Using the Internet for Survey Research: A Case Study

ARTICLE *in* JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE · JANUARY 2000

Impact Factor: 2.01 · DOI: 10.1002/(SICI)1097-4571(2000)51:13.0.CO;2-W · Source: DBLP

1 AUTHOR:

Yin Zhang
Kent State University
54 PUBLICATIONS **388** CITATIONS

Using the Internet for Survey Research: A Case Study

Yin Zhang
School of Library and Information Science, Kent State University, P.O. Box 5190, Kent, OH 44242-0001.
E-mail: yinzhang@slis.kent.edu

The Internet provides opportunities to conduct surveys more efficiently and effectively than traditional means. This article reviews previous studies that use the Internet for survey research. It discusses the methodological issues and problems associated with this new approach. By presenting a case study, it seeks possible solutions to some of the problems, and explores the potential the Internet can offer to survey researchers.

Introduction

The survey is a widely used research method in the areas of marketing, business, education, psychology, political science, sociology, and social work, as well as in library and information science (LIS) (Babbie, 1990; Busha & Harter, 1980; Church, 1993; Mertens, 1998; Powell, 1997). However, over the years, many drawbacks have been associated with this method. For example, postal mail surveys tend to suffer from low response rates and slow response times (Fowler, 1993; Fox, Crask, & Kim, 1988; Oppenheim, 1992). Today, the most challenging aspect of survey methodology is how to conduct studies efficiently and effectively while retaining validity.

The Internet, popular for its rich sources of information and powerful means of communication, provides survey researchers with many new opportunities. Even though many studies have employed the Internet for survey research, little research has been done on this approach itself. For example, validity and appropriateness of the method have seldom been discussed. It is unclear under what conditions Internet surveys can be effective, what factors may influence their validity, how the implementation of some techniques may improve response rate and quality of survey data, and how respondents react to Internet-based surveys during a survey session. This study reviews previous studies that use the Internet for survey research. It discusses the methodological issues and problems associated with this new approach. By presenting a case study, it seeks a better understanding of using the Internet for survey research, provides and evaluates possible solutions to some of the problems, and explores the potential the Internet can offer to survey researchers.

Received July 1, 1998; revised February 8, 1999; accepted February 8, 1999.

Previous Studies

Although still evolving, the Internet has been used for survey research in a variety of ways and for different purposes. The examples given in this section were largely selected from LIS literature. They serve to outline how the Internet has been used for research in LIS and other related areas. However, the advantages as well as potential problems and concerns of using the Internet for survey research are discussed in general, which should be applicable to other areas as well.

E-mail has been one of the most frequently used Internet tools for survey research because of its convenience. It has been used alone as well as in combination with other methods. In her study of e-mail use for scientific communication, Chu (1994) used e-mail as the only means to collect survey data. In their study of the use of computer-based tools by members of the Modern Language Association (MLA), Shaw and Davis (1996) sent a questionnaire to 1,000 MLA members, 500 by electronic mail and 500 by postal mail. German and Oppenheim (1996) contacted the potential respondents by both e-mail and postal mail to investigate periodical lending policies in university libraries. In a study of a public telecomputing system, Anderson and Gansneder (1995) developed a methodology that used both an e-mail survey and computer-monitored data. Their e-mail survey offered several options for potential respondents to receive, complete, and return the survey via e-mail as well as postal mail so that "users without the technical expertise needed to complete the survey on-line would not be excluded" (p. 38). This multiple-option e-mail survey has also been used in other studies, including Schoch and Shooshan's (1997) research on the use and users of a mailing list MEDLIB-L, and Gill and Yates-Mercer's (1998) research on the use of the World Wide Web by local authorities. In addition,

Roselle and Neufeld (1998) used e-mail at the follow-up stage of a postal mail survey, and they found e-mail was "as effective as postal mail in terms of both the speed and size of the survey response" (p. 153).

Mailing lists and newsgroups can be used efficiently to reach potential respondents because their subscribers usually share common interests or concerns. In most previous studies, questionnaires were sent to multiple mailing lists and/or newsgroups that were most likely to include the target individuals. Researchers have used this approach to study professional activities (Kovacs, Robinson, & Dixon, 1995; Ladner & Tillman, 1993; Powell, 1993; Schiller, 1994), to seek insights and concerns about new technologies (Cronin, Overfelt, Fouchereaux, Manzvanzvike, Mikyeong, & Sona, 1994; Miller, 1994), to investigate use and users of library systems and Internet-related technologies (Bane & Milheim, 1995; Cibbarelli, 1997; Ciolek, 1998; Cronin et al., 1994; Kovacs et al., 1995), and to study the readership of an electronic journal (Berge & Collins, 1996). In addition, mailing lists and newsgroups have been used to broadcast survey requests and to recruit potential respondents (Teo, Lim, & Lai, 1997).

At this point, the World Wide Web (the Web) has not been used as widely as e-mail, mailing lists, or newsgroups in formal research, especially in LIS. The Web provides new opportunities to conduct survey research more efficiently. For example, HTML forms and CGI programs can be used as a front end to receive survey data; hyperlinks can make complicated surveys easy to follow. Researchers have used the Web as the only means to collect survey data. Such examples include a study of users' searching behavior in using a Web search engine conducted by Spink, Bateman, and Jansen (1998). Also, researchers have used the Web in combination with other approaches to collect survey data. In their 1996 national survey of public libraries and the Internet, Bertot, McClure, and Zweizig (1996) sent a questionnaire via postal mail and offered its Web version as an optional completion mode for those libraries with Web access. In a study of users and uses of the Internet in Singapore, Teo, Lim, and Lai (1997) used an interactive Web survey for data collection. To recruit participants, they sent survey requests to several newsgroups and local newspapers while linking the survey to two popular Web sites. Web-based surveys have also been used along with server log analysis and in-person focus groups to conduct user-based assessment of Web-based information services (Dolenko, 1998). In addition, the Web has been used in collecting medical records (Subramanian, McAfee, & Getzinger, 1997), studying drug dealers (Coomber, 1997a, 1997b), and conducting laboratory experiments in behavioral and social sciences and psychology (Piper, 1998; Smith & Leigh, 1997).

Compared to conventional mail surveys, the advantages of Internet-based surveys can be summarized as follows: (a) the research costs for sending questionnaires and coding data are relatively low for Internet-based surveys (Berge & Collins, 1996; Goree & Marszalek, 1995; Kiesler & Sproull, 1986; Parker, 1992; Schmidt, 1997; Sproull, 1986). (b) Internet-based surveys usually have a short turnaround time (Berge & Collins, 1996; Goree & Marszalek, 1995; Kiesler & Sproull, 1986; Parker, 1992; Schmidt, 1997; Sproull, 1986). (c) They reach potential respondents in geographically remote areas readily (Kiesler & Sproull, 1986; Parker, 1992; Roselle & Neufeld, 1998; Sproull, 1986). (d) When a research topic is of a sensitive nature, Internet-based surveys offer a means to reach a group that is normally difficult to identify or access, such as drug dealers or gay, lesbian, and bisexual college students (Coomber, 1997a, 1997b; Goree & Marszalek, 1995; Marszalek & Goree, 1995). (e) They offer a means to efficiently survey larger numbers of individuals (Schmidt, 1997). (f) They may increase respondents' motivation to participate by providing a dynamic/interactive survey process (Kieley, 1996; Schmidt, 1997; Teo et al., 1997). (g) They may reduce errors from transcription and coding. In Internet-based surveys, most responses are in electronic format and have been pre-coded.

Potential problems and concerns unique to Internet-based surveys include the following:

Biased Sample and Biased Return

Biased samples and biased returns are a major problem with Internet-based survey research. Currently, respondents to Internet-based surveys are most likely to be individuals who have access to computer networks, who have the skills to use the survey tools, and who accept and feel comfortable with Internet surveys (Berge & Collins, 1996; Kiesler & Sproull, 1986; Parker, 1992; Sproull, 1986). Thus, findings from Internet-based surveys may not be generalizable to the whole population or even to the target sample if these limitations exist. However, the exact amount of the bias is essentially unknown due to many factors discussed below.

Access to the Internet and Survey. First, individuals in a population or sample may not have equal access to the Internet. Studies of Internet demographics indicate that, currently, certain social groups are underrepresented in Internet users. These underrepresented groups include women, people of limited financial resources, members of some racial and ethnic minorities, people at low education levels, and older age groups (CommerceNet & Nielsen, 1995, 1996, 1997a, 1997b; GVU, 1994a, 1994b, 1995a, 1995b, 1996a, 1996b, 1997a, 1997b). Second, not all individuals in a target population may have access to the specific software applications required for the survey. For example, a respondent needs a Web browser with appropriate plug-ins and/or applications to participate in a Web survey that involves multimedia messages. Third, given the fact that survey requests and questionnaires may be publicized and distributed in different ways, not everyone in a sample may receive the survey. In their study of readership of an electronic journal, Berge and Collins (1996) indicated that those readers, who read individual articles distributed by

libraries or shared by colleagues, might not be reached by the survey posted to multiple related mailing lists. Another possible limitation, which Berge and Collins did not mention, is that the survey might not reach readers who merely relied on alternative delivery modes (FTP and Gopher) of the electronic journal.

Comfort with the Internet survey format. It is noted that in surveys conducted on the Internet, whenever researchers offered multiple options for receiving and/or replying to surveys, some respondents chose to use the conventional means of completion—completing surveys on paper (e.g., Anderson & Gansneder, 1995; Berge & Collins, 1996; Bertot & McClure, 1996). If researchers use the Internet as the only means for collecting data, the representativeness of respondents may be questionable because the technological backgrounds of intended individuals and their access to and use of the network may vary widely (Anderson & Gansneder, 1995; Berge & Collins, 1996). Bertot and McClure (1996) surveyed the respondents who chose to reply to a print version of a questionnaire that was also available on the Web. The responses indicated that comfort with the survey format was a concern. Further, Kittleson (1995) reported that even active Internet users could prefer to answer a postal mail survey, and a considerable proportion of the respondents (33%) indicated that they were not comfortable answering the survey via e-mail. Kawasaki and Raven (1995) reported that users who had easier and less costly access to e-mail could return a survey by e-mail less readily. Thus, it cannot be assumed that Internet users prefer to complete surveys conducted on the Internet. Comfort with the survey format is still a concern.

Effect of self-selection in Internet-based surveys. Currently, most Internet surveys, especially those conducted via mailing lists, newsgroups, or public Web sites, mainly depend on self-selected respondents. However, how much these respondents represent the target population is usually unknown (Berge & Collins, 1996; Cronin et al., 1994). Some researchers have mentioned the possible bias of self-selection (e.g., Dolenko, 1998); some have conducted research to investigate the possible bias. By comparing the self-selected and randomly selected respondents in a computer network survey, Walsh, Kiesler, Sproull, and Hesse (1992) reported a difference in experience with networks between the two groups of respondents. Another key issue of self-selection is whether the group that would not participate differs in some manner from the group that would participate. Anderson and Gansneder (1995) used computer-monitored data to compare respondents and nonrespondents in their use of a computing system. Their results indicated that "respondents were more likely to use the system more often and for more time than were non-respondents" (p. 41).

Validity of respondents. In most Internet-based surveys, survey messages are very likely to reach unintended individuals. For example, participants can forward questionnaires to other unintended groups (Goree & Marszalek, 1995); also Web-based surveys posted at public Web sites are potentially reachable by anyone who has a Web browser with Internet connection. Yet there has not been an effective way to screen unintended participants.

Multiple responses from the same respondent. For Web-based surveys, particularly, participants can easily submit their replies many times. Consequently, the overall results may overrepresent these respondents. And this unequal representation problem has been mentioned by some researchers (e.g., Schmidt, 1997).

Variation and Difficulty in Reporting Response Rate

A disputed issue associated with Internet-based survey research is calculation of response rates—researchers have reported response rates in a variety of ways in current studies. In a traditional postal mail survey, when calculating response rate, researchers do not exclude unreachable respondents from a sample because they usually do not know. Internet-based surveys, however, usually allow researchers to recognize a large portion of undeliverable surveys or unreachable potential respondents. Consequently, Internet survey researchers have reported response rates by excluding cases of unreachable individuals from a sample (Anderson & Gansneder, 1995; Bertot & McClure, 1996; Chu, 1994). Furthermore, Swoboda, Muhlberger, Weitkunat, and Schneeweiss (1997) suggested that "under-reported undeliverable mail" existed in their study using an e-mail survey; they calculated the response rate by excluding those respondents associated with undeliverable e-mails from the sample as well as an additional 30% of estimated underreported portion.

Another cause for variation in reporting survey results in Internet-based survey research is difficulty in determining the size of a sample. In some cases, it is very difficult or even impossible for researchers to estimate the size if a survey is conducted on multiple related mailing lists or newsgroups, and if there is overlapping among subscribers (Bane & Milheim, 1995; Berge & Collins, 1996; Ciolek, 1998; Kovacs et al., 1995; Schiller, 1994). Therefore, instead of reporting the response rate, some researchers have only reported the number of responses (e.g., Bane & Milheim, 1995; Cibbarelli, 1997; Cronin et al., 1994; Miller, 1994; Powell, 1993).

Impersonalized Survey Request

Currently, it is usually hard and sometimes impossible to locate personalized contact information of potential respondents in Internet-based surveys. Survey research conducted in impersonalized communication, such as posting the survey on mailing lists, newsgroups, or public Web sites, usually has had relatively low response rates (e.g., Berge & Collins, 1996; Ciolek, 1998; Kovacs et al., 1995; Spink et

al., 1998). Some studies have mentioned techniques to acquire individual e-mail addresses from listservs (Berge & Colins, 1996; Schoch & Shooshan, 1997). Swoboda et al. (1997) developed a procedure to automatically retrieve e-mail addresses of newsgroup subscribers, but the addresses were limited to individuals who had posted message(s) during a certain time period. Most importantly, these efforts did not go beyond getting individual e-mail addresses; the names and titles of subscribers remained unknown.

More Expertise Required of Survey Researchers

Internet-based survey research usually requires more technical expertise than research conducted with traditional survey methods because of the nature of Internet-based surveys. For instance, Internet-based surveys are usually self-administered; respondents tend to "drop out" before they actually complete the surveys because of the lack of human contact (Dolenko, 1998). Besides, a large portion of survey replies are unusable because many respondents do not provide answers to some questions (Spink et al., 1998). Researchers have suggested to embed some advanced functions in Web-based surveys to improve response rate and quality of survey data. These advanced functions include reducing duplicate submissions from the same respondent, using a dynamic/interactive survey process to motivate potential respondents to participate, and reducing incomplete responses and unacceptable data (Schmidt, 1997; Teo et al., 1997). However, these features have seldom been implemented in previous studies because they require additional technical expertise of survey researchers. The exact nature and effect of these features are still unclear.

In summary, previous studies provide support for, as well as show problems with, the use of the Internet for survey research. Empirical studies addressing methodological issues are still lacking (Bertot & McClure, 1996; Walsh et al., 1992). In addition, we need a better understanding of how respondents react to Internet-based surveys during a survey session, how the principles of traditional survey methods can be applied to Internet-based surveys, and what implementation techniques are appropriate for data collection in Internet-based surveys.

A Case Study

For a study of scholarly use of Internet-based electronic resources, 203 researchers with in-press[1] papers to be published in eight LIS journals, both print and electronic, were selected for a survey research. This case study presents how the Internet, specifically e-mail and the Web, was used in the survey research. The details of survey sampling have been discussed elsewhere (Zhang, 1998; Zhang & Estabrook, 1998) and will not be repeated here. This study did not use the Internet for its survey sampling. Rather, it used the Internet as a tool at various stages and for different purposes, including presurvey preparation, survey pretest, data collection, and communication with potential respondents. In addition, respondents' reactions to the Web survey during the survey session were recorded in transaction log files, which would enable us to evaluate the advantages and disadvantages of the implementation techniques applied in this study. Using the survey data, this study compared the two groups of respondents who chose to reply to the survey via the Web and via postal mail or fax for their differences in age, gender, Internet experience, self-perceived overall ability to use the Internet, frequency of Web use, and access to the Web.

Method

Survey Instrument

A survey questionnaire was designed to collect data that would help us obtain a better understanding of the scholarly use of Internet-based e-sources among LIS researchers. The data included researchers' demographic information, current frequency with which they used Internet tools and protocols, means of obtaining access to various Internet tools and applications, strategies for locating e-sources for their research, opinions on citing e-sources, evaluation of Internet-based e-sources for research, and suggestions for improving their use of e-sources for research.

Web Survey Application

A Web survey application, whose system architecture is illustrated in Figure 1, was developed in Perl on an HP workstation with UNIX for this research project by the author. It was used to (a) generate personalized cover letters and questionnaires with hyperlinked e-sources where applicable, (b) collect survey data via the Web, (c) serve as a data entry interface for survey data received via postal mail or fax, and (d) generate log files for trouble-shooting, use and user analysis of the survey application, and evaluation of the system design.

The components, functions, and working procedure of the system are described as follows (see Fig. 1):

Authentication. A unique case ID is assigned to each potential respondent and is included in print copies of cover letters and questionnaires, and e-mail survey requests sent to respondents. Once a respondent submits his/her case ID via the survey main page (an HTML form), a CGI program will check the authorized ID list. If the ID is not on the list, neither the questionnaire nor any other requested information will be displayed. This module limits access to the survey page to authorized participants.

[1] In-press is defined as the status of papers that had bee acceptd for publication, but were not yet published as of July 1, 1997, in the eight LIS journals.

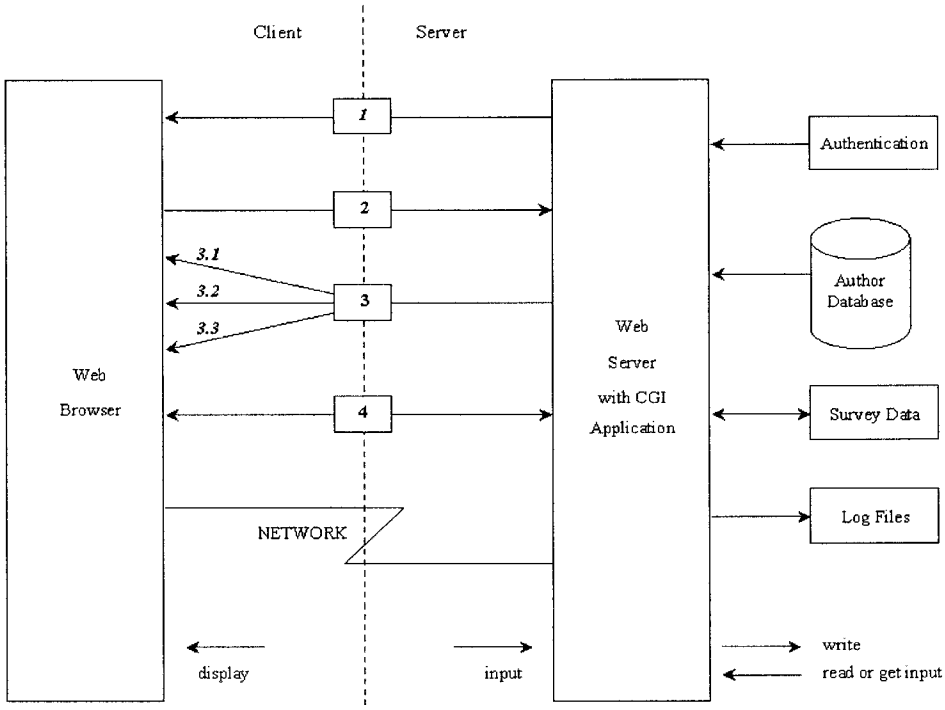

FIG. 1. The Web-based survey system architecture.

Author database. Part of the larger project is to investigate how scholars find e-sources and what factors are involved in their decisions on citing e-sources. Some questions in the questionnaire ask respondents to recall the related activities involved in writing the papers in the sample. Thus, this part of the questionnaire differs among respondents who have e-source citations, and needs to be personalized. The author database contains information about the authors and their papers, which is used to generate personalized cover letters and questionnaires. The database fields include: case ID, author's title and name, title of paper, source of paper (e.g., journal name, year, volume, issue, and page numbers), journal format, and e-citations if there any.

Survey data. Survey records are created or updated after the system checks the completeness of the replies based on the internal logic of each reply and confirms the results with respondents. Participants can view their own responses recorded in these data files.

Log files. The system log files generated automatically by the Web server (NCSA HTTPd NCSA/1.5.2a) are machine or host oriented, they cannot reflect accurately some user-related information. For example, they can tell which machines or hosts access the system, but they cannot tell if a respondent accesses the system from different machines or if several respondents access the system from the same machine. This Web survey application was designed to record and link all the transactions associated with each respondent during the survey process. The data were used to evaluate the design of the survey application and to investigate respondents' behavior during the survey session. One such example was case ID design evaluation. After a respondent submitted a case ID, this transaction was recorded regardless of whether the ID was valid or not. Subsequently, this information was used to evaluate the design of the case IDs by answering several questions, including: How many times did it take a respondent to enter the ID correctly, and how many respondents stopped at an unsuccessful ID entry?

Procedure. With a Web browser, a respondent can interact with the remote system in the following way (see Fig. 1):

Step 1: Load the survey main page. Once a respondent locates the survey Web site with a Web browser, the Web browser displays the survey main page, which is an HTML form, asking for the respondent's case ID and one of the following three options: complete the survey, view personal previous survey replies, or view the overall survey results.

Step 2: Input survey request. At this point, the respondent can enter his/her unique case ID, choose an option, and submit the form. If the respondent enters a wrong case ID, this process will repeat itself until the respondent enters the correct one. This prevents unintended respondents from accessing and completing the survey.

Step 3: Receive requested page. After receiving a valid case ID and an option for access, the CGI program delivers one of the following pages requested:
Option 1: a personalized cover letter and questionnaire form,
Option 2: the previous survey replies associated

with the case ID, or
Option 3: the overall survey results.[2]

Step 4: Complete the survey through an interactive process. Upon submission of the survey form, a step after 3.1 (option 1 in Step 3), the respondent is reminded of any items that may have been left blank. The respondent can then go back to fill in these fields. Once the respondent completes the survey form through this interactive process, a message appears on the screen indicating that the survey has been successfully completed and that the respondent can print a copy of the finished survey form and view the created survey record.

Survey Procedures

Presurvey preparation. On November 9 and 10, 1997, 37 authors with in-press papers, whose bibliographies were not available at that time, were contacted via e-mail for a complete reference list cited in their papers. With one follow-up 10 days after the initial request, a 100% response rate was achieved by January 15, 1998.[3] Most authors provided the data via e-mail, while seven authors (including an e-journal author) faxed or mailed the data, even though they had initially responded via e-mail to give a positive feedback. The results of the initial contact suggest that e-mail is an efficient way to communicate with this group of authors. More than half of the authors sent at least two e-mail messages asking questions or giving positive feedback before they sent the data. For the authors whose papers were published at the time, the bibliographical information of their papers was retrieved from the *Social Science Citation Index* database, and then included in the author database of this study.

For authors whose e-mail addresses were unknown, Yahoo White Page and Web sites of their affiliations were used to locate their e-mail addresses. In addition, some authors were asked for help in locating the unknown e-mail addresses of their co-authors. When the survey started in late February 1998, six of the 203 authors' e-mail addresses were still unknown.

Survey pretest. In early February 1998, the questionnaire and the on-line survey application were pilot tested and evaluated by six LIS scholars with backgrounds similar to the intended potential respondents; their feedback was incorporated into the final version of the questionnaire and the design of the survey application.

Initial mailing. On February 20 and 21, 1998, an initial invitation to participate in the survey was sent via e-mail to the 197 authors whose e-mail addresses were available. At the same time, print copies of the questionnaire were sent via postal mail to the six authors whose e-mail addresses were unavailable. In the following 7 weeks, 19 print copies were sent via postal mail or fax to the authors who requested a print copy. A self-addressed envelope with return postage was enclosed with all print copies of the questionnaire.

Follow-ups. There were three follow-ups: the first was sent to nonrespondents via e-mail on March 7, 1998; the second was also sent via e-mail on March 30; the third was sent via both e-mail and postal mail on April 30 and May 1, 1998.

Survey Data Processing

In this study, survey replies received via postal mail or fax were entered via an interface similar to the Web-based questionnaire described above. The only difference is that the transactions generated were not recorded in user log files. Once all the survey data were in electronic format, several programs written in Perl were used to transform the data into a format acceptable by Microsoft Excel—an acceptable format for most statistical software packages—for further data analysis. All the close-ended questions were pre-coded in the HTML questionnaire form; no further coding was needed. The answers to the open-ended questions were extracted from survey records for content analysis.

Results

Response Rate

Of the 203 potential respondents, two were unreachable due to affiliation change (there might be some other unknown cases). This reduced the target sample size to 201. By this study's cutoff date, June 15, 1998, 125 usable replies were received via the Web and 31 usable replies via postal mail or fax. Twelve potential respondents declined to participate. Eleven responded to the survey request, but their replies either were not received or were unusable.[4] In total, 179 researchers responded to the survey, and 156 completed replies were usable. With 201 potential respondents, the overall response rate was 89.1% (179 of 201), and the usable reply rate was 77.6% (156 of 201).

The daily cumulative responses by response channel are illustrated in Figure 2. The first two follow-ups sent via e-mail on days 14 and 37 and the third one via both e-mail and postal mail over days 68 and 69 generated obvious new waves of replies via the Web in the next few days right after the requests were sent, but the number of replies returned decreased with each follow-up. In contrast, the effect of the

[2] The preliminary survey results have been available at the survey Web site since April 30, 1998.

[3] One respondent reported that he had been out of the country and had no access of e-mail for 2 months during the survey period, which delayed his reply until January 15. In fact, by December 8, 1997, all the rest of the respondents had provided the data.

[4] One completed reply from a non-Internet user was unusable because most survey questions did not apply to the respondent and were left blank.

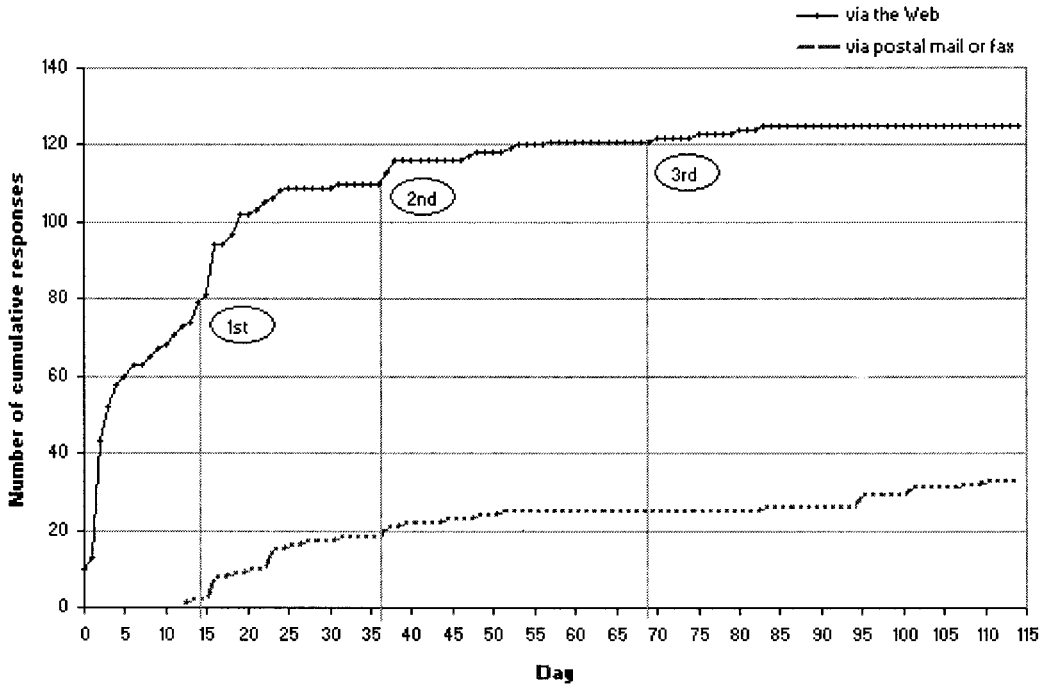

FIG. 2. Daily cumulative responses by response channel.

follow-ups was not very clear for responses received via postal mail or fax because it took a relatively long turnaround time and only a small number of respondents chose to reply in this way. It was interesting to note that the third follow-up generated a better wave of replies via postal mail. This suggests that this group of respondents would have preferred a print copy of the questionnaire—they had ignored the three e-mail follow-ups before they replied to the survey in print format.

Characteristics of Respondents

Respondents were roughly balanced by gender, 53.9% male and 46.1% female. Equally 6% of respondents were under age 30 or over 60, 24.7% were 30–39, 34.7% were 40–49, and the remaining 28.7% were 50–59 (Fig. 3). The average age of respondents was 45.4 years old. The overwhelming majority (94.2%) of respondents were p-journal paper authors, while only 5.8% were e-journal paper authors (Table 1). Given the sample distribution by journal format in which the potential respondents chose to publish their research (Table 1), it is not surprising to see this difference. Overall, 29.0% of respondents were authors who cited e-sources, while 71.0% of respondents were authors who did not cite e-sources in the papers (Table 1), even though they might have cited e-sources elsewhere.

In terms of technological background, respondents were asked how many years they had been using the Internet (including e-mail) and to rate their overall ability to use the Internet. Roughly half (46.1%) of respondents indicated that they had 5–9 years of Internet experience (Fig. 4), while 27.3% indicated that they had 10–14 years of experience. These two groups comprised 73.4% of the total respondents. Approximately 14% of respondents indicated that they had less than 5 years of Internet experience, while 12.3% indicated that they had over 15 years of Internet experience. On average, respondents possessed 8.9 years of Internet experience.

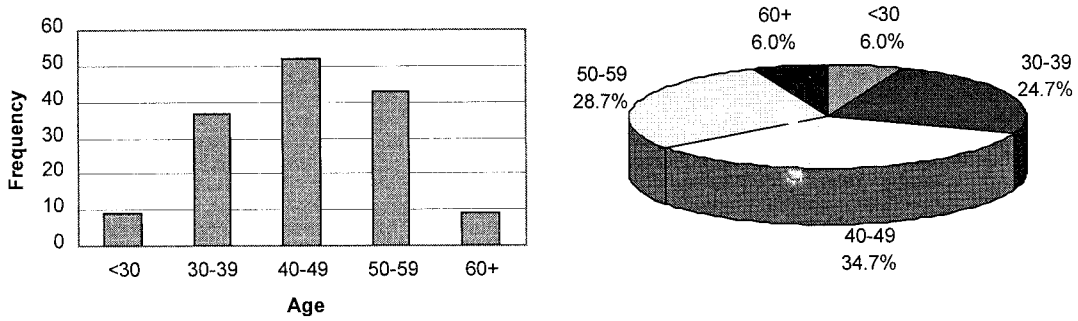

FIG. 3. Age distribution of respondents (n = 150).

TABLE 1. Sample and respondents by journal format, and presence/absence of e-citations in the sample articles.

	Sample		Respondents	
	Frequency	Percent	Frequency	Percent
e-journal paper author	14	7.0%	9	5.8%
p-journal paper author	187	93.0%	147	94.2%
Total	201	100.0%	156	100.0%
with e-citation	65	32.3%	45	29.0%
without e-citation	136	67.7%	111	71.0%
Total	201	100.0%	156	100.0%

Respondents' overall ability to use the Internet was rated on a five-point scale from 1 (beginner) to 5 (expert). Nearly half (47.4%) of respondents rated themselves as having "above average" overall ability to use the Internet, one-fifth (20.5%) as "average," 29.5% as "expert," and 4.4% as "below average"; no respondents rated themselves as "beginner" (Fig. 5). That is, 95.6% of respondents rated themselves as having at least average ability to use the Internet; only 4.4% rated themselves as "below average."

Comparison of Respondents by Response Channel

Using the survey data, this study compared the respondents who chose to reply to the survey via the Web with those who chose to reply via postal mail or fax for their differences in age, gender, Internet experience, self-perceived overall ability to use the Internet, frequency of Web use, and access to the Web. The following differences and relationships were found: (a) self-perceived overall ability to use the Internet was related to the choice of response channel ($\chi^2 = 7.602$, $p = 0.055$). Respondents who chose to reply via the Web had a higher self-perceived overall ability to use the Internet, and the difference was at a statistically significant level. (b) Web use was related to the choice of response channel ($\chi^2 = 19.020$, $p = 0.001$). Respondents who chose to reply via the Web used the Web more frequently than did respondents who replied via postal mail or fax. (c) There was a difference in age between respondents who chose to reply via the Web and those who replied via postal mail or fax (44.0 and 51.3 years old, respectively, as group age means; $F = 11.977$, $p = 0.001$). (d) There was a slight difference in the years of Internet experience between the two groups of respondents. However, the difference was not statistically significant (9.1 and 7.9 years, respectively, as group means of Internet experience; $F = 1.559$, $p = 0.214$). (e) Web access was not related to the choice of response channel ($\chi^2 = 1.955$, $p = 0.376$). The two groups of respondents had almost the same number of access points to the Web (from work, from home, and from libraries). (f) Gender was not related to the choice of response channel ($\chi^2 = 1.192$, $p = 0.275$).

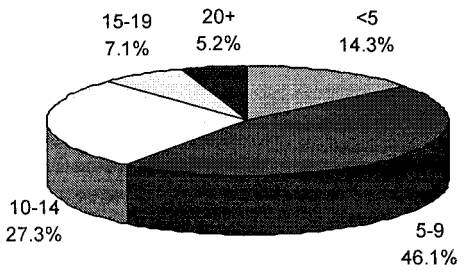

FIG. 4. Years of Internet experience ($n = 154$).

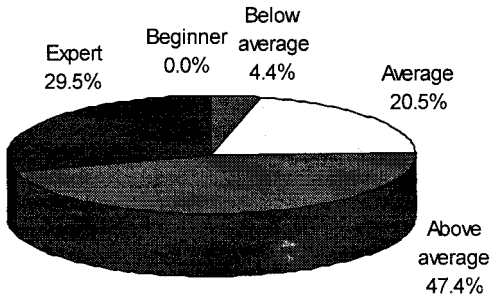

FIG. 5. Self-perceived overall ability to use the Internet ($n = 156$).

These findings echo results from previous studies on the characteristics of Internet survey respondents. That is, respondents tended to be younger and more skilled at using the Internet. These findings also suggest that the selection of response channel, a critical factor in participation in Internet-based surveys, might not merely depend on respondents' technological backgrounds or on their access to the Internet facilities, in this case, the Web. In fact, some respondents who were experienced and frequent Internet users chose to reply via postal mail or fax. Acceptance of the survey format might be a concern for these respondents.

Respondents' Reactions During the Web Survey Session

The transactions of each respondent during the Web survey session were recorded in several log files, which were generated by the Web server as well as by the Web survey application. The data were used to reconstruct the Web survey session of each respondent to obtain a better understanding of how respondents reacted to Internet-based surveys. The data were also used to evaluate the design and implementation techniques of the Web survey application.

Access and completion. Overall, 147 potential respondents attempted to access the Web survey. Most of them (132 or 89.8%) tried to finish the Web survey. The other 15 (10.2%) did not get to the stage of submitting the survey. Ten stopped at the questionnaire page; of these 10, five returned the survey via postal mail, three responded via e-mail to decline to participate, and two did not offer further feedback. Four of the 15 stopped at the main survey page without entering a case ID. One of the 15 reportedly stopped after an unsuccessful attempt to access the survey Web site.

Of the 132 respondents who attempted to finish the Web questionnaire, the majority (125 or 94.7%) finished it successfully, with two succeeding on their second try. Eight respondents (6.1%) submitted the questionnaire, but their replies were not created successfully on their first try. After becoming aware of the loss in the follow-up communications, two of the respondents redid the Web questionnaire and succeeded on the second try; two chose to redo the questionnaire on paper; and four gave up. One respondent reported having problems submitting the finished questionnaire, and chose to redo it on paper.

Case ID entry. In the pretest, 16-digit case IDs were used; however, some participants indicated that it was hard for them to type in the IDs correctly. Therefore, 10-digit IDs, appended with a lowercase letter, were used in the final design (e.g., 1234567890a). A Perl program was used to generate the initial 10-digit random numbers. These numbers were then sorted and adjusted to a coding scheme that included the journal format, journal name, multiple authorship, and status of presence/absence of e-citation. Lastly, the 10-digit numbers were appended with a lowercase letter to form the case IDs for the study. Among the 142 (132+10) respondents who tried their case IDs, 130 (91.6%) succeeded on their first try, 10 (7.0%) succeeded on their second try, and the other two (1.4%) succeeded after they tried four or more times. It was interesting to note that no respondents gave up once they tried to enter their case IDs; that is, no one stopped at an incorrect case ID entry. In addition, there was no indication of mixed use of case IDs; each respondent used the same ID throughout the study. This suggests that the design of the case IDs fits a sample size near 200 and could be even simpler.

The pattern of options selected. Research has suggested that disclosing survey results to respondents serves as an incentive for them to participate in postal mail surveys (Mertens, 1998). Bertot and McClure (1996) suggest that giving respondents a copy of their reply to an electronic survey might motivate them to participate. To examine the effects of these two incentives, this study offered respondents the opportunity to view their own replies (option 2 from the survey main page) and the overall survey results (option 3) in addition to completing the survey (option 1).

The results from this study suggest that offering the survey results might have served as an incentive for respondents to participate, especially for the Web survey respondents. Among the 125 respondents who replied via the Web with success, 76 (60.8%) requested to complete the survey only (option 1); four (3.2%) chose to reply to the survey and to view their own previous survey replies (option 1 and option 2); 33 (26.4%) chose to complete the survey and view the overall survey results (option 1 and option 3); 12 (9.6%) chose all the three options. That is, 16 (12.8%) of the Web survey respondents have viewed self previous survey replies, and 45 (36.0%) have viewed the overall survey results. For the 31 respondents who completed the survey via postal mail or fax, however, none of them chose an option other than option 1, even though some had indicated that they were very interested in the overall survey results. At this point, it is unclear what prevented these respondents from viewing the results.

Reactions to advanced functions for Web-based surveys. Some advanced functions, such as checking completeness of responses, reducing duplicate submissions from the same respondent, and reducing invalid answers, can be embedded in Web-based surveys to facilitate data collection and to improve the quality of survey data (Schmidt, 1997; Teo et al., 1997). However, these functions have seldom been included in previous studies. The nature and effect of these functions were unclear; their implementation guidelines were given less attention. The Web survey application in this study was designed to incorporate some of these functions, and respondents' reactions were recorded for evaluation.

(1) Incomplete responses. The CGI program in the Web survey application can check the completeness of each reply. The results, a list of unanswered questions or items evaluated based on the internal logic of each reply, would be shown on the screen after each submission. A respondent could then go back to fill in these fields and submit the survey again. During the first 11 days of the survey, the Web survey application had been designed in such a way that incomplete replies were not acceptable. After receiving suggestions from some respondents, the application was revised to accept incomplete replies as well from the 12th day.

One issue that arose from the initial design was that, as some respondents suggested, it might generate unintended answers to the questions that were not applicable to some respondents. To address this issue, this study compared the multiple submissions received under the initial design with the multiple submissions received after the design was revised. Among the 71 respondents who replied to the survey via the Web under the initial design, 21 (29.6%) supplied one or more subsequent answers to the questions they had missed in their previous submission(s). In addition, replies from eight respondents were lost—the respondents had submitted the replies, but their survey records were not created. It was quite likely that the lost replies were incomplete and, thus, unacceptable to the Web survey application, but some lost replies might also have been caused by system failure during that period. On the other hand, among the 56 (two on their second try) respondents who replied to the survey after the design was revised, 14 (25.0%) respondents provided answers to the fields they had missed in their previous submission(s).

Web-based surveys do provide opportunities to improve the quality of survey data through a dynamic/interactive process, which postal mail surveys cannot achieve. However, to take advantage of these opportunities, implementation techniques play a very important role. This study sug-

gests that the revised design proved almost as effective as the initial design in reducing incomplete responses and more effective in that it allowed respondents to skip those questions that were not applicable to them.

(2) Multiple submissions. As mentioned earlier, a problem associated with Web-based surveys is that respondents may submit the same reply several times. As a result, the overall survey replies may overrepresent these respondents. This study shows that this problem can be solved by providing each respondent with a unique case ID to separate replies from different respondents but to group replies from the same respondent together. This technique is especially useful for surveys where multiple submissions are necessary. For example, in this study, when respondents wanted to supply subsequent answers to previously missed fields, they needed to submit several replies. All submitted replies needed to be kept rather than to be overwritten by subsequent ones, because some respondents chose to provide subsequent answers by restarting the questionnaire, filling in and submitting the missed items only.

(3) Unacceptable responses. Web-based surveys can be designed to check and reduce unacceptable answers to some questions when the responses are expected to be in a specific format (Schmidt, 1997). For example, if a question is asking a respondent's age or years of experience, a valid response is most likely to be an integer less than 100. This is especially the case if survey researchers have suggested that respondents submit their replies in such a format in the question. Thus, if a response is in a format other than a numeral, or is a number out of the range, it is very likely to be an invalid answer. However, researchers need to be very cautious when taking advantage of this capability. Observations from this study show that respondents may answer a question in a variety of ways rather than in a given format. For instance, for the question about their years of Internet experience in this study, some respondents gave answers like ~5, 15?, 3–4, and >20. As such, a predefined response format is helpful to achieve uniformity of data, which will reduce the workload in data cleaning and processing. However, a flexible format may be more respondent-friendly.

Nature of E-mail Correspondence in This Study

E-mail was used extensively in this study and proved very effective for communicating with respondents and nonrespondents at various stages. Besides being used to collect bibliographic data at the stage of presurvey preparation, it was also used to (a) ask potential respondents about their co-author's contact information, (b) send the initial survey request and follow-ups, (c) provide answers to inquiries from respondents, (d) clarify some survey questions, (e) verify unclear responses, (f) seek and exchange research ideas and information, and (g) get feedback from respondents about their participation status.

Discussion and Conclusions

The Internet provides survey researchers with many new opportunities. This study offers evidence that Web-based questionnaires are of great potential value for LIS research. In fact, in this study, 80% (125 of 156) of the usable replies were received via the Web. However, at this stage, the Internet cannot serve as the only means to collect survey data if researchers need representative returns from a sample. This is even true for Internet-related studies, in which potential respondents are Internet users. In this study, for instance, 20% of respondents chose to complete the survey via postal mail or fax. Selection of response channel may be influenced by age, Web use (for a Web-based survey in this case), self-perceived ability to use the Internet, and comfort with the survey format. To reduce possible biased returns, researchers need to provide alternative ways to distribute surveys and to receive survey replies. By using the Web in combination with e-mail, postal mail, or fax, researchers can take advantage of the Internet's unique capabilities, and at the same time, reduce the risk of limiting responses to certain groups of individuals in a sample.

Internet-based surveys are still evolving; their full potential appears not yet to be realized. This study suggests that Web-based surveys can be designed to create user-centered transaction logs for trouble-shooting, use and user analysis, and system evaluation. Host-centered transaction logs generated automatically by most Web servers have limitations in reconstructing a complete survey session of each respondent in situations when a respondent obtains access to a survey site from different hosts or several respondents obtain access from the same host. In contrast, user-centered logs can link all the transactions of each respondent together and provide a better picture of each survey session. The data can also help researchers to communicate with respondents more efficiently. In fact, in this study, the logs proved to be very useful for trouble-shooting unsuccessful access and incorrect case ID entries. In addition, it is possible for Web-based surveys to dynamically generate personalized cover letters and questionnaires and to create individualized survey records that serve both respondents and survey researchers. With a properly written program, a Web-based survey can check the completeness of each reply based on its internal logic and give feedback through an interactive process so that inconsistent or missing data can be reconciled at the point of data collection.

Internet-based surveys provide great opportunities but pose challenges as well. One challenge is the presentation and interpretation of survey results, such as representativeness, validity, and response rate. Another challenge is that Internet-based surveys are usually more machine dependent. Certain factors that may influence response rate are beyond the control of survey researchers, such as the software and hardware respondents use to complete a survey, the network traffic when respondents get connected, and so on. In this study, respondents reported problems on the layout of the survey questionnaire on low resolution mon-

itors, problems of going back to previous parts of the questionnaire, problems of printing a completed questionnaire, and problems of loading the questionnaire from home computers via low-speed modems. Even though a questionnaire can be pretested on multiple platforms and with various browsers, still it may be limited by the resources available at the researchers' end. Further, Web surveys usually lack the readiness for respondents to mark up or make comments. In this study, some respondents typed such information in the fields of other open-ended questions nearby. Although a field might have been added at the end of each question for this need, it would be at the price of a quick load of the questionnaire, especially from a low-speed end.

Perhaps the most challenging aspect of using the Internet for survey research is that there have not been many research guidelines. More research is needed to expand our understanding of this new approach and to explore the full potential the Internet can offer to survey researchers.

Acknowledgments

The author would like to thank the members of her dissertation committee, Dr. Leigh S. Estabrook, Dr. Linda C. Smith, and Dr. Caroline A. Haythornthwaite, for their guidance. Dr. Edward Lakner at the Library Research Center of the University of Illinois provided valuable comments on the design of this study. The ALISE Research Committee and the jurors of the 1999 ALISE Methodology Paper Award provided valuable comments on an early version of this paper. Special thanks to all those who participated in the pretest and the survey of this study. Finally, the help from the editors and editorial staff of the eight LIS journals is gratefully acknowledged.

References

Anderson, S.E., & Gansneder, B.M. (1995). Using electronic mail surveys and computer-monitored data for studying computer-mediated communication systems. Social Science Computer Review, 13(1), 33–46.

Babbie, E. (1990). Survey research methods (2nd ed.). Belmont, CA: Wadsworth Publishing.

Bane, A.F., & Milheim, W.D. (1995). Internet insights: How academics are using the Internet. Computers in Libraries, 15(2), 32–36.

Berge, Z.L., & Collins, M.P. (1996). "IPCT Journal" readership survey. Journal of the American Society for Information Science, 47(9), 701–710.

Bertot, J.C., & McClure, C.R. (1996). Electronic surveys: Methodological implications for using the World Wide Web to collect survey data. In S. Hardin (Ed.), Proceedings of the 59th Annual Meeting of the American Society for Information Science (pp. 173–185), Medford, NJ: Information Today.

Bertot, J.C., McClure, C.R., & Zweizig, D.L. (1996). The 1996 national survey of public libraries and the Internet: Progress and issues (final report). Available at: http://istweb.syr.edu/~mcclure/nspl96/NSPL96_T.html.

Busha, C.H., & Harter, S.P. (1980). Research methods in librarianship: Techniques and interpretation. New York: Academic Press.

Chu, H. (1994). E-mail in scientific communication. In M. E. Williams (Ed.), Proceedings of 1994 National Online Meeting (pp. 77–86), New York: Learned Information.

Church, A.H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. Public Opinion Quarterly, 57, 62–79.

Cibbarelli, P. (1997). Cibbarelli's surveys: User ratings of Geac's Advance Integrated Library system. Computers in Libraries, 17(8), 38–39, 42.

Ciolek, T.M. (1998). The scholarly uses of the Internet: 1998 online survey. Available at http://www.ciolek.com/PAPERS/InternetSurvey-98.html.

CommerceNet, & Nielsen Media Research (1995, 1996, 1997a, 1997b). CommerceNet/Nielsen Media Research: Internet Demographic Study. Available at: http://www.nielsenmedia.com/interactive/commercenet.

Coomber, R. (1997a). Using the Internet for survey research. Sociological Research Online, 2(2). Available at: http://www.socresonline.org.uk/socresonline/2/2/2.html.

Coomber, R. (1997b). Dangerous drug adulteration—An international survey of drug dealers using the Internet and the World Wide Web (WWW). The International Journal on Drug Policy, 8(2), 71.

Cronin, B., Overfelt, K., Fouchereaux, K., Manzvanzvike, T., Mikyeong, C., & Sona, E. (1994). Internet-sourced competitive intelligence. In M. E. Williams (Ed.), Proceedings of 1994 National Online Meeting (pp. 87–92), New York: Learned Information.

Dolenko, M. (1998). Making user-based assessment work: Evaluating statistics Canada's Web site. In M.E. Williams (Ed.), Proceedings of 1998 National Online Meeting (pp. 105–114), New York: Information Today.

Fowler, F.J., Jr. (1993). Survey research methods (2nd ed.). Newbury Park: SAGE Publications.

Fox, R.J., Crask, M.R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. Public Opinion Quarterly, 52, 467–491.

German, R., & Oppenheim, C. (1996). Should academic libraries lend LIS journals? Journal of Librarianship and Information Science, 28(4), 191–202.

Gill, H.S., & Yates-Mercer, P. (1998). The dissemination of information by local authorities on the World Wide Web. Journal of Information Science, 24(2), 105–112.

Goree, C.T., & Marszalek, J.F., III (1995). Electronic surveys: Ethical issues for researchers. College Student Affairs Journal, 15(1), 75–79.

GVU's WWW user surveys (1994a, 1994b, 1995a, 1995b, 1996a, 1996b, 1997a, 1997b). GVU's WWW User Survey. Available at: http://www.cc.gatech.edu/gvu/user_surveys.

Kawasaki, J.L., & Raven, M.R. (1995). Computer-Administered surveys in extension. Journal of Extension, 33(3). Available at: gopher://joe.uwex.edu/00/joe/1995/june/rb3.

Kieley, J.M. (1996). CGI scripts: Gateways to World-Wide Web power. Behavior Research Methods, Instruments, & Computers, 28(2), 165–169.

Kiesler, S., & Sproull, L.S. (1986). Response effects in the electronic survey. Public Opinion Quarterly, 50(3), 402–413.

Kittleson, M.J. (1995). An assessment of the response rate via the postal service and e-mail. Health Values, 1995, 18(2), 27–29.

Kovacs, D.K., Robinson, K.L., & Dixon, J. (1995). Scholarly e-conferences on the academic networks: How library and information science professionals use them. Journal of the American Society for Information Science, 46(4), 244–253.

Ladner, S.J., & Tillman, H.N. (1993). Using the Internet for reference. Online, 17(1), 45–51.

Marszalek, J.F., & Goree, C.T. (1995). Practicing what we preach? Gay students' perceptions of student affairs. College Student Affairs Journal, 15(1), 80–86.

Mertens, D.M. (1998). Research methods in education and psychology: Integrating diversity with quantitative & qualitative approaches. Thousand Oaks, CA: Sage Publications.

Miller, J.P. (1994). Should you get wired? Library Journal, 119(2), 47–49.

Oppenheim, A.N. (1992). Questionnaire design, interviewing, and attitude measurement. New York: Pinter Publishers.

Parker, L. (1992). Collecting data the e-mail way. Training & Development, 46(7), 52–54.

Piper, A.I. (1998). Conducting social science laboratory experiments on the World Wide Web. Library & Information Science Research, 20(1), 5–21.

Powell, R.R. (1997). Basic research methods for librarians (3rd ed.). Greenwich, CN: Ablex Publishing Corporation.

Powell, J.H. (1993). Choosing engineering CD-ROM titles: Results of an Internet survey. CD-ROM Professional, 6(6), 34–41.

Roselle, A., & Neufeld, S. (1998). The utility of electronic mail follow-ups for library research. Library & Information Science Research, 20(2), 153–161.

Schiller, N. (1994). Internet training and support. Academic libraries and computer centers: Who's doing what? Internet Research, 4(2), 35–47.

Schmidt, W.C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. Behavior Research Methods, Instruments, & Computers, 29(2), 274–279.

Schoch, N.A., & Shooshan, S.E. (1997). Communication on a listserv for health information professionals: Uses and users of MEDLIB-L. Bulletin of the Medical Library Association, 85(1), 23–32.

Shaw, D., & Davis, C.H. (1996). The Modern Language Association: Electronic and paper surveys of computer-based tool use. Journal of the American Society for Information Science, 47(12), 932–940.

Smith, M.A., & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. Behavior Research Methods, Instruments, & Computers, 29(4), 496–505.

Spink, A., Bateman, J., & Jansen, B.J. (1998). Users' searching behavior on the Excite Web search engine. In M.E. Williams (Ed.), Proceedings of 1998 National Online Meeting (pp. 375–386), New York: Information Today.

Sproull, L. (1986). Using electronic mail for data collection in organizational research. Academy of Management Journal, 29(1), 159–169.

Subramanian, A.K., McAfee, A.T., & Getzinger, J.P. (1997). Use of the World Wide Web for multisite data collection. Academic Emergency Medicine, 4(8), 811–817.

Swoboda, W.J., Muhlberger, N., Weitkunat, R., & Schneeweiss, S. (1997). Internet surveys by direct mailing—An innovative way of collecting data. Social Science Computer Review, 15(3), 242–255.

Teo, T.S.H., Lim, V.K.G., & Lai, R.Y.C. (1997). Users and uses of the Internet: The case of Singapore. International Journal of Information Management, 17(5), 325–336.

Walsh, J.P., Kiesler, S., Sproull, L.S., & Hesse, B.W. (1992). Self-selected and randomly selected respondents in a computer network survey. Public Opinion Quarterly, 56, 241–244.

Zhang, Y. (1998). The impact of Internet-based electronic resources on formal scholarly communication in the area of library and information science: A citation analysis. Journal of Information Science, 24(4), 241–254.

Zhang, Y., & Estabrook, L. (1998). Accessibility to Internet-based electronic resources and its implications for electronic scholarship. In R. Larson, K. Petersen, & C.M. Preston (Eds.), Proceedings of the 61st ASIS Annual Meeting (ASIS'98) (pp. 463–473). Medford, NJ: Information Today.