

OPPORTUNITIES AND CHALLENGES OF CONDUCTING WEB SURVEYS: RESULTS OF A FIELD EXPERIMENT

Kumiko Aoki & Michael Elasmr, Boston University
Kumiko Aoki, 640 Commonwealth Ave., Boston, MA 02215

Key Words: Web Survey, Internet Survey

As Internet usage is becoming increasingly common, the Internet as a survey data collection tool is also becoming an increasingly hot topic of discussion among researchers. As a data collection tool, the Internet has evolved from a simple ASCII-test based email survey to an interactive web survey where both response content and response characteristics are interesting to observe. Internet surveys, as a concept, present new opportunities for researchers who want to collect and analyze data as efficiently as possible. Though there are still sampling limitations to overcome when using the web for surveying general populations, the web survey presents distinct advantages over traditional modes of data collection.

This article first reviews the literature about the advantages and disadvantages of web surveys, then reports the various challenges that researchers face when planning and implementing a survey via a secured web site. This article reports the results of a field experiment that examined the impact of data collection mode (web versus traditional mail) on response rate, response speed and other characteristics of survey response.

LITERATURE REVIEW:

In contrast to the number of studies that have examined email surveys, there is relatively little scholarly writing on the topic of web surveys. Existing information focuses on the limitations that prevent a researcher from using the web as a tool for survey research. Stanton (1998) identifies three areas where the limitations lie: sampling problems, response consistency problems, and participant motivation problems. Among the most discussed problems is sampling. Internet/web surveys still cannot use a representative sample of the general population, as a significant proportion of the general population is still not connected to the Internet.

Even if a researcher is not interested in the general population but rather in the population of web users, he/she will still face severe sampling handicaps. There is no comprehensive sampling frame of web users from which to draw a probability sample and there is no widely accepted method equivalent to Random Digit Dialing (RDD) in telephone surveys. In addition, the most common

form of a web survey involves placing a questionnaire on a web page and hoping that web surfers will take the time to fill it out. This poses a data integrity or access control problem, as the questionnaire placed on a public web site can be accessed by anyone, even participants whose interest might be to intentionally distort the data. This problem, however, can be overcome by employing a proper methodological design for the web survey. Such design includes implementing access control mechanisms such as password protection for the survey so that only individuals sampled by the researcher can access it.

There are also a myriad of technical issues that need to be considered before implementing a web survey as discussed in Dillman, Tortora, and Bowker (1999), Pitkow and Recker (1994) and Smith (1997). The lack of standardization among operating systems, servers, and browsers causes a number of problems and requires time-consuming efforts to ensure the compatibility and the practicality of a web survey.

Web surveys are often compared to mail surveys since they are both self-administered. Some researchers argue that web survey respondents might exhibit response characteristics that do not exist in traditional paper and pencil surveys. MacElroy (1999) termed such characteristics as "E-Personality". According to MacElroy, the characteristics of E-Personality are: the tendency to display aggressive behavior, the tendency to have stronger opinion positions, intense candor, cynicism, and the higher expectation for compensation.

Though the researcher should be aware of the many limitations associated with using the web for collecting survey data, he/she should also be aware that this mode of data collection offers some distinct potential advantages. Advantages cited in the literature include: low cost, "quick turnaround time, collapsed geographical boundaries, user-convenience, and more candid and extensive response quality" (Smith, 1997). In addition, in an interactive web survey using Java or Javascript for handling questions, the survey designer can impose constraints upon the respondent's response options or implement various interactive features that allow complex skip patterns, which eliminate data entry errors and can ensure a higher quality response (Schaefer and Dillman, 1998; Stanton, 1998).

In terms of response rate, past studies about using the web as a survey tool have reported inconsistent findings. Kwak and Radler (2000) found the response rate of web surveys to be lower than that of traditional mail surveys. By contrast, other studies found the response rates of web surveys to be higher than those of mail surveys.

The present study's general research question is: How do the response characteristics of a web survey differ from those of a traditional mail survey?

METHODOLOGY:

Overall Research Design:

From August, 1999, to February, 2000, the Communication Research Center at Boston University conducted a study to detail the processes that companies use when launching new products. In order to identify the sequence of events that marketing managers follow when launching new products, a database of marketing managers was purchased from a known vendor. The resulting sampling frame included the names and telephone numbers of marketing managers in consumer products organizations. Those marketing managers on the list were contacted by telephone to be recruited to participate in the survey. Once recruited to participate in the survey, each marketing manager was informed that the survey was being administered via the World Wide web. When a marketing manager refused to fill out the survey via the web, he/she was given the option of receiving a mail version of the instrument.

Since the purchased sampling frame did not contain email addresses, those marketing managers who agreed to fill out the questionnaire on the web were then asked to provide the recruiter with an email address through which they can be contacted. Those who opted to take the survey online subsequently received an email message explaining the purpose of the survey and how to connect to the web survey site and describing the incentives offered. Email messages were personalized in order to establish a good rapport with the respondents.

For those who preferred to receive a mail survey, a mail package containing a personalized cover letter explaining the purpose of the survey, the questionnaire, a postage-paid business reply envelope, and an incentive form was sent out via standard U.S. mail.

As an incentive to take part in this study, all respondents were offered a free CD of their choice and an opportunity to get the executive summary at the end of the project.

Web Site Design:

Selection of web Survey Software

Many web survey software packages were considered for this project. Based upon the description of various web survey packages provided on their respective web sites, only two of these packages were purchased and evaluated for this project. First, we evaluated the SPSS Data Entry web Server 1.0 (the SPSS web Server), a product that had just come on the market. As we were familiar with SPSS for statistical analysis and SPSS Data Entry Station for manual data entry, it was a natural choice for us to go with the SPSS web Server for creating a web survey site. It promised valuable features that no other web survey creation software had: the seamless connection between the questionnaire on the web and the SPSS dataset file for data analysis, data validation based on SPSS variable value definitions, etc.

However, during our evaluation of this package, we encountered two severe limitations. First of all, the SPSS web Server used Java programming language for data manipulation, which did not work with older versions of web browsers. Even with newer versions of browsers, it unbearably slowed down the loading of the questionnaire page, as it required a considerable amount of computer memory. Second, the limited capability of editing and formatting the questionnaire after uploading the questionnaire onto the server (the questionnaire could not be pretested locally and had to be uploaded to the SPSS web Server for viewing the actual format and layout that appears in a browser) made it very difficult to format the questionnaire exactly the way we wanted. The SPSS web Server would have been a good web survey creation/execution tool if the format and layout of the questionnaire were not an issue and if we could guarantee that all respondents had the latest web browser with a very powerful computer. The reality is that such guarantees could not be made at that point; therefore, the SPSS web Server could not be used for this project.

Being fully aware of the handicaps to overcome, we started looking for web survey software that could create a simple HTML form-based survey. A simple HTML form-based survey cannot include interactive features such as constraints and automatic skip patterns as discussed earlier. However, as noted in Dillman, Tortora, and Bowker, (1999), such a simple web survey will ensure a higher response rate and will reduce the likelihood of nonresponse error. In addition, a package with this feature enables questionnaire editing and formatting via any regular HTML authoring program.

In short, there were two conditions that we needed the software to fulfill: 1) an ability to format the questionnaire on the web identically to its printed version; and 2) an ability to be displayed on the most basic browser on the market. This latter condition was necessary in order to avoid getting responses from only those with the latest browsers and fastest computers. Based on these two conditions, we selected SurveySaid by Marketing Masters.

Design of the web Survey Questionnaire

The paper questionnaire spanned eight pages using a 10-point font. In order to avoid the possibility of page layout effects when comparing the results of the paper questionnaire to its web equivalent, we needed to mimic the paper version as closely as possible. In the process of finding the most optimal way of achieving this, we needed to make several choice decisions:

Web page layout:

A choice had to be made between laying the questionnaire on one long web page that the respondent had to scroll through, and dividing the questionnaire across multiple web pages. The issue involved the difference in the media: paper versus CRT screen. With the paper version, a respondent had to flip from one page of the questionnaire to another. Each paper page was clearly distinct from the other. The respondent does not have to look for the beginning of the next page as flipping the previous page takes you directly to the next one. On a printed questionnaire, once a respondent flips a page, there is no waiting as the next page is immediately available to them. On the web, however, putting the entire document on one page would require the respondent to scroll down and pay much more attention to not skipping questions due to fast scrolling. Putting the document on multiple pages also had its problems. Depending on web congestion, a respondent might have to wait several minutes when attempting to move from one page to another.

Web page submission:

In addition to page layout considerations, another concern to the researchers involved the submission of questionnaire results to the server. If the web survey consists of multiple web pages, the data will be sent to the server whenever the respondent completes a page and hits the "next" or "submit" button. Every time the respondent hits the "next" button, the new page has to be downloaded from the server. Having multiple pages gives the respondent the possibility of giving up in the middle of the survey and submitting partial data. By

contrast, if the web survey consists of a single web page the data will be sent to the server all at once when the respondent hits the "submit" button at the end of the survey. In order for this to happen, however, the respondent has to scroll all the way to the end of the survey to click on the "submit" button. This task may seem easy but could be problematic when a respondent gets fatigued and decides to quit before the end of the survey. In such a case, no data will be sent to the server at all.

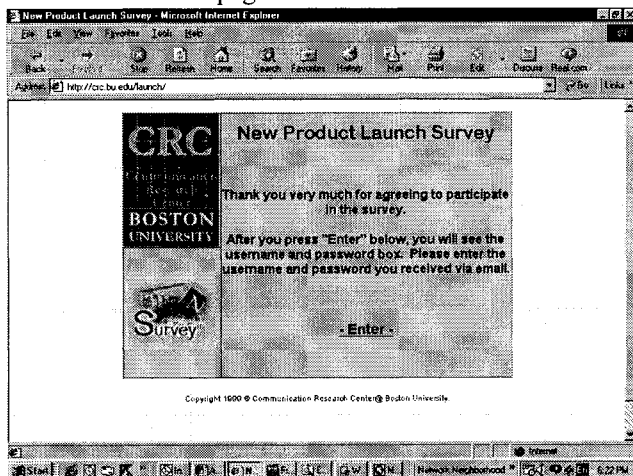
The layout and submission considerations led the researcher to adopt a hybrid approach to the web survey: 1) a single page was used to avoid the problems involved in a multiple page layout; and 2) the illusion of multiple pages was created through page skip icons labeled "next" and appearing at the bottom of each section; this minimized the problems associated with scrolling. Clicking on a "next" icon took the respondent to another portion of the same page. Using a single page ensured that the respondent's transition from one section to another would be instantaneous. The illusion of a multiple page layout was created by inserting a blank screen between sections so that the respondent can't see any text beyond the screen immediately before him/her.

In order to make the questionnaire appear similar to the printed version, we avoided the use of drop-down menus for multiple-choice responses. We used radio-buttons for multiple-choice (i.e., only one choice could be checked) questions and check boxes for multiple-response (i.e., check-all-that-apply) questions. We also used alternating bands of color (light green in the cream yellow background) in a large table to help respondents align questionnaire items in the computer screen. For skip questions, hyperlinks were provided so that by a click of the highlighted text, the respondent can skip to the appropriate next question. As the questionnaire contained a number of different tables and layouts, the formatting of the questionnaire web page in HTML was the most tedious part of the web site construction.

The questionnaire file was about 405KB, taking about one minute to download using a 56kbps modem but almost instant download in a 10Mbps corporate LAN environment. If the majority of respondents are expected to take the survey at home, one minute of downloading time might discourage response. However, in this study, the majority of respondents were expected to take the survey in their offices where a higher-speed Internet connection is usually in place. Therefore, a file size of 405KB was not expected to be a discouraging factor for their participation.

Construction of an Entire Web Survey Site

The web survey site had to have additional web pages besides the actual questionnaire: one to welcome visitors and another to thank the respondents upon completion of the survey. We created two layers before the respondent sees the actual questionnaire and two layers after the completion of the survey. The very first page a visitor saw was the welcome page. This was displayed when anyone typed the URL of the survey. Anyone who pointed his or her browser to the URL was able to see the page below.



From this page, only those who had been invited (i.e., those with the appropriate username and password) could see the next page that was the instruction page. Creating a secured web site is a very important procedure in creating a web survey so that the researchers can restrict the access to the site to only those who have been selected in the sample and thereby ensure both sampling criteria and integrity of the data.

There were three ways available to us for limiting access to the web server only to survey participants:

- 1) create an individual account for every respondent;
- 2) create a group account for all the respondents; or
- 3) generate a unique identifier for each respondent on the fly and embed the number in the email message sent to the respondent.

We created a special group account in our Windows NT server and used the account's username and password exclusively for the web survey. In other words, by creating a group account all the respondents use the same username and password to access the survey. Using this system ensured the anonymity of the respondents and eliminated the

complexity of managing multiple accounts for the survey.

Following this username and password page, we included a page on which appeared the same information that was sent in the cover letter to the mail survey respondent. This cover page included the definition of terms and statements reiterating confidentiality and anonymity. The cover page is followed by the actual questionnaire. After a respondent completed the survey and hits the "submit" button, the incentive page appeared.

As an incentive to complete the survey, we offered all respondents a CD of their choice (among five different top-of-the-charts CDs) as well as the opportunity to receive a copy of the executive summary of this study. We created another form for this that contains radio buttons for different CDs and text boxes for the respondent to give his/her name and address to which the CD can be mailed. Again in order to ensure the anonymity of the survey responses, we created a separate database for the incentive forms that separates respondents' names and addresses from their survey responses.

The last page a respondent saw is the "thank you" page that appeared after the respondent fills out the incentive form and hits the "Submit Information" button.

RESULTS:

Mode of Data Collection and Response Rate:

A total of 575 marketing managers were contacted by telephone. Among them, 170 (29.6%) agreed to take part in the survey, 100 (17.4%) refused, and 181 (31.5%) were left voice mails three times. Among those who agreed to participate in the survey, 98 (57.6%) preferred to take the mail survey and 72 (42.4%) people preferred to take the survey on the web.

Out of those 98 people who agreed to take the mail survey, 32 (32.7 %) actually returned the completed questionnaire via mail after a varying number of follow-up letters. Six of those who decided not to take part in the survey after the questionnaire was mailed to them replied to us with courteous letters informing us of their inability to participate in the study.

Out of the 72 who agreed to take the survey on the web, 29 (40.3 %) completed their questionnaire. Four of those who decided not to take part in the survey after the email invitation message was sent to them sent us reply messages indicating their inability to participate. A few respondents told us that their companies' firewall policy did not allow them to visit the survey web site. One of them

requested the questionnaire to be emailed as a document file attachment, and the person returned the completed questionnaire through email.

Mode of Data Collection and Response Time:

With respect to the time span from the initial questionnaire mailing to the receipt of the completed questionnaire, the average number of days for the mail survey was 50 days. Actually none of the mail survey responses arrived before the follow-up letter that was sent two weeks after the initial mailing of the questionnaire and 13 of 32 were returned after the third follow-up mail which contained the questionnaire and a one-dollar gift certificate for a cup of coffee from a known coffee house as a token of appreciation. By contrast, web survey responses were received within three days of the original contact. This clearly shows a major time advantage for a web survey method over a mail survey method.

In order to increase the sample size, we also sent out questionnaires to those who never answered their phone but to whom a message was left on their answering machines or voice mailboxes. We sent out the questionnaires to 181 of those and received 17 (9.4%) completed questionnaires from them.

After three months from the initial mailing of the questionnaire and two months after the last follow-up, another effort was made to increase the sample size. We sent out another letter with the questionnaire to those who had agreed on the phone to take the survey, but never returned it. The letter promised a new incentive of a \$50 gift certificate upon the receipt of the completed questionnaire. We sent out the questionnaire package with the new incentive to 68 of the people who had agreed to take the mail survey and 43 of those who had agreed to take the survey on the web. Ten out of 68 (14.7%) returned the completed questionnaire via mail and two out of 43 (4.7%) completed the questionnaire on the web. It seems that the higher incentive works better for the mail survey takers than the web survey takers.

Cost per Response:

It is difficult to make a direct cost comparison between a mail survey and a web survey as both involve a number of intangible costs as well as tangible costs. In our study, having a small sample size for both mail survey and web survey, the per-response cost of the mail survey and that of the web survey seemed approximately equal. The main cost of the web survey was associated with the creation of the web site. This is a one one-time cost regardless of the sample size. The main cost in the mail survey was associated with the printing and mailing activities (including tangible costs for postage,

letters, and envelopes and intangible labor costs for creating and assembling the mail package) and the data entry costs, both of which are incremental costs that grow with the size of the sample. For a large-scale survey, a web administration is more cost effective than a mail survey. The larger the sample size, the greater the differences in cost effectiveness between web and mail surveys and the greater the advantages of the web over the traditional mail approach.

Differences in Response Characteristics Between the Mail and Web Versions:

Originally there were 35 responses returned for the web survey. However, after looking at the data closely, we found that there were four sets of duplicate responses based on their identical responses and the same originated IP addresses. Within each set of responses, only one set was totally complete. It seems that some respondents started the questionnaire and then stopped in the middle, but nonetheless submitted the responses. Then, later they revisited the questionnaire site, started it over, and completed the questionnaire.

Five out of 28 (17.9%) of the web survey respondents did not complete the entire questionnaire, but proceeded to submit the data without filling out the rest of the questionnaire. On the other hand, all the mail survey respondents completed more than 80% of the entire questionnaire. In other words, in mail surveys it is more likely that a respondent will not return an incomplete questionnaire while in web survey it is more likely that researchers receive partial responses as respondents can submit even incomplete responses with a click of the "submit" button.

Contrary to our expectation and also being inconsistent with the findings from past studies (Schaefer and Dillman, 1998; Kwak and Radler, 2000), respondents of the mail survey provided significantly more open-ended answers than respondents of the web survey. For three completely open questions asking for descriptions and explanations, on average 54.8% of all the mail survey respondents answered each open-ended question in comparison to 31.4% of all the web survey respondents. This may be attributable to the fact that for web survey takers to answer open-ended questions, they have to move their hand from the mouse to the keyboard. As all the lengthy open-ended questions are placed at the end of the questionnaire, after answering many questions by just a click of the mouse, this may become a little too cumbersome for them.

CONCLUSION:

In this paper, we describe our experience with a field experiment to conduct a web survey and to compare that mode of data collection with a traditional mail survey, in terms of the challenges a web survey faces in creating and executing the survey and the differences in response rate, response time, cost, and response characteristics. The major findings from this study are:

- More respondents preferred to take the survey via a printed instrument sent by mail while a significant proportion chose to complete the questionnaire on the web.
- The web survey had a slightly higher response rate than the mail survey.
- With respect to time span of actual survey completion, web survey respondents completed the survey in an overwhelmingly shorter period of time.
- With respect to cost, the larger the sample size, the greater will be the differences in cost effectiveness between web and mail surveys with web surveys being more cost efficient.
- web survey respondents submitted more partial response questionnaires than those who chose to take the survey by mail.
- Mail survey respondents provided significantly more open-ended answers than those who chose to take the survey on the web.

In sum, the study showed major advantages of a web survey: higher response rate, shorter response time, and lower cost with a larger sample. However, the study also showed disadvantages of a web survey at present: technical complexities, page design effects, greater chance of getting partial responses, and tenuous open-ended responses. Some of the disadvantages of a web survey may be overcome as the technology more advances; people become more used to using a computer and the web; and more research is done to determine best practices.

REFERENCE:

- Couper, M. P., Blair, J., and Triplett, T. (1999). A comparison of mail and e-mail for a survey employees in Federal Statistical Agencies. *Journal of Official Statistics*, 15, 39-56.
- Dillman, D. A., Tortora, R. D., and Bowker, D. (1999). Principles for constructing web surveys.
- Kwak, N. and Radler, B. (2000). Using the web for public opinion research: A comparative analysis between data collected via mail and the web. Paper presented at the annual meeting of American Association for Public Opinion Research, Portland, OR.
- MacElroy, B. (1999). The effect of E-Personality on research results. Socratic Technologies, Inc.
- Pitkow, J.E. and Recker, M M. (1994). Using the web as a survey tool: Results from the second WWW user survey. Available online at http://www.cc.gatech.edu/gvu/user_surveys/User_Survey_Home.html
- Schaefer, D. R., and Dillman, D. A. (1998). Development of a standard e-mail methodology: Results from an experiment. *Public Opinion Quarterly*, 62, 378-97.
- Smith, C. (1997). Casting the Net: Surveying an Internet population. *Journal of Computer-Mediated Communication*, 3 (1). Available online at <http://www.ascusc.org/jcmc/vol3/issue1/smith.html>
- Stanton, J.M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, 51, 709-725.