

Guidelines for developing a robust web survey

Pranav Naithani

Business Faculty, Higher Colleges of Technology, Sharjah, U.A.E.

Email: pranavnaithani@gmail.com

Abstract: A web based survey is an effective tool which is used frequently in academic and non academic researches. Increase in internet usage and easy access to web technology facilitate the growing popularity of web surveys but absence of exhaustive literature on web surveys presents a significant challenge. This paper presents basic guidelines for developing a robust web survey. Aspects related to data quality, coverage bias, questionnaire design, non response bias, response bias, processing error, data duplication and pilot testing are discussed in detail. The paper closes with suggestions on key areas in which advance research is required to enhance the data quality of web questionnaires.

Keywords: Web survey, web questionnaire, questionnaire design, data quality.

1. Introduction

A web based survey is an effective tool which is used frequently in academic and non academic researches. Increase in internet usage and easy access to web technology facilitate the growing popularity of web surveys but absence of exhaustive literature on web surveys presents a significant challenge. This paper presents basic guidelines for developing a robust web survey questionnaire. Web surveys bring along benefits of a self-administered questionnaire as such the preliminary section of this paper presents the salient features of a self-administered questionnaire. Following section dwells into the benefits provided by a web based survey as well as key issues which influence the data quality of a web survey.

2. Benefits of a self-administered questionnaire

Telephonic interviews enhance coverage error (Reis and Judd 2000) and are not suitable for a large number of questions (Manly 1992). Face-to-face interviews are expensive (Langbein 2006) and depend upon trained interviewers (Manly 1992). Though interviews have higher response, yet if self-administered questionnaires are distributed properly they can generate high response rates (Kusek and Rist 2004). Self-administered questionnaire costs relatively lesser, does not depend on trained data collectors and can have a higher level of structure in comparison to interview method (Axinn and Pearce 2006). Axinn and Pearce (2006) suggest data collection through the highly structured survey as a preferred solution for a study where the target audience is spread geographically widely.

Structured surveys have many benefits as discussed above, but they also lack flexibility to gather freely expressed views of respondents and may suffer from low

response rates (Jonassen et al. 1999). Overall, the benefits of self-administered structured questionnaires outweigh the drawbacks and that is the reason why self-administered questionnaires are gaining popularity. One of the most popular mediums through which self-administered questionnaires can be communicated to the respondents is through internet. Following section presents the key reasons for the increasing popularity of web surveys.

2.1. Web-based self-administered questionnaire

Technology provides an economic medium for conducting surveys through internet instead of through the postal mail (Andrews et al. 2003) and data collected through web surveys is comparable to that of other modes of survey (Sternberg et al. 2003). Systematically designed web questionnaires result into a smaller number of missing data (Waltz et al. 2004) and eliminates the need of direct contact with the sample (Axinn and Pearce 2006) which translates into reduced costs of conducting a survey.

Additional reasons for selecting web survey are, absence of volunteer bias and high ethicality due to voluntary participation, absence of interviewer's bias, low-cost (Sternberg et al. 2003), easy administration with lower chances of error due to the automation factor, and reduced need for manual edit checks as many of the checks can be inbuilt on the web based questionnaire (Waltz et al. 2004). Some other reasons for selecting a web questionnaire are speed, immediate indication of non-response rate and higher internal consistency (Goldstein 2007; Wessner 2007).

3. Data quality of self-administered web questionnaire

While web based surveys are cost effective, fast and require a lesser number of edit checks yet there still remain several other issues which need to be addressed before a researcher initiates a web survey. Such key issues which influence the data quality of a web survey are discussed below.

3.1. Web questionnaire: Coverage bias

It has been observed that online population may suffer from age bias as people above the age of 60 years are highly unrepresented and people below the age of 20 years are highly represented in the online population (Knaption and Myers 2004). This fact of online population may sometimes influence the coverage bias. Web questionnaire may also induce coverage bias if the target audience does not have easy access to the internet or if the target audience is short of necessary skills to use the internet. Thereby web surveys may not be equally effective for surveys related to older respondents, respondents in rural areas who have poor accessibility to the internet and respondents with lower levels of literacy.

3.2. Web questionnaire: Mode of contact

Basic e-mail approach with questionnaire incorporated in the body of the e-mail, or questionnaire attached to the e-mail or a website link of the questionnaire incorporated in the body of the e-mail, are the convenient ways available to approach the sample for web questionnaire (Wimmer and Dominick 2005). But many a times approaching through email may not deliver the desired results as the email may be identified as a junk mail and the targeted respondent may not get to read the content. It is suggested to clearly mention the survey topic in the subject of the email and to send email from at least two sender addresses. As suggested by Dillman (2006), a personalised cover letter must be preferably written to enhance the response rates. The content of the introductory letter, noticeably should communicate the salience of the questionnaire (Dillman 2006).

3.3. Questionnaire Design: Questionnaire Segments

Walonick (1997) suggested that, *“items on a questionnaire should be grouped into logical coherent sections. Grouping questions that are similar will make the questionnaire easier to complete, and the respondent will feel more comfortable. Questions that use the same response formats, or those that covers a specific topic, should appear together”*. It is pertinent to note that just as it is important in a paper questionnaire to group the related questions together, in web questionnaires also grouping and sequencing has to be logical and the design and layout to be respondent friendly.

3.4. Questionnaire Design: Length of the questionnaire

Lengthy questionnaire demands more time from respondents and may shrink response rates, but the same questionnaire may also be perceived of higher relevance and importance to the respondents due to its length (Rabin 2007 et al.). As stated by Rabin (2007), earlier research has produced mixed results on the relevance of the length of the questionnaire. It is commonly accepted practice to enhance the response rates of the questionnaire by keeping it preferably short.

As suggested by Rabin (2007), long, complex, multipart questions and open-ended questions should be avoided to achieve higher response rates for lengthy questionnaire. In web surveys the questionnaire may be divided into sub groups and each group may be presented in one web page with a facility to save the responses. A facility of the saving the responses will facilitate the respondent to respond to a lengthy questionnaire in more than one sitting, thereby enhancing the overall response rate.

3.5. Questionnaire Design: Other important elements

To maximise the response rate and to minimise response errors, following relevant guidelines, should be taken into consideration. Close-ended questions are favoured for speedy processing (Bradburn et al. 2004) and the length of the questions should be kept around twelve words (Boynton and Greenhalgh 2004). The questionnaire should steer clear of double-barrelled and leading questions (Burgess 2003) and questionnaire must assure anonymity for higher response rate (Wansink et al. 2003).

Questions involving negatives should not be considered (Burgess 2003) and demographic questions should preferably be placed at the end (Frascara et al. 1997). The questionnaire must have an uncomplicated design (Puleo, 2002). Graphics heavy web questionnaires take longer time to download and may disengage the respondent. The design and colour scheme of the web questionnaire must be simple to minimise dropouts and to realize higher response rates (Rabin et al. 2007).

3.6. Prevention of non-response bias

Krosnick (1999) presented an argument validated by his research output that higher non-response rate in surveys does not necessarily enhance non-representativeness. Brick and Bose (2001) also presented similar arguments by stating that the non-response rate may not all the times result into bias. Still plenty of research publications present the acceptable response rate for surveys in the range of 50 per cent and above (Hill and Alexander 2006) and this acceptable response is expected to minimise the non-response bias. Necessary steps must be taken while designing the web questionnaire to minimise non-response bias. A common source of non-response bias is addition of questions to the questionnaire after the instrument has already been circulated and answered by a part of the sample (Edwards et al 2007). To prevent such

bias, no question should be added after the questionnaire is circulated.

3.7. Prevention of response bias

Refusals and inability to answer questions, the unwillingness of the respondent to show their ignorance, memory biases, inaccurate information and increased level of respondent burden, protection of privacy, integrity and interest are common reasons for response bias (Abbasi 2000). Absence of specific question, use of technical terms and difficult jargons, interpretation of the questionnaire, the wording of the answer/ incomplete alternatives in choices, leading and loaded questions are the other key reasons for response bias (Manly 1992) and the same should be avoided in web surveys.

3.8. Prevention of processing errors and editing checks:

Common processing errors such as data grooming, data capture, editing and estimation (Abbasi 2000) do not apply to the web survey as the survey executed through a web questionnaire carries an inbuilt database. The complete database file from the web server are afterwards imported to the statistical software without manual entering/re-entering of database thus eliminating possibility of processing errors such as data omission and data duplication. Abbasi (2000) highlights following five main editing checks - structure checks, range edits, sequencing checks, checks for duplication and omissions and logic edits. Web questionnaire can easily incorporate inbuilt features for all these checks.

For structure checks, the programmer who designs the web questionnaire can test the web questionnaire and its linked database file for correct recording and labelling of the responses. The web questionnaire and its linked database file can also be easily pre-tested for duplication and omissions checks to prevent duplication and omission of response. Range edit is not required for a web questionnaire as it is not a pen and pencil response where the respondent can give answer outside the valid range. For example, if the answers are to be given on a 5-point Likert type scale then respondents cannot enter any other response/response value, as the web questionnaire will not allow it. Sequencing check and logic edit features can be easily inbuilt on the web based questionnaire for preventing errors such as unmarried respondents and respondents without childcare burden attempting the questions asked from married respondents with childcare burden.

3.9. Prevention of data fraud, duplication and omission

To prevent errors arising out of data fraud and duplication, as suggested by Waltz et al. (2004), username and password should be issued and the web questionnaire may have a provision of cancelling the password once the completed response is submitted. To

prevent data omission, for unintentional non-response to questions, as suggested by Waltz et al. (2004), a prompt may be included in each segment of the web questionnaire, to intimate the respondent about the unanswered question in that segment. But care has to be taken while incorporating the prompts, as too many prompts may disengage the respondent thereby enhancing the dropout rate.

4. Pilot testing the web questionnaire

A preliminary pilot test on a web survey may be administered with the help of a pen and paper version of the web survey in which the questionnaire can be reviewed for the language, structure and sequencing of the different segments. This test may be conducted on a limited representative of the target audience. If required, some alterations may be made in the language and sequencing of the web questionnaire according to the feedback received. Web questionnaire needs to be essentially tested for structure checks, duplication and error checks and if any error is found then the programmer who has developed the web survey make incorporate necessary adjustments. Pilot testing of the web survey should also include measuring of the average time taken to respond to the complete questionnaire so that the target audience may be pre informed about the same while being approached through email.

Reliability refers to the consistency of a measurement technique (O'Leary, 2004). For testing the reliability of the responses collected through a web questionnaire, Cronbach's Coefficient Alpha may be selected as it is a preferred tool for paper based surveys (Cortina 1993; Zimmerman 1993) and equally qualifies for web questionnaires. Internal consistency reliabilities vary from a low of zero to a high of one. Response range close to or above 0.70 is acceptable for Cronbach's Alpha test (George and Mallery, 2003) as such if the responses pass the Cronbach's Alpha test the web survey may be used for further data collection.

5. Conclusion

Web based questionnaires have already become a key medium for conducting survey through self-administered questionnaires. With the increase in global internet penetration, computer literacy and decrease in cost of information technology hardware and software the usage of web surveys will gain more prominence in academic as well as other surveys. Quality of data gathered in a survey is a key issue of any researcher and the same applies to web surveys. Coverage bias of web surveys with reference to male female participation is already witnessing decline as more and more female now participate in web surveys.

In future, coverage bias related to age and geographical location (rural versus urban) will also witness decline as more elder and rural population will have easy access to

the internet. Still, a large part of the global population is illiterate and even amongst literate population a significant part is not computer literate as such the coverage bias related to literacy will take some time to witness any significant decline. To address this issue partly, web surveys can incorporate voice and video to communicate with such audiences and use pictorial data gathering technique or touch screen technique to gather data from such respondents. Though the computer is still a dominant mode of contact for web surveys, availability of the internet on mobile phones and new products such as iPads will make it necessary for the web survey designers to adapt the surveys according to the technical needs of a mobile phone user.

Rapid changes in internet technology are providing additional facilities for effective and efficient implementation of web surveys through self-administered questionnaires and their usage will continue to grow exponentially. As the literature on web surveys is still in a nascent stage, trial and error will be a major source of understanding of new challenges and dimensions of web questionnaires.

References

- Abbasi, Z (2000) Reducing measurement error in informal sector surveys, Informal Paper No - 17, Ministry of Statistics and Programme Implementation, Govt. of India.
- Andrews, D., Nonnecke, B., Preece, J. (2003). Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*. Vol. 16, No. 2, pp 185-210.
- Axinn, W & Pearce, L.D. (2006). *Mixed Method Data Collection Strategies*. New York : Cambridge University Press.
- Boynton, P.M., & Greenhalgh T (2004). Hands-on guide to questionnaire research: Selecting, designing, and developing your questionnaire, *BMJ*; Vol. 328, pp. 312-315
- Bradburn, N.M., Sudman, S., Wansink, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design (Revised Edition)*, John Wiley & Sons, USA.
- Brick, J.M. and Bose, J. (2001) Analysis of potential non response bias. Paper presented in the Annual Meeting of the American Statistical Association, 20th August.
- Burgess, T. F. (2003). A general introduction to the design of questionnaires for survey research, *Guide to the design of questionnaire*, ISS, University of Leeds, U.K.
- Cortina, J. M. (1993). What is Coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology*, Volume 78, pp. 98-104.
- Dillman, D.A. (2006) *Mail and Internet Surveys: The Tailored Design Method 2007 Update with New Internet, Visual, and Mixed-Mode Guide*, Wiley-IEEE.
- Edwards, J.E., Scott, J.C., Raju, N.S. (2007) *Evaluating Human Resources Programs: A 6-phase Approach for Optimizing Performance*, John Wiley and Sons.
- Frascara, J., Meurer, B., Toorn, J., Winkler D. and Strickler, Z. (1997) *User-centred Graphic Design: Mass Communications and Social Change*, Taylor & Francis.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 4th edition. Boston: Allyn & Bacon.
- Goldstein, B. (2007) *The Ultimate Small Business Marketing Toolkit: All the Tips, Forms, and Strategies You'll Ever Need*, McGraw-Hill Professional
- Hill, N. and Alexander, J. (2006) *Handbook of customer satisfaction and loyalty measurement*, Gower Publishing
- Jonassen, D.H., Tessmer, M., Hannum, W.H. (1999) *Task analysis methods for instructional design*, Lawrence Erlbaum Associates
- Knapton, K. and Myers S. (2004). Demographics and online survey response rates. *Quirk's Marketing Research Review*.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, Vol. 50, pp. 537-567.
- Kusek, J.Z. and Rist, R.C. (2004) *Ten Steps to a Results-based Monitoring and Evaluation System: A Handbook for Development Practitioners*, World Bank Publications
- Langbein, L.I. (2006) *Public Program Evaluation: A Statistical Guide*, M.E. Sharpe.
- O'Leary, Z. (2004). *The Essential Guide to Doing Research*, Sage: London
- Puleo E., Zapka J., White M.J., Mouchawar, J., Somkin, C., Taplin, S. (2002). Caffeine, cajoling, and other strategies to maximise clinician survey response rates. *Evaluation and the Health Professions*, Vol. 25, No. 2, pp. 69-184.
- Rabin, J., Hildreth, W.B., Miller, G. (2007) *Handbook of public administration*, CRC Press
- Sternberg, R.J., Dietz-Uhler, B., Leach, C. (2003) *The Psychologists companion: A guide to scientific writing for students and researchers*, 4th edition, Cambridge University Press.
- Walonick, D.S. (1997). *Survival Statistics*, StatPac Inc.
- Waltz, C.F., Strickland, O. and Lenz, E.R. (2004) *Measurement in Nursing and Health Research: Third Edition*, Springer Publishing Company
- Wansink, B., Cheney, M. M., Chan, N. (2003). Understanding Comfort Food Preferences Across Gender and Age. *Physiology and Behavior*, Vol. 53, pp. 459-478.
- Wessner, C.W. (2007) *An Assessment of the Small Business Innovation Research Program*, National Academies Press.
- Wimmer, R.D. and Dominick, J.R. (2005) *Mass media research: An introduction*, Cengage Learning
- Zimmerman, D. W., & Zumbo, B. D. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions *Educational & Psychological Measurement*, Vol. 53, pp. 33-50