

Kim Sheehan and Mariea Hoy

ON-LINE SURVEYS

From 'Using e-mail to survey internet users in the United States: methodology and assessment,' *Journal of Computer Mediated Communication* 4(3) March 1999

THE INTERNET PRESENTS ENORMOUS potential for interaction between on-line users and researchers. . . . [We present] evidence based on previous research that discusses the strengths and limitations of web page-based surveys and assesses the viability of using e-mail as a survey data collection method. . . .

Web-based surveys

To date, the Internet offers both web page-based surveys and e-mail for prospective researchers to use for data collection. Web page-based surveys tend to collect broad-based data from individuals all over the world who self-select to respond to surveys that are posted on web sites. These web page-based polls can collect demographic information, as well as other types of purchase, psychographic and opinion data. Numerous benefits to web-based surveys have been noted.

A web page-based survey can take advantage of the graphic power available through programming languages such as HTML and JavaScript to create an attractive, interesting, and compelling survey that is inviting to respondents. . . . The use of CGI scripts allow adaptive questioning, which means that the questions that a respondent is asked depend on his or her answers to previous questions (Kehoe and Pitkow 1996). This allows for follow-up questions that can enrich responses as well as easier navigation for respondents.

Web page-based polls have been noted for their ability to generate a high number of responses (Kehoe and Pitkow 1995); the GVU polls at Georgia Institute of Technology generate more than 10,000 responses per poll. The sheer number of responses suggests that the results represent a diverse set of users. For example, it was estimated that one out of every 100 on-line users responded to each of the GVU polls (Kehoe and Pitkow 1996).

This high volume of responses can be collected very quickly. . . . For example, studies have shown that several hundred responses can be generated over the course of a single weekend (McCullough 1998). This time factor alone suggests huge benefits over traditional surveying techniques in terms of being able to collect and analyze data quickly, and implement decisions based on the findings.

The costs of both data collection and analysis can be minimized by the use of web-based surveys (McCullough 1998). Outside of high start-up costs for equipment and web page design, the actual implementation of a survey can be almost free, with no costs for paper or postage. Data analysis can be simplified by a direct transfer from the form to the analysis software, where limited data cleaning would be necessary (McCullough 1998).

Web page-based surveys allow for anonymity in responses, since the respondent can choose whether to provide his or her name or not. Previous research (Kiesler and Sproull 1986) has indicated that anonymity may affect response rates positively, as respondents may be more willing to respond without fear that their answers may be identifiable to them.

Since respondents type in their answers directly to a form on a web page, there is no need for an interviewer to have contact with the respondents. . . . Therefore, survey responses will be free from errors caused by interviewers, resulting in cleaner data (McCullough 1998).

Similarly, the lack of an interviewer eliminates any potential for bias that the interviewer brings to the survey. An interviewer's mood, prejudices or opinions will not be reflected in the data (McCullough 1998).

However, web-based surveys do present some limitations that researchers must recognize when they are considering this method.

Web page-based surveys must attract respondents to the web page with messages posted in news groups, links on other web pages, banner ads, and other types of methods. As a result, all segments of a Web population may not be represented in the sample (Kehoe and Pitkow 1996). All Internet users do not use the same browsers, and different browsers may not present images and text on web pages in the same manner. For example, some users (such as those subscribing to freenets) use only a text-based web browser (such as Lynx), and may not be able to respond to the survey. Some web based-polls are announced in Usenet newsgroups. Therefore, if potential respondents are not a frequent visitor to newsgroups, they may not be aware of the survey announcement posted in newsgroups, and thus may not have the opportunity to complete the survey. The self-select nature of web page-based surveys also may affect their generalizability. . . .

Web page-based polls generally allow for multiple responses from a single individual, as well as responses from individuals outside of the population of interest (e.g. persons in countries where a product or service is not available, or from

persons who are younger or older than the population of interest). This could also bias the results.

One way to validate a method is to compare it to other methods that are accepted within the research community. Since it is almost impossible to develop response rates to web page-based surveys (Kehoe and Pitkow 1996), it is difficult to compare web page-based survey methods to traditional survey data collection methods such as postal mail and telephone surveys. This leads to another generalizability issue. Without an understanding of the size of the respondent pool in comparison to the size of the universe and the sampling pool, it is also difficult to generalize research findings beyond the universe of those responding to the survey.

E-mail as a data collection method

Using e-mail as a survey data collection method comparable to postal mail may ameliorate some of the issues inherent in web page-based data collection. Previous research presents several reasons to support the idea the e-mail offers much promise as a means of administering surveys as well as pitfalls to be avoided.

Today, as many as 100 million people worldwide have access to e-mail (DOC 1998). Eighty per cent of all users use the Internet daily, with many reporting that 'surfing' replaces 'TV viewing' as entertainment (Kehoe, Pitkow and Morton 1997). The sheer number of individuals using the medium coupled with the frequency and ease with which they could be contacted suggest that e-mail is a viable survey method.

A lack of a national directory of e-mail addresses could be seen as a limitation to e-mail surveys. For example, Schuldt and Totten (1994) reported a problem with obtaining names for their sample. This situation has changed in recent years. Many content providers compile their own databases and should be able to access names quickly from these sources. Some organizations (such as universities and trade associations) publish directories, both paper and on-line, with e-mail addresses. On-line search engines such as Lycos provide 'People finders' for e-mail addresses.

When respondents use the 'reply' function of their e-mail programs to return their completed surveys, their names and e-mail addresses can be automatically written on the electronic message (i.e. the survey) the researcher would receive. While previous research (Kiesler and Sproull 1986) has indicated that anonymity may have affected response rates positively, other researchers (Couper, Blair and Triplett 1997) suggest that the lack of anonymity may not have any effect on response rates. With e-mail surveys, anonymity could be guaranteed through the use of encryption technology, and confidentiality can be guaranteed through confidentiality assurances. This study chose to guarantee confidentiality. Assuring that responses will be confidential throughout the data collection process should help to build respondent trust and enhance response rates.

An additional benefit to using e-mail is that duplicate responses can be eliminated. Steel, Schwendig and Kilpatrick (1992) suggested that duplicate responses can become problematic since researchers using postal mail often send out multiple copies of questionnaires to their entire sample in order to increase response rates.

E-mail presents a benefit over postal mail, then, since e-mail responses can be tracked and previous respondents can be eliminated from follow-up e-mail.

E-mail surveys may allow the researcher to develop a profile of non-respondents. Depending on the search engine used and the respondents' server, some demographic information about persons with e-mail accounts is available on-line and some demographic information such as gender and location may be compiled. It might also be possible to attempt to contact non-respondents using an alternative method (such as postal mail or telephone) to solicit responses that could be compared to the e-mail sample for similarities. It should be noted that demographic information about persons with e-mail accounts may not be completely accurate, as individuals may have changed locations or jobs since the information was provided. However, the availability of such data allows for options that the researcher can consider when assessing non-response.

As with web page-based surveys, there appears to be some cost savings inherent in using new technology. Parker (1992) indicates that cost savings from e-mail compared to traditional mail and telephone surveys are based on low transmission costs and elimination or reduction of paper costs. E-mail may also present cost savings over web page-based surveys, as costs for page design and posting to a server would not be incurred. However, some savings may be offset by the on-line server used (costs vary by Internet service provider) and time considerations (transmission costs may increase by the minute, which may impact the length of the survey).

When respondents perceive technology as easy to use, they seem more likely to respond (Parker 1992). As more people become familiar with the Internet, these individuals should become comfortable using the technology to answer surveys. An additional advantage to e-mail is that respondents can return it in one of three ways: e-mail, fax or postal mail (Parker 1992). This flexibility may enhance the perception of ease of use. Unless the respondent purposely deletes the survey, it cannot be accidentally tossed or misplaced like a mail survey. Yet, comparable to a mail survey, the respondents still have the benefits of completing the survey at their own pace and convenience.

There is not clear evidence that new technology produces a higher response than postal mail. In a review of nine studies that have used both postal mail and e-mail four studies show postal mail achieving higher response rates than e-mail, three studies indicate that e-mail response rates are higher than postal mail, and two studies did not show significant differences in response rates. Researchers indicated that the lack of familiarity with the technology may have impacted some of the response rates. It is also important to note that many of these studies are from small, homogenous populations, and thus may not represent larger population groups' response tendencies.

Past studies found direct marketers can collect data more quickly using e-mail than with postal mail methods. In the five studies that reported response time results, e-mail responses were collected significantly faster than postal mail responses. The variety of populations used in these studies suggest that this rapid rate of response might be seen among larger Internet populations.

Current research has also identified two key limitations unique to e-mail that must be considered when planning an e-mail survey. First, researchers must recognize that unsolicited surveys may be considered aggressive by respondents, and not

in keeping with Internet culture (Mehta and Sivadas 1995). Minimizing a perception of intrusiveness should help to address this problem (Schillewaert, Langerak and Duhamel 1998). Second, the changing nature of the Internet suggests that it is possible that e-mail addresses may become out-of-date fairly quickly (Smith 1997). Addressing this issue early on can prepare the researcher for dealing with delivery failures. . . .

Considerations

. . . It would not be possible to generalize results to mass markets including both Internet users and non-Internet users based on knowledge attained solely from on-line respondents. This has also been shown as a limitation to web based polls (Coomber 1997; Kehoe and Pitkow 1996). However, depending on the research question, it is possible that sample information can be used to generalize to the on-line population.

One of the most challenging limitations is the changing nature of the Internet. The composition of the Internet changes daily with new individuals logging on and others adding or switching Internet service providers. Thus, some directories may contain information that is out of date or incomplete. . . . The changing nature of the Internet is also seen in changes to how search engines operate. Any ownership changes of a search engine or other web content provider may result in unanticipated changes to this methodology. Additionally, the technology allows individuals to set up mail filters, which delete messages from those senders not on the receiver's 'approved' list. This deletion may or may not be reported to the sender. As use of mail filters grows, response rates may be affected. Researchers should anticipate these changes by testing search engines prior to address generation to make sure that the method is still appropriate and pre-testing the study with a random sample of names to determine and plan for non-deliverable mail.

While response rates now appear promising, respondent distrust of data collection may influence response rates in the future. . . . One respondent wrote, 'if you are a student then I am the Emperor of Japan'. The novelty of using e-mail to collect data may be partly to blame. Until this method becomes more ingrained with academics and popularized among on-line users, respondent concern and distrust is likely to continue.

Additionally, individual ISPs have policies and procedures that may limit the success of e-mail surveys. We encountered one ISP that monitored the number of e-mails delivered to its users that originated from a single address. If the number was very large, the ISP assumed that the sender was 'spamming', and the system operator blocked the originator from sending additional messages to the ISP's subscribers. . . .

How government regulation will affect the promise of e-mail remains to be seen. Federal courts have barred specific companies from sending unsolicited e-mail advertisements to subscribers of CompuServe (Kanaley 1997). The courts are ruling that ISPs have the right to restrict access by 'spammers', mostly for economic reasons. Users who pay hourly access rates complain about spending too much time and money reading messages they have no interest in. How this will affect mailing in

the future is not yet clear. Options being discussed include charging mailers a fee for each piece of mail sent. Some believe this will cause companies to be more selective in the addresses to which they send mail. Obviously, this would increase the costs of e-mail surveying. . . .

While e-mail surveying will probably never replace the broad-based data available via postal mail surveys, it will probably provide adequate data for the study of on-line populations, and given the propensity of 'hard to reach' individuals to respond, may provide richer data about on-line behavior than postal mail surveys. As on-line usage continues to grow, and as more and more consumers have access to e-mail, it is conceivable that this method may be eventually used in place of postal mail to gather information about broad-based consumer segments.

References

- Coomber, R. (1997) 'Using the Internet for survey research.' *Sociological Research Online* 2 (2) [On-line]. Available: <http://www.socresonline.org.uk/2/2/2.html>
- Couper, M.P., Blair, J. and Triplett, T. (1997) 'A comparison of mail and e-mail for a survey of employees in federal statistical agencies.' Paper presented at the annual conference of the American Association for Public Opinion Research, Norfolk, Va.
- Kanaley, R. (1997, February 5) 'Judge bars bulk mailer from online.' *Philadelphia Inquirer*, p. C1.
- Kehoe, C., and Pitkow, J. (1996) 'Surveying the territory: GVU's five WWW user surveys.' *The World Wide Web Journal*, 1, (3). [Also on-line]. Available: http://www.cc.gatech.edu/gvu/user_surveys/papers/w3j.html
- Kehoe, C., Pitkow, J., and Morton, K. (1997) *Eighth WWW user survey* [On-line]. Available: http://www.gvu.gatech.edu/user_surveys/survey-1997-10/
- Kiesler, S., and Sproull, L. S. (1986) 'Response effects in the electronic survey.' *Public Opinion Quarterly*, 50: pp. 402-413.
- McCullough, D. (1998) 'Web-based market research, the dawning of a new era.' *Direct Marketing* 61 (8), pp. 36-39.
- Mehta, R. and Sivada, E. (1995) 'Comparing response rates and response content in mail versus electronic mail surveys.' *Journal of the Market Research Society* 17 (4), pp. 429-440.
- Parker, L. (1992) 'Collecting data the e-mail way.' *Training and Development*, July: pp. 52-54.
- Schillewaert, N., Langerak, F., and Duhamel, T. (1998) 'Non probability sampling for WWW surveys: a comparison of methods.' *Journal of the Market Research Society* 4, (40), pp. 307-313.
- Schuldt, B.A., and Totten J.W. (1994) 'Electronic mail vs. mail survey response rates.' *Marketing Research*, Winter: pp. 1-7.
- Smith, C. (1997) 'Casting the net: Surveying an Internet population.' *Journal of Computer Mediated Communication* 3 (1). Available: <http://www.ascusc.org/jcmc/vol3/issue1/smith.html>.
- Steele, T.J., Schwendig, L.W., and Kilpatrick, J.A. (1992) 'Duplicate responses to multiple survey mailings: A problem?' *Journal of Advertising Research* 32 (1): pp. 26-33.