5-2007

Design Effects in the Transition to Web-Based Surveys

Don A. Dillman
Washington State University, dillman@wsu.edu

Jolene D. Smyth
Washington State University, jsmyth2@unl.edu

Published in *American Journal of Preventive Medicine* 32:5S (May 2007), pp. S90–S96; doi:10.1016/j.amepre.2007.03.008

Published online April 26, 2007.

Design Effects in the Transition to Web-Based Surveys

Don A. Dillman[1] and Jolene D. Smyth[2]

1. Social and Economic Sciences Research Center, Washington State University, Pullman, Washington, USA
2. Department of Sociology, Washington State University, Pullman, Washington, USA

Corresponding author – Don A. Dillman, Social and Economic Sciences Research Center, PO Box 644014, Washington State University, Pullman WA USA 99164-4014, email dillman@wsu.edu

Abstract

Innovation within survey modes should always be mitigated by concerns about survey quality and in particular sampling, coverage, nonresponse, and measurement error. This is as true today with the development of web surveying as it was in the 1970s when telephone surveying was being developed. This paper focuses on measurement error in web surveys. Although Internet technology provides significant opportunities for innovation in survey design, systematic research has yet to be conducted on how most of the possible innovations might affect measurement error, leaving many survey designers "out in the cold." This paper summarizes recent research to provide an overview of how choosing the web mode affects the asking and answering of questions. It starts with examples of how question formats used in other survey modes perform differently in the web mode. It then provides examples of how the visual design of web surveys can influence answers in unexpected ways and how researchers can strategically use visual design to get respondents to provide their answers in a desired format. Finally, the paper concludes with suggested guidelines for web survey design.

Introduction

In the early 1970s telephone technology provided a new and innovative medium to conduct surveys. The telephone quickly became a major survey mode because it was relatively

inexpensive and had nearly 90% household coverage rates. In addition, the development of random-digit-dialing procedures allowed efficient random sampling of households, and research indicated that telephone survey results were quite similar to those obtained via face-to-face interviews.[1] As telephone surveying advanced, however, methodologists quickly realized that how this new technology was used to conduct surveys was limited by concerns about survey error. Many new techniques were developed to deal with the limitations of the telephone mode (e.g., shortening response scales and converting fully labeled scales to polar point–labeled scales) and a vast body of research addressing the strengths and limitations of telephone surveying has been developed.

Surveyors are now in the beginning stages of another major change in survey methods as cultural changes in how the telephone is viewed and used (e.g., reliance on answering machines) have led to reductions in the advantages that previously drew them to the telephone (i.e., coverage and cooperation rates), and technological innovation has produced a new mode of data collection, the web survey.[2] The Internet provides significant opportunities for innovation, making the web survey an incredibly exciting data collection tool. However, as surveyors incorporate web surveys into their major data collection repertoire, it is useful to keep in mind the lessons of the past. Use of the various innovative techniques possible in web surveys should always be mitigated by concern with minimizing survey error, which is now conceptualized as sampling (i.e., smaller sample sizes yield less-precise estimates), coverage (i.e., error resulting from part of the population not having a known, non-zero chance of being sampled), nonresponse (i.e., error resulting from differences between sampled individuals who completed the survey and those who did not), and measurement error (errors resulting from poor question or questionnaire design).[2,3]

Each of these sources of error is influenced by the choice of survey mode and by how the survey is designed within a mode; however, the challenge for web surveyors in general and eHealth methodologists in particular is that the body of research addressing how survey error is affected in web surveys is in its infancy. As a result, there is little information about the effects of using techniques developed for use in other modes or of using new web innovations on survey error in web surveys. The research that has been done, however, is instructive. This paper focuses primarily on one of the four sources of survey error—measurement error—and provides an overview of how choosing the web mode affects the asking and answering of questions. In doing so, special attention is paid to the transition from telephone surveys to web surveys, as this transition poses significant challenges because of fundamental differences between these two modes and because of the increasing use of mixed-mode survey designs.

Factors External to Web Survey Design That Affect Measurement

A number of cultural or social challenges affect how respondents interact with web surveys. Many people are wary of the Internet because they feel that they have little control over it. As more people learn about Internet fraud such as "phishing" (i.e., an identity theft scam involving the impersonation of a legitimate business/organization to get one's personal information), the spread of computer viruses, and the collection of personal information through "spyware," they become increasingly distrustful, especially of unknown

senders of e-mail. Moreover, the Internet is approached by some users as a source of entertainment and anonymity, which means that their interactions there may lack sincerity and honesty and be less governed by societal norms of communication, interaction, and exchange.

Coupled with the cultural and social challenges that accompany web surveys are technological challenges. Whereas telephone technology has always been relatively standardized and easy to use, the complexity faced by potential respondents to web surveys is enormous and ever-changing. Differences in Internet connection speeds, computer configurations (i.e., platforms, processing power, screen sizes, browser software), and computer and Internet skills may increase or decrease the willingness of respondents to complete web surveys. Perhaps more important though, technological variation as well as variation in the ability to use the technology may result in respondents experiencing the stimuli of a web survey in significantly different ways. For example, altered visual presentation of elements or different mental/emotional responses may lead them to interpret and respond to survey elements in significantly different ways resulting in measurement error. Certain specialized programming steps can be undertaken to minimize differences across software and hardware configurations; however, these do not work in all instances. For example, some methods require Java Script to be enabled in order to be effective, but many computer users disable Java Script on their computers.

Because these factors exist external to survey design, surveyors are limited in the extent to which they can immediately and directly influence them. In the short run there is not much that can be done to ease people's distrust of the Internet. Although certain steps can be taken to reduce privacy and legitimacy concerns (e.g., the use of encryption methods for secure data transmission, providing telephone numbers for validation and sending the survey through a trusted source), these steps are not complete remedies. Likewise, surveyors cannot control the computer and Internet connection configurations that people use. Nor can they affect the computer/Internet skill levels of potential respondents. What they can do though is design user-friendly surveys that minimize measurement error based on what is known about respondent processing and response behavior once inside the survey instrument.

What Is Known about Respondent Processing and Response Behavior?

Schwarz[4] argued that survey respondents conduct themselves in survey situations as if they are involved in a conversation, with the survey itself representing the researcher's contribution to the conversation. Respondents abide by universally observed rules of communication that dictate, for example, that one should be understandable, clear, concise, honest, and not repetitive,[5] and they expect the researcher to do the same. In applying conversational rules to the survey process, respondents assume that all of the information contained in the questionnaire is important. Thus, they glean information about question meaning and expectations from what survey designers often consider to be formal features of the questionnaire (i.e., question order, format, and, in self-administered surveys, graphic presentation) in addition to the obvious verbal or written cues.[6] As a result, design ele-

ments that seem inconsequential to the survey designer may hold great meaning for respondents and may significantly affect how they interpret and answer questions in the survey.

Formal Feature #1: Question Format

Tradition is a major driving force behind most survey construction. Within a single mode, it is logical to use the question formats that work best, and over time specific question formats and survey modes become coupled and tradition is born. The web is no exception. Despite its youth, web survey traditions have begun to develop as researchers adjust to the absence of an interviewer who could interpret, prompt, cajole, or otherwise motivate respondents to provide complete and accurate answers.

At the same time that reliance on web surveying has rapidly expanded, so too has the use of mixed-mode surveying as a means to increase response rates at minimal cost. For example, in assessments of eHealth interventions, baseline data are often collected via one mode (e.g., paper), while follow-up assessments are conducted via another (e.g., online). In other instances, users who fail to respond to paper or web assessments may be contacted by telephone. This use of mixed-mode surveys poses a direct challenge to mode-specific traditions. While surveyors might use mode-specific question formats intending to collect the same information from all respondents, these formats may actually provide different stimuli to respondents and therefore artificially produce different responses (i.e., measurement error).

In one example, a transition from the telephone mode to the web mode led to persistent differences in answers to a simple question. In an open-ended format, the telephone question asked, "What is your marital status?" In the web survey the question was identically worded but asked in a closed-ended format with the response options single, married, separated, divorced, or widowed. In the web mode, 4.5% fewer respondents reported being single or married while 3.5% more reported being separated and 1% more reported being divorced or widowed each month (R. Tortora, The Gallup Organization, personal communication, May 9, 2002). The obvious explanation for the change was the mode-specific switch from an open-ended question on the telephone to a closed-ended question with suggestive response options on the web. Although this difference seemed easily resolved, the response from the well-established telephone unit was to note that this was the way the marital status question is always asked.

A similar situation often occurs with multiple-answer questions. In telephone interviews such questions are usually posed as a series of forced-choice items to each of which respondents answer "yes" or "no." However, on the web, a response feature that allows more than one answer to be selected, the check box, is typically employed so that respondents can simply go down the list and check the appropriate items[7] (see Figure 1 for an example).

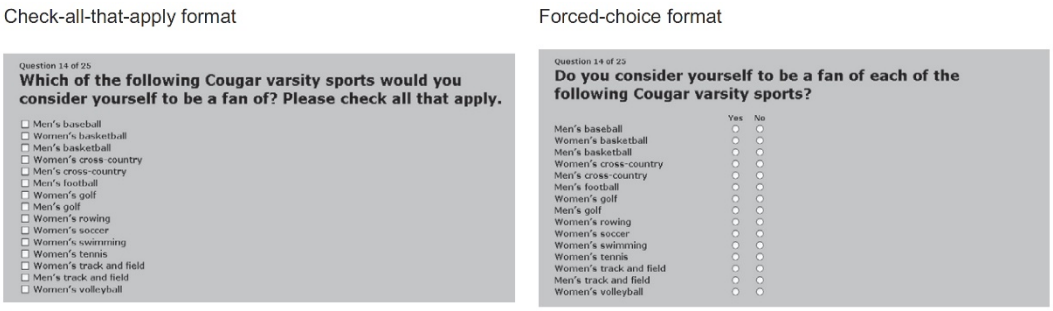

Figure 1. Example of multiple-answer question formats.

A recent experimental comparison of these two formats **within a web survey** (Figure 1) showed that for 16 questions respondents to the forced-choice format endorsed more options than respondents to the check-all format; 15 of the 16 comparisons were significant. Forced-choice respondents marked an average of 5.0 response options while check-all respondents marked only 4.1. Additionally, 91% of the individual response options were endorsed more often in the forced-choice format. Moreover, a segment of the check-all respondents who answered the questions in the mean time period or less (66%), were more likely to mark options when they appeared in the top of the list, a finding that suggests they were employing a satisficing response strategy (i.e., choosing the first few defendable responses just to meet the requirements of the question and failing to exert the effort required to provide optimal responses).[8]

A follow-up study replicated the above **within-web** findings and then compared responses from telephone forced-choice and web check-all respondents **across modes**. This study found that the telephone forced-choice respondents endorsed significantly more items than the web check-all respondents, and that individual response options were more likely to be endorsed in the telephone forced-choice format.[9] Further experimentation showed no difference, however, in the endorsement of items when the forced-choice question format was used across modes, a finding that challenges unyielding reliance on mode-specific conventions.

Formal Feature #2: Visual Elements in Web Surveys

Even when the conventional practice of varying question format across telephone and web modes is avoided, other differences that are more difficult to resolve may arise because these survey modes rely on fundamentally different types of communication. In telephone surveys, respondents engage in a conversation, not too unlike those they have every day. The essential form of communication here is aural. The words that respondents hear are their major source of information about what is being asked of them and how they should respond, but they may also gather such information from inferred interviewer characteristics (e.g., gender, race/ethnicity) or from paralinguistic cues (e.g., voice inflection, tone or emphasis, and timing).[10,11]

Web respondents, however, take their cues from visual instead of aural stimuli. Visual cues can include both verbal messages (text), and seemingly formal features of the questionnaire made up of numeric (the use of numbering), symbolic (the use of symbols with culturally defined meanings), and graphic elements (spacing, color, brightness, size, font),[12,13] the visual counterparts to aural paralanguages. The use of numbers, symbols, and graphic features affect how respondents interpret questions and instructions by evoking cultural rules about their meaning (e.g., items close in proximity must be related).[13,14] For example, Jenkins and Dillman[13] noted that Gestalt psychology laws of perception apply to the survey situation. According to these laws, figures that are located close to one another or are similar in appearance are more likely to be seen as belonging to the same group (law of proximity and law of similarity) and figures that are simple, regular, and symmetrical are easiest to perceive and remember (law of Pragnanz). Tourangeau et al.[14] further argued that respondents use interpretive heuristics (they discuss five) to draw meaning from visual features of questionnaires. For example, respondents applying the "near means related" heuristic will conclude that items located in close proximity on a screen or page are conceptually related to one another.

The following examples show the dramatic differences that such visual features can make in web surveys. Each of these examples focuses on a different problem, but all of them illustrate how visual features can cause respondents to answer differently than the researcher intended. Consequently, they each illustrate difficulties involved in transitioning from an aural- to a visual-based survey mode.

From Single- to Multiple-Column Scalar Presentations

Oftentimes web surveyors take advantage of the horizontal orientation of screens by displaying response options in columns rather than in a single vertical display. Using data from a 2001 paper survey experiment containing the same question and scales presented in Figure 2, Christian and Dillman12 found that respondents were more likely to choose "good" and less likely to choose "very good" when the options appeared in multiple columns. In a web survey, Christian[15] found the same significant response differences between single- and multiple-column scale arrangements. These findings suggest that respondents to the multiple-column versions may be reading from left to right, focusing their attention on only the top row of options (i.e., excellent, good, poor). In contrast, the vertical arrangement seems to encourage respondents to read from top to bottom and to process all of the scale points. Thus, it appears that by taking advantage of the horizontal orientation of computer screens in this way researchers may be unwittingly compromising the quality of their data.

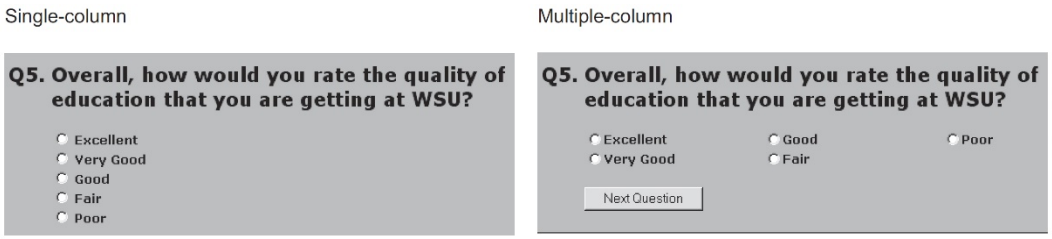

Figure 2. Example of linear and nonlinear scale layouts.

Unintentional Effects of Grouping

In web surveys researchers can use visual features to assist respondents; however, the use of such features can cause unintended effects. Smyth et al.[16] reported on a series of experiments designed to address the effects of using spacing and headings to subgroup response options in a single-answer question (Figure 3), a practice used in an ongoing national survey. Using data from three web surveys, they reported that subgrouping increased the number of options respondents chose and the likelihood that they chose an option from each subgroup. These effects occurred regardless of how the subgroups were arranged relative to one another (i.e., vertical or horizontal), regardless of question type (opinion vs fact), and regardless of whether headings were present. Respondents even continued to select options from both subgroups when an instruction to "Please select the best answer" was added to the question. These results suggest that visual features used intentionally or unintentionally can have unknown response effects.

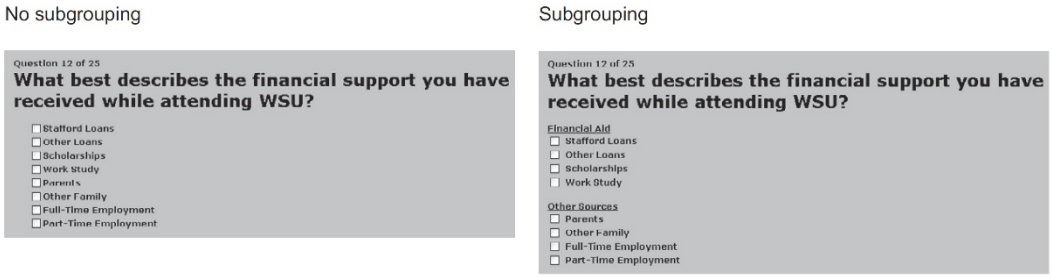

Figure 3. Example of subgrouping response options.

Other research has shown that how questions themselves are grouped or separated from one another can significantly influence responses. Results of an experiment cited by Schwarz[4] (Schwarz and Hippler [1992]), for example, showed that a question about marital satisfaction was **less** correlated with a question about general satisfaction when the questions were grouped together by placing them in a single box than when they were presented separately in two boxes. In contrast, Tourangeau et al.[14] found eight items to be **more highly** correlated when they appeared all together on one webpage as opposed to

either split over two pages or in a one-question-per-page design. This apparent contradiction in findings can be traced to the relationships between the items in these studies. Because the items in the experiment cited by Schwarz[4] did not measure the same underlying construct, they were prone to assimilation or contrast effects. In contrast, all items in the Tourangeau et al.[14] study were indicators of a latent construct, and as such were not prone to assimilation or contrast effects.[17]

Encouraging Specific Response Formats

Since web survey respondents lack interviewers to convert their answers into acceptable formats, web survey designers often build error messages into their questionnaires to get respondents to correct their mistakes. Such error messages, however, increase respondent frustration and survey termination, so it is necessary to use visual elements to help respondents answer correctly the first time and avoid error warnings. But, as demonstrated above, visual features can have undesirable effects. It is important, therefore, to determine which design elements yield desired effects while minimizing undesired effects.

In a recent example, Christian et al.[18] set out to find the best way to get respondents to report the date they began their studies using a desired format, a two-digit month and a four-digit year. Through a series of manipulations of symbolic and graphic elements they increased the percentage of respondents using the desired format from a low of 45% to a high of 96%. Figure 4 shows the formats that induced the lowest and the highest use of the desired format. To maximize use of the desired format, they adjusted the size of the answer boxes to reflect the expected number of digits, moved the answer spaces closer to one another to encourage use of the same response format (numbers) in both boxes and to discourage use of alpha characters in the month box, replaced "Month" and "Year" with the symbols "MM" and "YYYY" to communicate the desired number of digits, and moved the symbolic instructions into the natural reading path. This series of manipulations and the resultant nearly universal use of the desired format show that it is possible to use visual design to reduce respondent errors and, on web surveys, to help respondents avoid error notices.

Figure 4. Example of using visual design to induce reporting in the desired format.

These examples demonstrate the difficulty of transitioning from telephone to web surveys or of trying to combine the two in a mixed-mode research design. It is simple to copy and paste questions from telephone survey software into web design software, but one

cannot assume that such questions communicate the same information to respondents. In moving a question from an aurally- to a visually-based medium, the meaning and/or expectations communicated by the question may be altered. Such changes occur because respondents apply cultural rules to visual features in the questionnaire, resulting in interpretations that are altered from what is communicated in the wording of the question alone. Consequently, the graphic manipulations pointed out by Jenkins and Dillman[13] can be used strategically to help respondents answer correctly and avoid receiving error messages.

Conclusion

Just as telephone survey practices were constrained and shaped by concern with survey error, the use of the abundant potential for innovation that the Internet provides should also be mitigated by a concern with the four main sources of survey error: sampling, coverage, nonresponse, and measurement. While minimizing total survey error (i.e., the aggregation of all four types of error) is the ultimate goal, the present conversion from a predominantly aural survey mode, the telephone, to a predominantly visual mode, the web, makes measurement error particularly salient. The examples described above provide a brief demonstration of just some of the measurement challenges that web survey designers and eHealth researchers must contend with.

This is not to say that telephone surveys provide the gold standard for data quality; rather, the point here is that understanding the particularities of each survey mode is important for optimal design within and across modes. That this paper has focused on the particularities of web surveys does not mean that other survey modes are immune from such challenges. Nor does it mean that web surveys should always be subordinated to other survey modes. In some instances, such as when respondents are asked to report sensitive data, computer-based surveys seem to provide better data than other modes and perhaps this benefit should be exploited in the interest of data quality.[19] However, in other instances, such as when asking multiple-answer questions, formats that were originally developed for the telephone (i.e., forced choice) may produce higher-quality data in web surveys than those that tradition suggests should be used (i.e., check all). The work of sorting out when it is more advantageous to utilize features only found in web surveys and when it is more advantageous to forsake these features for mode comparability has only just begun.[2]

While still in its infancy, the body of research pertaining to measurement issues in web surveys has grown significantly over the last six years[14,16,20,21]; however, many design elements still need to undergo systematic experimentation and testing to determine how they affect measurement. In the meantime, web survey designers can turn to several sources for theoretically grounded web survey design guidelines. First, both an article by Crawford et al.[22] and the book *Internet Data Collection* by Best and Krueger[7] lay out theoretically grounded and, where possible, research-based, guidelines for web survey design. Second, web survey designers can consult the self-administered survey design literature.[2,23] Many guidelines established in this literature for paper surveys are relevant to web survey de-

sign, especially for web surveys using a static design (i.e., a design where multiple questions are included on one page). Finally, researchers can consult the literature on website usability,[24] including industry-specific web-based usability guidelines such as those developed by the National Cancer Institute.[25] However, as Crawford et al.[22] are careful to point out, "researchers must always consider the goals of survey research when applying knowledge gained from other fields of study, such as website usability" as the goals driving survey design differ significantly from those driving website design. Some suggestions for general website usability may not be appropriate in the web survey context.

While researchers can consult these resources for information about specific design decisions, the following are some more general web survey design guidelines:

- Articulate questions and response options clearly, keeping in mind that respondents use more than just words to determine question meaning and response expectations.
- Use the "bells and whistles" sparingly. In other words, use graphics, sound, animation, and so on only when necessary. The effects of many of these multimedia elements on data quality are largely unknown, but the few studies that do exist indicate that they have a substantial influence on answers (Couper et al.[21]). Also, to the extent that they increase download time, they may contribute to nonresponse.
- Avoid simply borrowing question wording and formats from other survey modes. Sometimes it may be desirable to use the exact question wording and format from another mode, but other times it may be necessary to alter question wording or format to achieve the same question meaning across modes.
- Evaluate the survey before fielding it using techniques such as cognitive interviews and pre-test prototypes.
- Above all else, use design elements with consistency and regularity. A design element (e.g., boldface text) should mean the same thing at the beginning of the survey as it does at the end. Consistency and regularity in design will help the respondent more efficiently process information throughout the questionnaire.

Acknowledgments – This paper is a revision of a presentation made at the Critical Issues in E-Health Research Conference in Bethesda, Maryland, June 9, 2005. This research was supported by the Social and Economic Sciences Research Center and the Department of Community and Rural Sociology. Don A. Dillman is regents professor and the Thomas S. Foley distinguished professor of government and public policy in the Department of Community and Rural Sociology, Department of Sociology, and the Social and Economic Sciences Research Center. Jolene D. Smyth is a graduate assistant in the Department of Sociology and the Social and Economic Sciences Research Center. No financial conflict of interest was reported by the authors of this paper.

References

1. Groves RM and Kahn, RL. Surveys by telephone: a national comparison with personal interviews. New York: Academic Press, 1979.
2. Dillman DA. Mail and Internet surveys: the tailored design method. New York: Wiley, 2000.
3. Groves RM. Research on survey data quality. Public Opin Q 1987;51:S156–72.
4. Schwarz N. Cognition and communication: judgmental biases, research methods, and the logic of conversation. Mahwah NJ: Lawrence Erlbaum, 1996.
5. Grice PH. Logic and conversation. In Cole P, Morgan JL eds. Syntax and semantics. Vol. 3. Speech acts. New York: Academic Press, 1975:41–58.
6. Schwarz N, Grayson CE, Knäuper B. Formal features of rating scales and the interpretation of question meaning. Int J Public Opin Res 1998;10:177–83.
7. Best SJ, Krueger BS. Internet data collection. Thousand Oaks CA: Sage, 2004.
8. Smyth JD, Dillman DA, Christian LM, Stern MJ. Comparing check-all and forced-choice question formats in web surveys. Public Opin Q 2006,71:66–77.
9. Smyth JD, Dillman DA, Christian LM. Does "yes or no" on the telephone mean the same as "check-all-that-apply" on the web? Paper presented at the Second International Conference on Telephone Survey Methodology, Miami FL, 2006.
10. de Leeuw ED. Data quality in mail, telephone, and face to face surveys. Amsterdam: TT Publications, 1992.
11. de Leeuw ED. To mix or not to mix: data collection modes in surveys. J Off Stat 2005;21:233–55.
12. Christian LM, Dillman DA. The influence of graphical and symbolic language manipulations on responses to self-administered questions. Public Opin Q 2004;68:58–81.
13. Jenkins CR, Dillman DA. Towards a theory of self-administered questionnaire design. In: Lyberg L, Biemer P, Collins M, de Leeuw E, Dippo C, Schwarz N, Trewin D, eds. Survey measurement and process quality. New York: John Wiley and Sons, 1997:165–98.
14. Tourangeau R, Couper MP, Conrad F. Spacing, position, and order: interpretive heuristics for visual features of survey questions. Public Opin Q 2004;68:368–93.
15. Christian LM. The influence of visual layout on scalar questions in web surveys. Master's thesis, Washington State University, 2003.
16. Smyth JD, Dillman DA, Christian LM, Stern MJ. Effects of using visual design principles to group response options in web surveys. Int J Internet Sci 2006;1:6–16.
17. Peytchev A. Effect of multiple questions per page in web-based surveys on data quality: a measurement error approach. Paper presented at the Midwest Association for Public Opinion Research, Chicago, 2003.
18. Christian LM, Dillman DA, Smyth JD. Helping respondents get it right the first time: the influence of words, symbols, and graphics in web surveys. Public Opin Q 2007;71:113–25.
19. Wright DL, Aquilino WS, Supple AJ. A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. Public Opin Q 1998:62:331–53.
20. Couper MP, Traugott MW, Lamias MJ. Web survey design and administration. Public Opin Q 2001;65:230–54.
21. Couper MP, Tourangeau R, Kenyon K. Picture this! Exploring visual effects in web surveys. Public Opin Q 2001;68:255–66.

22. Crawford S, McCabe SE, Pope D. Applying web-based survey design standards. J Prev Interv Community 2005;29:43–66.
23. Dillman DA, Gertseva A, and Mahon-Haft T. Achieving usability in establishment surveys through the application of visual design principles. J Off Stat 2005;21:183–214.
24. Nielsen J. Designing web usability. Indianapolis: New Riders Publishing, 2000.
25. National Cancer Institute. Research-based web design and usability guidelines 2001. Available at: http://usability.gov/guidelines.