in all countries of the world for the foreseeable future. In addition, the potentially useful tool of e-mail solicitation is plagued by the lack of standardized addresses and an appropriate method for generating random samples of addresses in the way that random digit dialing facilitated the design of probability based telephone survey methods. For web surveys, the lack of such procedures prevents researchers from ensuring that each "e-mail address" has a known probability of being selected into a random sample of some defined population. It therefore poses a major challenge to web survey methodologists wishing to select scientifically valid samples of households, businesses, or other units using e-mail.

However, this does not mean that researchers are unable to conduct any scientifically valid web surveys. Some populations—employees of certain organizations, members of professional organizations, certain types of businesses, students at many universities and colleges, and groups with high levels of education—do not exhibit large coverage problems. When nearly all members of a population have computers and Internet access, as is already the case for many such groups, coverage is less of a problem.

Against this background, the combined problem of coverage and sampling error in web surveys becomes strongly evident. Unless coverage considerations are carefully attended to and random samples drawn, there is no scientific justification for using inferential statistics to make statements about larger populations. Instead, such arguments seem to be moving outside the scientific realm, becoming an *article of faith* that if we receive tens of thousands of responses they must represent something larger—even though we have no knowledge of what that larger "thing" is. It seems likely then that the "weighting" of underrepresented groups will introduce other errors, e.g., it will be assumed that those few low-income minority respondents who have Internet access are similar to such people who lack that access. Based on that assumption, inferences will be made about the behaviors and opinions for the larger number of those who do not have access to the web.

Reasons for Nonresponse to Web Surveys

Another serious source of potential error in web surveys is the nonresponse problem associated with posting a web questionnaire and inviting people to respond. Response rates are likely to be very low and heavily influenced by interest in the topic and/or the technology of responding (as noted by Vehovar (in press) and the authors of other chapters in this book (Bosnjak & Batinic, this volume; Tuten, Urban & Bosnjak, this volume). Number of contacts (or call-backs) has always been a major influence on response rates to other survey methods, and the tool most depended upon to reduce nonresponse error.

There is little doubt that procedures can be developed for achieving response rates to web surveys that are reasonably comparable with those obtained by other methods (Dillman, 2000). But, is also clear that such procedures will add significantly to the cost of doing web surveys. Even if such costs are similar to those for doing other types of surveys, web surveys will still retain significant advantages such as speed, the elimination of postage and stationary costs, and low processing costs. Nonetheless, as the desire to use web surveys expands, the gap between costs of broadcasting for volunteers to complete web surveys and implementing multiple contacts and incentives to limited samples will remain large, and continue to tempt designers to ignore issues of coverage and nonresponse. Although research is needed for developing appropriate protocols for maximizing response rates from web survey samples, previous research on other survey modes can be drawn upon to determine possible contact strategies for reducing nonresponse rates and ultimately nonresponse error.

Another source of nonresponse to web surveys stems from problems experienced by individuals as they attempt to complete web surveys. Observation of people while they attempt to complete web surveys suggests the existence of at least two sources of significant frustration—lack of basic computer knowledge and poor questionnaire design—lead to premature termination of the survey. Specifically, we have observed respondent frustration from the following:

- People who lack computer experience do not know how to provide and erase certain answers, e.g., radio buttons, which require clicking on a different answer choice vs. HTML boxes which, require reclicking the same box.

- Not knowing what to do with a drop-down menu.

- Not being able to see all of the answer choices without scrolling the page up and down.

- Being forced to answer every question, even when none of the answer choices seemed appropriate.

- Not knowing how close to the end of the questionnaire they were.

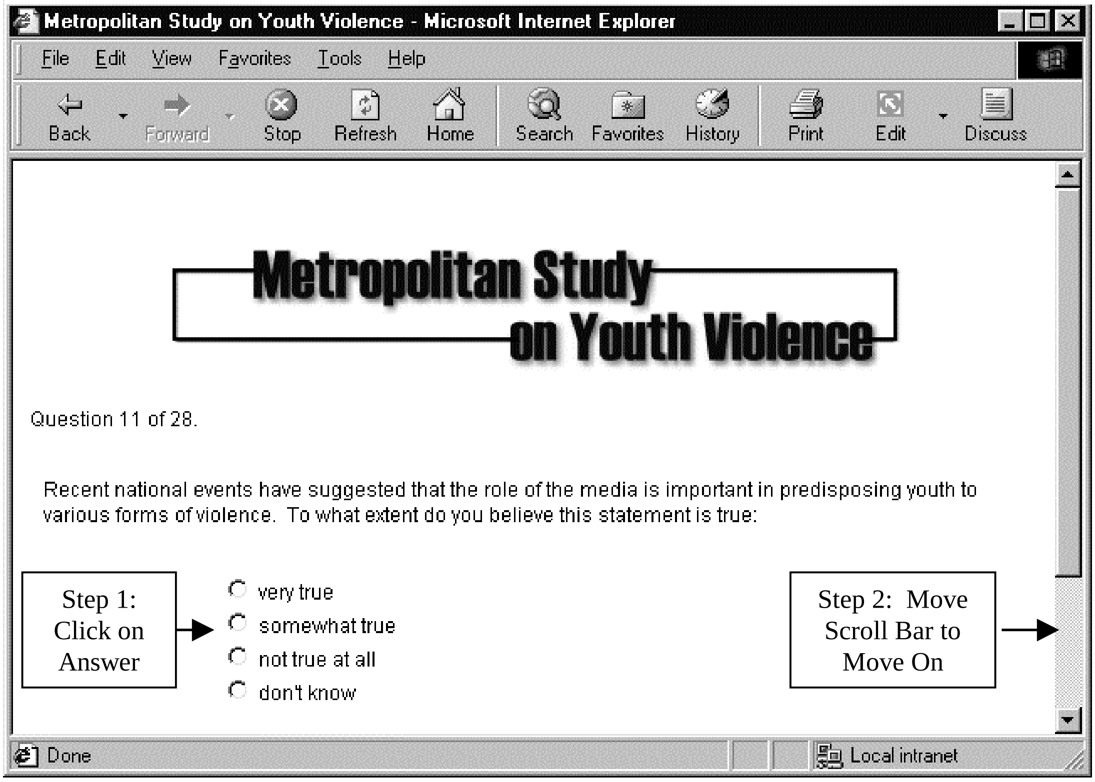
- Only being able to see one question at a time, so that when their concentration was interrupted they had to figure out how to back up and see a question in order to answer the current one.

- Having to take multiple actions to answer each question (e.g., clicking on an answer choice, moving to the scroll bar in order to reveal a "click for next page instruction," and then clicking on that instruction to make the next question appear).

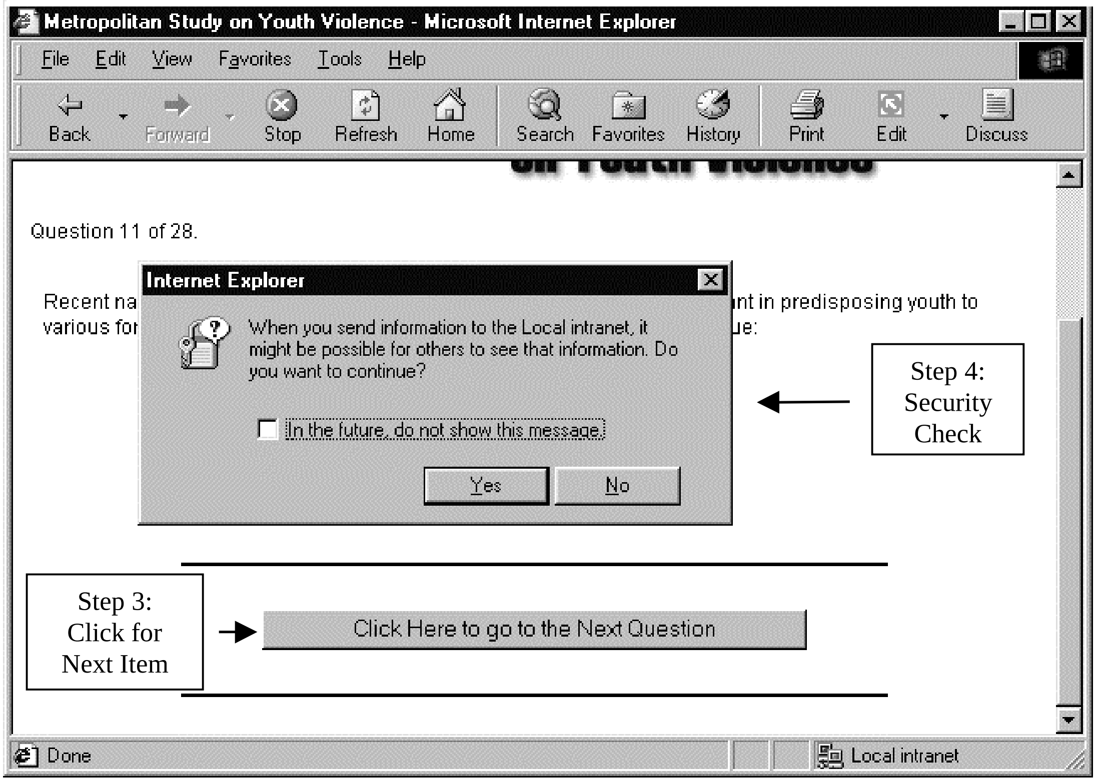
Figure 1 illustrates a web survey that contains excessive navigational controls, as just described. The respondent must engage in a total of four movements to advance to the next question. First, the respondent must process the information presented and select a response category by clicking on the appropriate response. Second, the mouse must be used to scroll down the screen in order to see what is to be done next. Third, the respondent must click a button to advance to the next question. And fourth, when the respondent clicks to submit the response, a fourth window appears warning of possible security issues. Although this last step of security verification can be avoided by setting the browser so as not to display such a message, a certain level of computer literacy is required for making that change which a respondent may not have. Repeating these steps for answering each of the twenty to thirty questions that might constitute a survey is likely to frustrate even the most dedicated of respondents.

Figure 1: Example of a questionnaire using one question per page instruction which required four separate clicks to answer (view before scrolling)

View Before Scrolling:



View After Scrolling:



Nonresponse may also occur because a respondent who has been successfully recruited to participate in a web survey is prevented from accessing some aspects of the questionnaire, the result of incompatibilities in hardware or software. One of the striking differences between web and paper surveys is that what the designer of a web questionnaire may see on his or her screen, may not be what the respondents see on their screens. For example, if the web survey was created using a higher level of HTML language (i.e., 4.0 or 5.0) and advanced features of this language are employed in the questionnaire, then respondents who access the survey with an older browser may not be able to access the questionnaire. And, if they are able to access it, some of the response features may be disabled or perhaps rearranged on the screen.

The web has long been recognized for its ability to provide surveys with a unique avenue for incorporating complex and dynamic interaction controls. Several programming languages exist that allow the designer to implement complex branching, variable stimuli based on previous respondent selections, and an infinite number of creative and visually entertaining navigation controls. Consequently, web survey programmers, who by the nature of their profession strive to remain on the cutting edge of their technology, have often used their advanced programming skills to create surveys that are inaccessible to significant portions of the survey population. A example of this problem was reported in a survey conducted by the U.S. Census Bureau (Nichols and Sedivi, 1998). These authors found that the use of Java for a survey of businesses made it impossible for many members of the survey population to access the survey. In addition, many of those who thought they had adequate equipment and agreed to respond encountered problems in doing so. Whereas 84% of those asked completed equivalent paper questionnaires, only 68% of the experimental group submitted completed web questionnaires.

Programmers who use certain features of advanced programming techniques that are incompatible with a respondent's software, or those that take especially large amounts of time to download, may end up with questionnaires that cannot be received and responded to—at least without extraordinary patience and effort. An illustration of the latter issue is found in an experimental test of a "plain" questionnaire requiring 317k of memory, compared to a "fancy" questionnaire that required three times the memory, 959k (Dillman, Tortora, Conradt and Bowker, 1998). The required time for transmission on a 28.8 modem was at least 120 seconds for the plain version compared to 345 seconds for the fancy one. Individuals whose Internet connections were slow, were undoubtedly likely to spend a longer time receiving the question-naire, and in some cases their computers became overloaded and crashed. Consequently, not only did it take less time to complete the plain version, but respondents were less likely to have to return to the questionnaire at least once in order to complete it. Whereas 93.1% of those who logged into the plain version completed all of it, only 82.1% of those entering the fancy version finished.

If one could ensure that all respondents would have the necessary software installed and configured to participate in a web survey, then much of our argument here would be moot. In fact, those computer assisted interview methods that use the web for personal interviews are not susceptible to the restraints discussed here since respondents answer through the same medium, usually a laptop or other portable computer configured by the research team to be consistent across all interviews. However, given the rapid evolution of Internet technology and the varying rates with which people will choose to adopt such technology, survey methodologists must accept the uncertainty in delivering instruments consistently to the survey population and find ways to minimize its effect.

New Sources of Measurement Error

The advent of web surveying presents measurement challenges not previously faced by survey methodologists, for which research has not yet provided solutions. The enduring problem is that what the designer of a web questionnaire sees on the screen may differ significantly from what some, and in other cases perhaps most, respondents would see on their screens. Evaluation of the web surveys located by Bowker (1999) revealed much variability in methods of construction. When tested on various levels and types of web browsers, operating systems, screen configurations, and hardware, the visual stimulus of the survey items (i.e., physical placement and presentation) was often different than what was originally intended by the designer. Though it could not always be determined what the creators of these surveys had in mind, it can be assumed that none would want to introduce the potential of measurement error due to unequal delivery of item stimuli. Specifically, this review of web questionnaires resulted in finding many questionnaires that exhibited the following characteristics:

- Attitude scales with physical distances between points on the scale which changed as a result of: (1) changing the screen configuration from 800 x 600 to one of the other commonly used sizes, (e.g., 640 x 480 or 1024 x 768); (2) tiling from full-size to half-size screen; and/or (3) use of a lower level browser than was probably used in the construction of the designer.

- The inappropriate use of different background colors for each of several points on an attitude scale such that the visually apparent distance between radio buttons appeared different than it actually was.

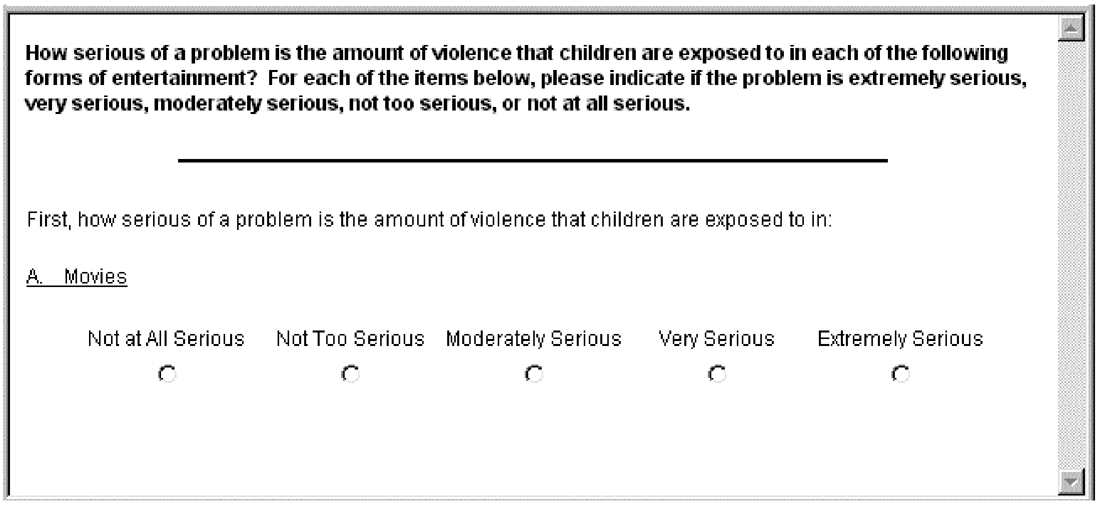
- Different construction procedures for the use of wrap-around text so that when a full-size screen was tiled, the distances between points on attitude scales changed differentially.

- When viewed in a tiled position it was impossible to see the entire question, so that unless the screen size was changed certain answer categories could not be viewed.

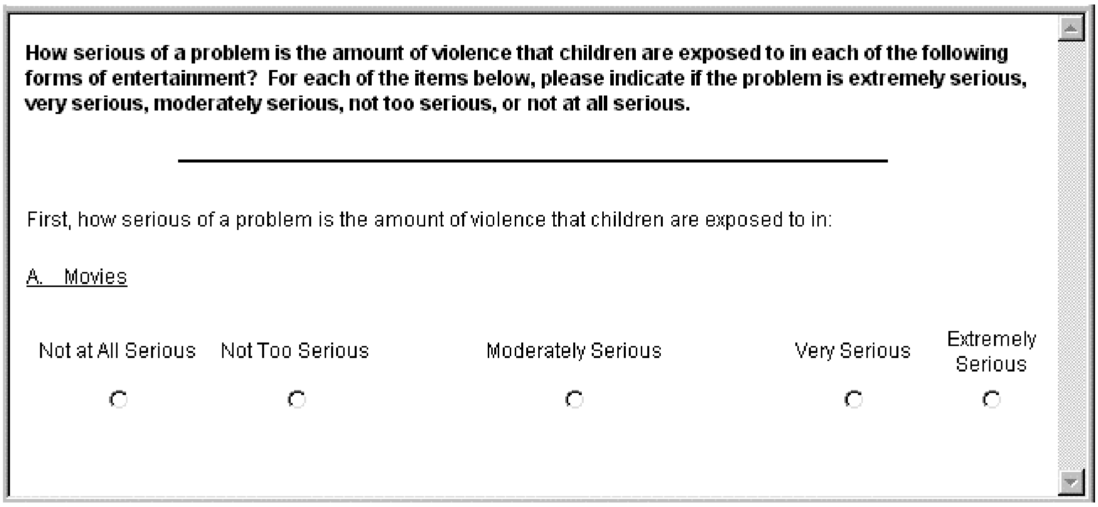
While many factors contribute to the difficulty in delivering the same stimulus to survey respondents (varying web browsers, screen configurations etc.) these factors, though independent in their origin, are not self-governing in their effects. In fact, it is more likely that many of the issues raised above will interact with one another, creating an indeterminable amount of variation in stimuli between and among respondents, which consequently increases the likelihood of measurement error. Figure 2 shows an example where both configuration and browser differences conspire to create a differential effect on the response categories. In the first part of the figure, the questionnaire item categories are evenly distributed when viewed with an 800 x 600-dpi-screen configuration. However, as can be seen in part two, the categories have been distorted due to the change in screen configuration *and* a change in both the level and proprietor of the web browser.

Figure 2: Example of combined effect of screen configuration differences and browser differences resulting in differences in amounts of space between response categories, plus text wrap-around of one of the category labels, so that a respondent sees a different display than did the programmer.

800x600 screen as viewed on Internet Explorer 5



640x480 screen as viewed on Netscape Navigator 4



Although measurement error effects represent one of the most serious threats to the conduct of quality web surveys, it is also one of the most easily addressed through various design controls by programmers. For example, the web survey shown in Figure 2 could easily have been created so as to be viewed properly on all screen configurations and browsers if the designer would have restrained the screen width (preferably to 640 x 480) and the size of response categories within tables. Most web designers are familiar with both the need to deliver a web site that is accessible and viewable to all its viewers, and the inherent problems in doing so. However, because it is usually a trade-off between undesirables for programmers, a much-needed dialogue must occur between researcher and programmer to eliminate unintended measurement effects.

The use of web surveys makes possible entirely new ways of presenting information to respondents. Not only can multiple colors be applied, but also several additional navigational features can be added to questionnaires. Full-color graphics can be used to create more realistic environments of interaction and evaluation for respondents. For example, a person can be asked to choose the most desirable products from among a collection of objects, such as automobiles or appliances, each of which is a different color. Variations in the way, in which the graphics were developed, and even more so in the way in which it is received, can result in significant differences in item stimulus. For example, if the designer created the images with a color palette that uses millions of colors and the respondent can only view the images with 256 colors, then what the respondent sees will be different from what the designer intended for them to see.

Another unique feature of web survey displays is that navigational controls (e.g., drop-down menus), which were unavailable for other methods, can be placed beside items for which respondents are asked to give their opinion. With drop-down menus, a respondent must go through a two-step process of clicking on a box in order to bring a list of response categories into view, and then select their answer choice.

The Need for Research-Based Principles of Design for Web Questionnaires

Procedures for designing web questionnaires in ways that reduce survey error need to be developed and tested. One attempt to develop such procedures, begun by Dillman, Bowker and Tortora (1998), has resulted in the statement of 14 principles of design for web questionnaires (Dillman, 2000). Table 1 lists each of these principles, and also specifies the type of survey error which each was intended to address.

Table 1. Principles for the design of web surveys (Dillman, Tortora and Bowker, 1998; Dillman, 2000) and their relationship to traditional sources of survey error.

	Type of Error			
	Sampling	**Coverage**	**Measure-ment**	**Non-response**
1. Introduce the web questionnaire with a welcome screen that is motivational, emphasizes the ease of responding, and instructs respondents on the action needed for proceeding to the next page.				X
2. Provide a PIN number for limiting access only to people in the sample.	X	X		
3. Choose for the first question an item that is likely to be interesting to most respondents, easily answered, and fully visible on the first screen of the questionnaire.				X
4. Present each question in a conven-tional format similar to that normally used on paper self-administered questionnaires.			X	X
5. Restrain the use of color so that figure/ground consistency and read-ability are maintained, navigational flow is unimpeded, and measurement properties of questions are maintained.			X	
6. Avoid differences in the visual appearance of questions that result from different screen configurations, operating systems, browsers, partial screen displays and wrap-around text.		X	X	X

Type of Error

	Sampling	Coverage	Measure-ment	Non-response
7. Provide specific instructions on how to take each necessary computer action for responding to the questionnaire and other necessary instructions at the point where they are needed.				X
8. Use drop-down boxes sparingly, consider the mode implications, and identify each with a "click here" instruction.			X	
9. Do not require respondents to provide an answer to each question before being allowed to answer any subsequent ones.				X
10. Provide skip directions in a way that encourages marking of answers and being able to click to the next applicable question.			X	
11. Construct web questionnaires so they scroll from question to question unless order effects are a major concern, and/or telephone and web survey results are being combined.		X	X	X
12. When the number of answer choices exceeds the number that can be displayed in a single column on one screen, consider double-banking with an appropriate grouping device to link them together.			X	
13. Use graphical symbols or words that convey a sense of where the respondent is in the completion process, but avoid ones that require significant increases in computer memory.		X		X
14. Exercise restraint in the use of question structures that have known measurement problems on paper questionnaires, e.g., check-all-that-apply and open-ended questions.			X	X

Consideration of these principles from a survey error perspective leads to several observations. First, sampling error, which is mostly a reflection of sample size, is addressed only somewhat indirectly by one of the principles. This principle (#2) provides a PIN number for limiting access only to people in the sample and is mostly concerned with coverage error, that is, giving each solicited sample member a known nonzero, chance of being included in the survey.

Three other principles are also concerned with coverage error. These principles (#6, 11 and 13), as described in detail elsewhere (Dillman, 2000), are concerned partly with technological issues and limiting the amount and frequency of information transfers between the respondent and host computers. However, at first glance they may seem to have much more to do with measurement and nonresponse than they do with coverage. In fact an aspect of each of these principles concerns whether the web questionnaire is designed in a way that makes it possible for participants to receive and respond to the survey instrument. The aforementioned study by Dillman, Tortora, Bowker and Conradt (1998) provides an illustration of survey coverage being unintentionally limited by the technology in a way that prevented sampled individuals from being able to respond. The design of web questionnaires must take into account this technological impediment to coverage, which is no less serious or different in its implications, for example, than people being unable to respond to a telephone survey because of not having a telephone that works.

Eight of the principles shown in Table 1 are concerned with the reduction of measurement error. Certain of these principles (numbers 5 and 6) depict problems associated with differences in what the programmer sends to potential respondents and that which appears on the respondents' screens. For example, principle six notes differences in screen configurations, operation systems, browsers, the use of wrap-around text, and partial screen display preferences that may affect the appearance of attitude scales, and the relative distances between categories. Those differences can also be magnified by the inappropriate use of colors used to surround radio buttons, which make categories appear closer or further apart, a topic addressed by principle 5. Two of the remaining measurement principles (8 and 10) concern measurement dilemmas unique to web questionnaires: the use of drop-down boxes and skip directions which, when clicked upon, take respondents to the next appropriate question. In contrast, the other measurement-related principles (4, 12 and 14) concern compatibilities important when web surveys are a component of mixed-mode surveys in which data are also being collected by paper questionnaires.

Nine of the principles in Table 1 directly concern nonresponse error. Each of these principles is aimed at reducing potential frustration among respondents that can lead to nonresponse. Among the concerns they address are complicated introductory screens that discourage less computer literate individuals from continuing with the remainder of the questionnaire (Principle 1), and initial questions that require individuals to perform a response task with which they are unfamiliar. They also concern potential frustrations associated with the inappropriate use of drop-down menus (Principles 3 and 8), and letting respondents know how close they are to the end so they do not become discouraged and quit even though they are nearly finished with the questionnaire (Principle 13). Taken together, these principles suggest that the potential sources of respondent frustration with web surveys are many and varied.

The list of principles presented here is far from complete. In particular, it does not include survey implementation procedures, for example, number of contacts, methods for contact, and use of incentives, which are necessary for obtaining high response rates, and decreasing the likelihood that respondents are different than nonrespondents. And, some of the goals, e.g., limiting access through to the use of PIN numbers, can be accomplished in different ways.

The formulation of these principles was based upon considerable, but unsystematic, observations of questionnaires and individuals attempting to complete them. For most of the principles, there is as yet little or no experimental evidence to confirm quantitatively the importance or lack of importance of their use in designing most web surveys. It is important that such research be done.

However, the need for research goes much further that simply evaluating such principles. For many years scientists have conducted research in which some data are collected by one mode and some by another (Dillman and Tarnai, 1989). More recently, greater use of mixed-mode surveys is being promulgated by the opportunity to improve response rates through the use of mixed-mode designs and apparent declines in response rates to telephone survey methods in particular. It is also being encouraged by the ease with which word files can be moved from one form of questionnaire to another, so that adding an additional mode to a data collection project is much easier to accomplish, than in the past. This trend, in combination with the major coverage problem facing conductors of web surveys, suggests to us that most early efforts to conduct scientific surveys via the World Wide Web are going to be mixed-mode in nature. As of yet, we have little knowledge of whether answers people provide to a web survey will differ significantly from those provided on paper and/or on the telephone. This too represents an important area of research for which advancements need to be made.

The research that is needed in order to facilitate the development of scientific underpinnings for conducting quality web surveys will be diverse. Some of the needed research involves conducting cognitive interviews, in which people with varied computer experience and skills, are observed as they attempt to fill out web questionnaires using equipment and access procedures that are much different from those of the questionnaire designers. Other research will consist of traditional experiments in which alternative design and implementation methods aimed at improving response are tested. The needed research will be guided both by past survey research on other methods of self-administration (c.f. Jenkins and Dillman, 1997) as well as new concepts from the computer usability literature (c.f. Schneiderman, 1997).

Summary and Conclusions

Throughout the 20th century survey methodologists have faced a series of methodological challenges, from the development of efficient sampling methods and personal interview methods to the creation of procedures for computer assisted surveying by telephone. As the 21st century begins, another set of challenges exists, one of which is the development of principles for the effective and efficient design of self-administered web surveys. To date, survey methodologists have invested little time or effort on this topic.

The potential efficiency and cost associated with doing surveys on the world wide web has led to the rapid deployment of such surveys, with little attention being given to consideration of the essential scientific underpinnings needed for conducting quality sample surveys. In this paper we have argued for the need to give explicit attention to four types of survey errors—sampling, coverage, measurement and nonresponse—in designing such surveys. We have also argued that inherent characteristics of web surveys, from low costs to the fact that technology may intervene between what the designer thinks has been created, and what the respondent sees, raises new issues not previously addressed by survey methodologists. Thus, rather than simply applying what is already known, new research is needed for identifying the best of means of helping web surveys to achieve their potential as an effective means for conducting sample surveys. The outcomes from that research will determine whether web surveys can be depended upon for making precise estimates of the distributions of characteristics in specific populations, the function performed so well by other types of surveys throughout much of the 20th Century.

References

Bandilla, W. {--OSS Chapter}

Bosnjak, M. & Batinic, B. {--OSS Chapter}

Bowker, D. (1999). *Constructing the client-computer interface: Guidelines for design and implementation of web-based surveys.* Summary Report 99-15, Social and Economic Sciences Research Center. Washington State University, Pullman, WA.

Brenner, V. {--OSS-Chapter}

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

Dillman, D.A. & Tarnai, J. (1989). Administrative Issues in Mixed Mode Surveys. In R.M. Groves, P. P. Biemer, L. E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.) *Telephone Survey Methodology* (pp. 509-528) New York: Wiley-Interscience.

Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). *Principles for constructing web surveys: An initial statement.* Technical Report 98-50, Social and Economic Sciences Research Center. Washington State University, Pullman, WA.

Dillman, D. A., Tortora, R. D., Conradt, J., & Bowker, D. (1998). *Influence of plan versus fancy design on response rates for web surveys.* Unpublished paper presented at annual meeting of the American Statistical Association, Dallas, TX.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York, Wiley.

Jenkins, C. R. & Dillman, D.A. (1997). *Towards a Theory of Self-Administered Questionnaire Design*. In L. Lyberg, P. Biemer, M. Collins, L. Decker, E. deLeeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), Survey Measurement and Process Quality (pp. 165-196) New York: Wiley-Interscience.

Lawrence, J.M. (1999). *Internet Proves a Web of Addiction for 11 Million Worldwide,* Boston Herald. August 23rd.

National Telecommunication and Information Administration. (1999). *Falling through the net: Defining the digital divide,* Available: www.ntia.doc.gov/ntiahome/net2/falling.html.

Nichols, E. & Sedivi, B. (1998). *Economic Data Collection Via the Web: A Census Bureau Case Study.* Paper presented at the meeting of the American Statistical Association, Dallas TX.

Nua Internet Surveys. (1999). http//www.nua.ie/surveys/

Schneiderman, B. (1997). *Designing the User Interface*. Reading, MA:Addison-Wesley.

Tuten, T., Urban, D. & Bosnjak, M. {--OSS Chapter}

Vehovar, V., Zenel, B., Lozart, K., and Zaletel, M. (in press). Nonresponse in Web Surveys. In Groves, R.M., Dillman, D.A., Eltinge, J. , and Little, R. *Survey Nonresponse.* New York: Wiley-Interscience.