

USING INTERNET-BASED SURVEYS FOR ACADEMIC RESEARCH: OPPORTUNITIES AND PROBLEMS

Narczyz Roztocki, State University of New York at New Paltz

Abstract

This paper presents a preliminary investigation of the use of Internet-Based Surveys, also called E-Surveys, as tools for academic research. A brief review of existing literature is included. An introductory analysis of the advantages and disadvantages of this method of data collection are discussed, and details of an Internet-Based Survey, which investigated and evaluated current involvement in e-commerce, are given. Company managers' attitudes, fears, and associations about engagement in e-commerce were measured in this survey, the main purpose of which was to provide the authors with practical experience in Web-based data collection. Furthermore, the preliminary results and experiences from the test phase of the survey are presented. This paper ends with suggestions for future research in the field of Internet-Based Survey methodology.

Introduction

Surveys and interviews are perhaps the most popular methods of primary data collection for academic research. For years, traditional paper-and-pencil surveys have been used to test research hypotheses in the fields of marketing, economics, psychology, and political science. However, paper-and-pencil surveys have always had the disadvantage of a limited number of potential respondents, and a slower distribution and return rate. Now, however, the rapid growth of the Internet has expanded research survey opportunities.

Despite its vast potential and the fact that Internet-Based Surveys have been in use for several years, there has been relatively little published research on this methodology (Zhang, 2000). Academic researchers intending to explore the potential of Internet-Based Surveys for the collection of primary data are left with limited guidance. Therefore, most researchers attempt to map traditional paper-and-pencil surveys into Internet format, thus disregarding both the opportunities and the problems associated with the Internet environment. Without understanding how to make use of the opportunities and counteract the problems, it is likely that the resulting academic research will be flawed or comprised in some manner. Potential researchers, therefore, need to be aware of the advantages and, how to make the most of them, and how to solve problems caused by the methodology disadvantages.

The most significant advantages of Internet-Based Surveys as compared to the traditional paper-and-pencil approach, are its efficiency and speed. The use of Internet-Based Surveys:

- eliminates mailing costs for questionnaires;
- reduces costs for coding respondents' data;
- reduces human-error;
- effectively reaches respondents in different geographic areas; and
- reaches respondents in a relatively short amount of time (Bertot and McClure, 1996; Schmith and Leight, 1997; Teo, Lim, and Lai, 1997; Schmidt, 1997; Sackmary, 1998; Zhang, 2000).

Hypothetically, these advantages make Internet-Based Surveys less expensive, more accurate, and, in general, more practical than their paper-and-pencil counterparts.

They are not without potential problems, however. These problems are usually directly associated with the general characteristic of computers and the nature of the Internet. Internet-Based Surveys may generate incorrect results when:

- multiple responses from the same participant are present; or when
- empty (blank) responses;
- incomplete responses;
- invalid responses; or
- biased responses are submitted (Strauss, 1996; Schmidt, 1997; Zhang, 2000).

These problems can be experienced in traditional paper-and-pencil surveys, but are intensified by the special conditions of the Internet environment, which may include perceived anonymity, less control over respondents selection and transmission errors.

To an extent using a properly designed database system for the collection, storage, and analysis of data can solve such problems, however, with a database, submissions can be checked for "multiples" from the same participant, using a relatively simple query. This query can, for example, check for similar responses with near-identical submission times, and keep only the first response, while deleting subsequent ones. Databases can also check for internal consistency, effectively filtering responses that are the result of certain transmission errors, or problematic respondents.

Furthermore, it can be assumed that as the number of Internet users increases, the bias of data collected by Internet-Based Surveys will potentially decrease. It

had been argued that Internet-Based Surveys did not reach the “average” population, since only those with a certain level of income owned computers. Now, however, as the price of owning a personal computer and obtaining reliable Internet access has decreased, the potential level of economic bias associated with Internet-Based Surveys has decreased as well.

The intention of this paper, then, is to explore the use of Internet-Based Surveys specifically for academic research. The advantages and disadvantages of Internet-Based Surveys are investigated, with some attention to how these might be optimized, or overcome, respectively. Because this paper describes a research project in progress, most of the findings are preliminary in nature. However, these findings appear to be important for both academics who employ Internet-Based Surveys for the purpose of data collection, and researchers focusing on improving the efficiency of this method. The methodology for designing an Internet-Based Survey for academic research is presented, along with a case study, from which implications and recommendations for future research are drawn.

Research Methodology

This section describes our methodology for preparing, designing, implementing and analyzing an Internet-Based Survey for academic research. The project was completed in five major steps: Literature Research, Internet-Based Survey Design, Database Design, Internet-Based Survey Posting and Data Collection, and the Examination of Collected Data. These steps represent the preliminary stages of an ongoing research project.

Literature Research. In order to gain more of an understanding of Internet-Based Surveys and database design, the existing literature on its purpose, design, advantages, and disadvantages was studied. The information gained was used to design the Internet-Based Survey for this project.

Internet-Based Survey Design. In order to gain practical, first-hand knowledge of Internet-Based Survey methodology, a survey to elicit information about involvement in e-commerce was developed. In addition to questions about their company’s industry, geographical location, and size, managers were asked about their attitudes, fears, and associations regarding e-commerce. Hyper Text Markup Language (HTML) was used to design the survey.

Shortly after the first (inactive) draft of the E-Survey was posted on the Internet, the managers of three companies, with whom we had worked before, were contacted and asked to provide their feedback. Their feedback allowed us to make improvements in

the survey’s design. For example, several of the questions were reformulated to be more easily understood by potential respondents, and the purpose of the survey was more clearly presented in the introduction.

The questionnaire contained two major parts. The first part contained mainly demographic questions, as shown in Exhibit 1. The second part asked questions about the respondents’ E-commerce involvement, as presented in Exhibit 2. At the end of the questionnaire, a text box was provided to elicit “open” feedback from the respondents.

Exhibit 1. Demographic Profile Questions

Question
What type of company do you work for?
How large is your company?
Where are you located?

Exhibit 2. Questions about E-commerce Involvement

Question
Is your company involved in E-commerce?
How do you rate the importance of E-commerce for your company?
How do you rate your level of trustfulness in the information received from the Internet?

Since the objective of our research was to gain practical experience in creating and implementing Internet-Based Surveys, we were not overly concerned with the topic of our survey. The topic of E-commerce was chosen because it appeared to be a current and interesting one, and one which we felt was liable to draw in a large number of respondents.

In addition, we attempted to include only questions which were regarded as benign. We looked for questions which did not create discomfort in the respondents and did not put them at risk of disclosing information which could result in dismissal.

Database Design. After the active version of the Internet-Based Survey was designed and tested, a relational database system was created in order to store responses. This database contained only a single table, which was designed so that each respondent who pressed the “submit” button created a new record, along with a survey number and submission time.

Internet-Based Survey Posting and Data Collection. The survey was uploaded to the server, and checked for functionality. Minor errors were corrected, and its Web address was submitted to five major search engines. Responses to the survey were collected over the course of approximately one year, using the created database. After submission, the responses and their submission times were stored in the table. The submission time

was used to identify new entries and to track multiple submissions from a single participant.

Examination of Collected Data. After one year of active data collection, the responses were examined both for informational content and general feedback about the survey itself. Because the objective of this study was to gain hands-on experience in using Internet-Based Surveys for academic research, we did not design our survey with the goal of a testing hypothesis in mind. Our examination of collected data was therefore more qualitative than quantitative. The results of the data collection are presented in the Preliminary Results section of this paper.

Preliminary Results

The results of the survey were examined in terms of whether or not they confirmed or disproved or upheld our ideas about the advantages and disadvantages of Internet-Based Surveys. Our method of database design, as to how it impacted the results, was also examined.

Responses to the Survey. The survey, whose objective was to measure degrees of trust in the Internet and levels of engagement in e-commerce, elicited 102 responses during its one-year run. 93 of those submissions were regarded as usable for future research, while nine were not. Eight submissions contained mainly empty spaces. Two submissions were classified as multiple submissions from the same respondent. These submissions, aside from being identical in response, had only a two-minute difference in submission time. Therefore, the eight mostly-blank surveys were discounted, and the two identical submissions were counted as one.

Demographic Profile. The 93 responses which were considered usable were examined for their demographic profile. The results are summarized in Exhibit 3.

Exhibit 3. Demographic Profile

Company Type:	
Service	28
Manufacturing	27
Banking	6
Healthcare	6
Other	25
No response	1
Company Size: (Number of Employees)	
1-5	9
6-20	11
21-50	9
51-100	10
101-500	17
501-1000	10
More than 1000	26
No response	1
Geographical Area:	
North America	26
Europe	15
South and Central America	9
Australia and New Zealand	7
Asia	16
Other	19
Not specified	1

The results suggest that we reached a wide variety of respondents in companies of different sizes, industries, and geographical locations. The relatively high number of “Others,” may be due the fact that we only specified 18 countries to choose from.

E-commerce Involvement. The 93 valid responses were examined for the participants’ involvement in E-commerce. Exhibit 4 summarizes the results.

Exhibit 4. E-commerce Involvement

E-commerce Involvement:	
Using	27
Planning to use	41
Not interested	19
No response	6
E-commerce Importance for Respondent's Company:	
Extremely important	17
Very important	27
Somehow important	27
Not important	19
No response	3
Level of Trustfulness in the Internet:	
Extremely trusting of	7
Very trusting of	32
Somewhat trusting of	45
No trust in	3
No response	6

The number of "No response" entries increased in the second section of the survey. The reason for this dramatic increase (3 non responses in section one as opposed to 15 in section two) may be that the questions about e-commerce involvement were included after the questions regarding demographic profile. After filling in the first three questions, some participants might have chosen not to respond further or not to provide an answer to a particular question, but may still have felt inclined to submit the survey. The increased number of non responses from the first to the second section could also be attributed to the fact that the demographic questions were easy to respond to, because they were objective. In the second section, which was more subjective, the respondent had to make their own justification, which was perhaps more time-consuming.

Preliminary Findings. The results of our survey confirmed several observations by other researchers about the cost savings, accuracy, comprehensiveness, and convenience of Internet-Based Surveys. We did not spend any money mailing out questionnaires and did not register any human errors. We were able to reach respondents in different industries and different geographic areas. In addition, the collected data was available in a database file format, which was convenient for further data analysis.

The degree of voluntary feedback also confirmed that the use of a text box in an Internet-Based Survey is very valuable, because it provides all types of feedback. For example, based on these entries in the

text box, we found out that many of our respondents wanted to learn about the final results. This indicated that an explanation about how and when to obtain the final results would be helpful. Several questions in the text box also asked about the purpose of our survey. Therefore, another of the preliminary findings was that an Internet-Based Survey must have a clear, short explanation of its purpose at the outset.

Even though much of our initial expectations about the advantages and overall accuracy of the Internet-Based Surveys were confirmed, we also came up against a confirmation of its disadvantages, some which we had not fully anticipated. The results showed that an Internet-Based Survey is not always able to reach many respondents in a short time. The number of usable responses (93) was rather small, considering the existence of millions of Internet users, and the relatively long time (approximately one year) that our questionnaire was posted on the Web. One reason for the lack of responses may have been our recruitment method: submitting the Web address to major search engines, which was rather passive in nature. This led us to investigate methods of improving the survey and our implementation methodology in order to counteract the disadvantages.

Improvements

Based on the results of the Test Run (five initial steps, discussed in the Research Methodology section, and considered to be a part of a larger project dedicated to the field of Internet-Based Surveys), a number of improvements in survey design as well as database design are planned. These improvements are intended to address the complications we encountered.

Perhaps the biggest problem with this survey was the lack of responses, which we directly attributed to our recruitment method. It could be assumed that with more active method of recruitment, and direct benefits for respondents, the number of submission would be increased. To make our Internet-Based Survey known to managers, and to reach more potential respondents, we plan to submit the survey to additional search engines, e-groups, and to ask popular business Internet sites to host a link to our survey. This more aggressive recruitment measure will hopefully increase the number of submissions.

It was decided that an additional question asking the respondents of the survey how they found out about the survey would be useful to be provided. This would help to better identify successful and unsuccessful recruitment methods.

We also realized that improvements in database design are needed. As previously described, only a single table was used to store responses during the Test Run. After the completion of the Test Run, a second database with multiple tables. The main purpose of the

multiple tables would be to optimize data storage space. While using a single table for each response, we provided an extensive space for storing respondents' comments from the text box. This procedure meant that the space in a table was "held" even if no comment was submitted, which resulted in unnecessary storage space requirements, and increased file size. An additional table containing only comments (when submitted) would be more efficient. The database will be created using the common database design process. This process includes four major steps: Requirements Collection and Analysis, Conceptual Design, Logical Design, and Physical Design (Elmasri and Navathe, 1994). To learn more about this well-accepted and efficient database design process, interested readers may be referred to the cited literature.

After the period of more extensive data collection is finished, we plan to perform a more complex data analysis. Based on the final results, we also expect to gain extended knowledge about Internet-Based Survey methodology and plan to make additional recommendations regarding the existing advantages and disadvantages.

Conclusions and Future Research

As we have discovered in the course of our case study and our investigation of the literature, there are a large number of research opportunities in the area of Internet-Based Surveys.

One future research project might be the creation of an "intelligent user interface." Instead of relying solely on HTML form to create an Internet-Based Survey, this "intelligent" interface could be supported by programming languages such as Java or JavaScript. As participants progressed through the survey, the intelligent interface could pose questions specifically tailored to their previous responses. If, for example, a respondent indicates that he/she works for a company which is engaged in global business, an additional question about the companies' most important markets outside the country could be asked. Such a question would not be included for these engaged in domestic business only. This would also cut down on the number of "not applicable" blank responses.

Another research opportunity might be to develop a methodology to measure and reduce self-selection bias, which may exist when part of a large population chooses to participate, and another part of this population declines participation. In most Internet-Based Surveys, the participants select themselves, rather than being as a selected by researchers. The use of password allows researchers to have some control over selection and composition of respondent, and thus, to reduce self-selection bias. The amount by which passwords reduce self-selection could be investigated,

and other methods of retaining control over selection could be explored.

Another project might be to compare traditional paper-and-pencil survey methodology with Internet-Based Survey methodology for factors such as validity of response, participants' profiles, or total costs needed for administration. The findings of this research project would help to quantify the cost savings and identify when cost savings are greatest.

Another project could focus more on the design methodology of Internet-Based Surveys. The impact of survey length, layout, appearance, and phrasing of questions and responses could be examined. This research would enable the creation of surveys which are more customized in attracting more respondents from targeted populations.

Overall, it may be concluded that Internet-Based Surveys are important research tools for the future. Modern Internet-Based Surveys are not traditional paper-and-pencil surveys simply mapped into HTML format and posted at the Web, however. They are intelligent user interfaces supported by efficient database systems to retrieve, process and analyze data. As Internet-Based Surveys became more intelligent, and therefore better able to deal with problems, while optimizing advantages, it could be assumed that the importance of the Web for academic data collection will increase.

References

- Bertot, J. C., and McClure, C. R. (1996). Electronic Surveys: Methodological Implications For Using the World Wide Web to Collect Survey Data. *Proceedings of the American Society for Information Science Conference*, 173-185.
- Elmasri, R., and Navathe, S. B. (1994). *Fundamentals of Database Systems*. Menlo Park, CA: Addison-Wesley Publishing Company.
- Sackmary, B. (1998). Internet Survey Research: Practices, Problems, and Prospects. *Proceedings of the American Marketing Association Summer Conference*, 41-49.
- Schmidt, W. C. (1997). World-Wide Web Survey Research: Benefits, Potential Problems, and Solutions. *Behavior Research Methods, Instruments, & Computers*, 29(2), 274-279.
- Schmith, M. A., and Leight, 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.
- Strauss, J. (1996). Early Survey Research on the Internet: Review, Illustration and Evaluation. *Proceedings of the American Marketing Association Winter Conference*, 222-228.

- Teo, T. S. H., Lim, V. K. G., and Lai, R. Y. C. (1997). Users and Uses of the Internet: The Case of Singapore. *International Journal of Information Management*, 17(5), 325-336.
- Zhang, Y. (2000). Using the Internet for Survey Research: A Case Study. *Journal of the American Society for Information Science*, 51(1), 57-68.

About the Author

Narcyz Roztocki is an Assistant Professor of Business at the State University of New York (SUNY) at New Paltz. He received his MS degree in Mechanical Engineering from the Technische Universität Hamburg-Harburg (TUHH), Germany and his Ph.D. in Industrial Engineering from the University of Pittsburgh. His research focus is in the area of Strategic Management, Activity Based Costing, Economic Value Added, Decision Support Tools and Information Systems. Dr. Roztocki is a member of ASEM and IIE.