DHĀRANA
BHAVAN'S INTERNATIONAL JOURNAL OF BUSINESS
Vol:4, 2 (2010) 77-82 ISSN 0974-0082

Web Based Surveys: An Analysis of Nonresponse Causes

Srivenkataramana T[1], Gopabandhu Mishra[2] and Saisree M[3]

Abstract

Web based surveys (WBS) are becoming increasingly common due to widening net connectivity. WBS mode has its positive and negative points. Cost and time reductions are on the plus side while the question of randomness of sample, nonresponse and quality of response are on the other. Thus WBS presents a mixed bag. This paper examines the reasons for nonresponse in WBS. The often quoted reasons are examined using Factor Analysis. It is found that design is a predominant factor. This aspect is further examined by analyzing the effect of positioning of demographic block, using ANOVA. It is noted that this block in the beginning of the questionnaire results in more dropouts as compared to it being later on. Also examined is the impact of sensitivity of questions and its interaction with the positioning of demographic block.

Key words : ANOVA; Demographic Block; Factor Analysis; Noncoverage; Nonresponse, Questionnaire Design; Web Based Surveys.

1. Introduction

In the 21st century, the internet is having a profound effect on the survey industry, as it is the case with several other areas of human enterprise. The rapid development of surveys on the World Wide Web has led some to argue that soon WBS will replace the traditional methods (Couper, 2000). Thus we stand at the threshold of a new era for survey research. The impact of web on survey data collection is worthy of serious research attention.

There are two main problems in WBS: *coverage error and nonresponse error*. For the former, there are two approaches. (1) Limit the study to those with access to web. (2)Overcome the limitation of restricted technology access by making it available to all those concerned. The latter problem, *nonresponse*, poses a bigger hurdle. Even if we could successfully identify a sampling frame which is of interest to clients or analysts, the problem of nonresponse may still threaten the utility of WBS.

Nonresponse has always been an issue of concern irrespective of the mode of data collection as missing values not only mean less efficient estimates because of the reduction in sample size but also mean that standard complete data analysis methods can not be immediately used. With increased scope of electronic media, WBS is

1. Bangalore University,Bangalore 2. Utkal University, Bhubaneswar 3. IBSB, Bangalore

becoming a common mode. With this, survey respondents and analysts face new and different challenges.

Nonresponse error arises through the fact that not all people included in the sample are willing or able to complete the survey. It is a function of both the rate of nonresponse and of the differences between respondents and nonrespondents on variables of interest (Groves and Couper, 1998).

For surveys where the frame cannot be identified, nonresponse is hard to define. For example, if an open invitation is issued on a web portal to participate in a survey, the number of all those eligible to participate is typically not known, and therefore the nonresponse rate is unknowable. Given this, there is currently little information on nonresponse in WBS. We must rely primarily on e-mail surveys to give us a handle. Several studies have compared response rates from email surveys with those from mail surveys of the same populations. These studies are summarized in Couper et al., (1999) and Schaefer and Dillman (1998). In all but one study, the email survey failed to reach the response levels of mail surveys. Several explanations account for this difference. One is that tried and tested motivating tools used in mail surveys can't be implemented in email surveys. The situation and experience with respect to nonresponse is similar in a WBS and functional equivalents are yet to be developed. There is at present little experimental literature on what works and what does not in terms of increasing response rates to WBS. Many of the techniques developed and tested over time to increase response rates in mail surveys may not work in WBS. Electronic equivalents of response stimulating factors are yet to be developed. As coverage problems are overcome in WBS, the problem of nonresponse is likely to become increasingly prominent (Couper, 2000).

2. Research on nonresponse in WBS so far

We first outline briefly some issues that have been already examined. Dillman and Knapp (1998), Heidingsfelder(1999) have mentioned the design specific causes of nonresponses. Dillman(2000) argues that graphically complex and fancy designs cause more nonresponse in comparison to plain designs. Frick et al. (1999) have verified the effect of the order of the topics on dropouts and found that sensitive questions in the beginning of the questionnaire lead to more dropouts. Bikart and Schmittlen (1999) illustrate that some respondents display a survey response propensity (an enduring personal characteristic). Nonrespondents may lack this propensity and may be suffering from survey response fatigue. Effect of incentives on response rate was studied by Frick and Reip (1999) followed by Bosnjak and Bandilia (2000). The results were partly contradicting: the former says that incentives have an effect while the latter negates this. Bosnjak and Tracy (2001) have classified response behavior and identified seven response categories as follows: a) *Complete responders*, who view all the questions and answer all, b) *Item responders*, view all but answer only some, c) *Item nonresponding* dropouts, view some and answer only a few of these, d) *Answering dropouts*, view some and answer all those, e) *Lurkers*, view all but answer none, f) *Lurking dropouts*, view only some without answering any, and g) *Unit nonresponders*, neither view nor answer any of the questions.

Two review studies report interesting findings. Knapp and Heidingsfelder (2001) reviewed nine unrestricted self-selected surveys by internet Rogator (Germany) in order to identify factors influencing dropout rates. They found that *longer surveys, sensitive topics* and *lack of incentives* led to higher dropout rates. MacElroy (2000) reviewed 19 such studies by Modalis Research Technology (USA) involving b2b technology decisions. He noted that dropout rates decrease with incentives and increase with questionnaire length.

The above studies are rather limited in scope. They mostly refer to some of the web (or email) survey design characteristics and not to all of them, Vehovar et al (2002). Also, they refer to only one outcome, response rate, defined as a percentage of respondents in the target population. However the web survey response process has three distinct stages *(email recruitment, access to the questionnaire, and questionnaire completion).* and several outcomes can be defined (Lozar Manfreda, 2001; Vehovar, 2002). Monica Pratesi *et al*. (2004) have identified several steps in the web survey process and modeled the survival of eligible respondents to find out which respondents come farthest in the process. The analysis helps to explore the nonresponse process better Vera Toepoel *et al*. (2009) have studied the effect of

layout in rating scales contributing to response rate in a web questionnaire.

While such studies are useful, they leave many questions unanswered. Certainly web survey methodology is still in its infancy. And, the additional information available when using the web to collect data can provide valuable insight into nonresponse behavior as nonresponse is an attitudinal problem to some extent.

3. Reasons for nonresponse

The often quoted reasons for nonresponse are listed below mode wise.

Type of Survey	Main reasons for nonresponse
Personal Interview	1. Address not found. 2. Door locked 3.Interviewer's inability to handle the respondent. 4. Sensitivity due to interviewers' presence. 5. "Busy now, come later" response. 6. Nonavailability of concerned person.
Telephone Survey	1. Number is incorrect or changed. 2. The number is busy or no answer. 3. No direct contact (Answering machine). 4. Concerned person is not available.5. "Busy now and call latter" response. 6. Person said "I don't give information over phone. 7. Contact established but language is a barrier. 8. Respondent hung up the phone.
Mail Survey	1. Address is incorrect or incomplete 2.Responder never returned the mail.
WBS	1. Surveys are often time consuming. 2. Down loading takes a lot of time 3. Topic of the survey is not of interest. 4. No incentives for participation. 5. No clarity in the questions. 6. Navigation between the questions is difficult. 7. Questions are sensitive. 8. Doubtful integrity of the source. 9. No benefit perceived in taking the survey 10. The link could be a virus.

Table 3.1 Reasons for nonresponse

These reasons may be compared for commonness as well as contrast to help in and in the choice of corrective steps. It can be seen that WBS contrast with the other three modes essentially with respect to some design specific issues, such as down loading time and navigation between questions. These are the counterparts of interviewer bias in traditional surveys. Fortunately, the researcher has some leverage over these variables which can be exploited for better survey implementation. Also the traditional methods of handling nonresponse may not succeed in dealing with nonresponse in WBS.

4. Design and methodology

Three Web based studies were carried out in a sequence. The first *(study 1*) aims to identify the key factors for nonresponse, while the other two studies *(studies 2 & 3)* examine these key factors causing nonnresponse.

Study 1 uses factor analysis to group the variables causing nonresponse. Three factors emerge: *Design factor, Benefit factor and Source factor*, with the first contributing almost 37% of the variance.

As a sequel, ***study 2*** examines the design factor further. *Demographic block* is the vital part in a majority of surveys and WBS is no exception. ANOVA was used to find out whether the positioning of demographic block has an effect on response rate. Three versions of questionnaires, *V1, V2, and V3* were prepared. *V1* had demographic block in the start, *V2* in the middle and *V3* had in the end. Each version had the same 20 questions.

Finally ***study 3*** was conducted as a 2-way ANOVA to assess the individual and interaction effect of block positioning and nature of questions, using a 2x3 design. The general questions included topics like holidaying and choice of holiday destination, while the sensitive questions related to income and health issues.

Study	Month & Year of Web request	Achieved sample size	Nature of questions
Study 1	May, 2009	1000	Causes for nonresponse in WBS
Study 2	Jan, 2010	2464	General
Study 3	July, 2010	2554	General and sensitive

Table 4.1 Profiles of Samples

5. Results

The results of the studies are outlined next.

5.1 Study 1

Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.780, indicating that the reduced set of variables collectively meet the necessary threshold of sampling adequacy. Bartlett's test of Sphericity, Approx. Chi-Square is 158.137 at, indicating significant correlation among variables.

Variables	Factor 1	Factor 2	Factor 3
Surveys are often time consuming	.732	-.226	-.375
Down loading the link is time consuming	.652	-.048	-.171
No easy navigation between the questions	.733	.311	-.340
Link could be a virus	.689	-.135	.317
No clarity of the questionnaire	.607	.463	-.292
Topic of the survey is not of interest	.502	-.555	-.024
No incentive for participation	.474	-.114	-.234
Integrity of the source	.593	.140	.596
No benefit perceived in taking the survey	.571	-.415	.427
Questions are sensitive in nature	.362	.737	.332

Table 5.1 Component Matrix

All the variables (except for variable 6) show a high factor loading (0.5 or above) with only one factor. In order to allow a simple interpretation, orthogonal rotation with Varimax procedure was employed leading to the following results:

Variable	Factor 1	Factor 2	Factor 3
1. Surveys are often time consuming	*.788	.294	-.144
2. Downloading the link is time consuming	*.602	.297	.070
3. No easy navigation between the questions	*.792	.067	.344
4.Link could be a virus	*.506	.181	-.061
5.No clarity of the questionnaire	*.672	-.047	.462
6.Topic of the survey is not of interest	.369	*.541	-.362
7.No incentive for participation	.309	*.679	.190
8.Integrity of the source	.060	*.677	.513
9.No benefit perceived in taking the survey	.134	*.813	-.047
10.Questions are sensitive in nature	.094	.088	*.876

Table 5.2 Rotated Component Matrix

**High loading*

Variables	Eigen values	% of Variance	Cumulative%	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
				Total	Total % of Variance	Cumulative %		Total % of variance	Cumulative %age
1	3.627	36.267	36.267	3.627	36.267	36.267	2.582	25.822	25.822
2	1.438	14.378	50.645	1.438	14.378	50.645	2.095	20.95	46.772
3	1.174	11.739	62.384	1.174	11.739	62.384	1.561	15.612	62.384
4	0.986	9.862	72.246						
5	0.76	7.601	79.846						
6	0.557	5.571	85.418						
7	0.524	5.236	90.654						
8	0.368	3.679	94.333						
9	0.36	3.597	97.931						
10	0.207z	2.069	100						

Table 5.3 Total Variance Explained

Scree plot:The variables under consideration are plotted against their eigen values to confirm the number of factors extracted. The components lying on the slope are considered to be prime.

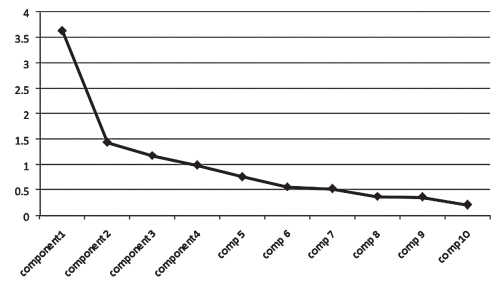

Fig 5.1 Scree Plot

The following points emerge from the analysis:

Factor 1- **Design Factor**: The variables such as *surveys are often time consuming, down loading link time, navigation between the questions & clarity in the questionnaire* have shown high factor loadings, with an **eigen value** of **3.62.**

Factor 2-**Benefit/Interest Factor**: The variables such as, *topic of the survey is not of interest, no benefit perceived in taking the survey, no incentive for participation* have high factor loadings with an **eigen value** of **1.43.**

Factor 3-**Sensitivity factor**: The variable *questions are sensitive in nature* have shown a high loading with an **eigen value** of **1.17.**

Since design turns out to be an influential factor, unlike in traditional surveys, it provides scope for better survey monitoring. *Study 2*, indicated significance of the positioning of demographic block as seen in the next two tables. The response rate is noted to improve as block position is deferred, and the standard error reduces.

Demo-graphic block position	Sample size	Mean	Std. Deviation	Std. Error	95% Confidence bounds for mean	
					L.Bound	U.Bound
Start	820	3.6585	1.89865	.06630	3.5284	3.7887
Middle	824	6.2233	1.66696	05807	6.1093	6.3373
End	820	8.3220	1.26713	.04425	8.2351	8.4088
Total	2464	6.0682	2.50866	.05054	5.9691	6.1673

Table 5.4 Descriptive Statistics of Response rate

Sources	Sum of Squares	df	Mean Square	F	Significance
Between	8946.238	2	4473.119	1679.559	.000
Within	6554.308	2461	2.663		
Total	15500.545	2463			

Table 5.5 ANOVA for Response Rate

Study 3 Two-Way ANOVA for Response Rate as dependent Variable

Question type (general, sensitive) and Demographic block positioning (end, middle, start) being taken as independent and Response rate as dependent.

Independent Variables	Label	Sample Size
Nature of Questions	General questions	1285
	Sensitive questions	1269
Dblock positioning	Start	850
	Middle	854
	End	850

Table 5.6 Between subject Factors

Nature of Questions	Dblock positioning	Mean	Std. Deviation	N
General questions	Start	5.40	1.021	415
	Middle	7.60	1.021	435
	End	9.20	.749	435
	Total	7.43	1.811	1285
Sensitive questions	Start	2.40	.633	435
	Middle	4.79	.750	419
	End	7.40	1.021	415
	Total	4.69	2.356	1269
Total	Start	8.32	1.899	850
	Middle	6.22	1.667	854
	End	3.66	1.267	850
	Total	6.07	2.508	2554

Table 5.7 Descriptive Statistics of Response rat

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	14090.010[a]	5	2818.002	3648.259	.000	.877
Intercept	93912.957	1	93912.957	121582.181	.000	.979
questiontype	4546.883	1	4546.883	5886.514	.000	.698
DBP	9009.890	2	4504.945	5832.220	.000	.821
questiontype * DBP	277.974	2	138.987	179.936	.000	.124
Error	1968.136	2548	.772			
Total	110102.000	2554				
Corrected Total	16058.146	2553				

a. R Squared = .877 (Adjusted R Squared = .877)

Table 5.8 Tests of Between Subject Effects

From the ANOVA table we can infer that Demographic block positioning continues to matter irrespective of nature of questions in the sense that its deferring improves the response rate. (Partial Eta Squared values of Demographic Block positioning i.e. DBP and Question Type is 0.821 and 0.698 respectively. So, clearly DBP contributes more to response rate than Question Type). The significance of the

interaction effect is demonstrated by the F Value 179.936, which is supported by the fig 5.2 as well.

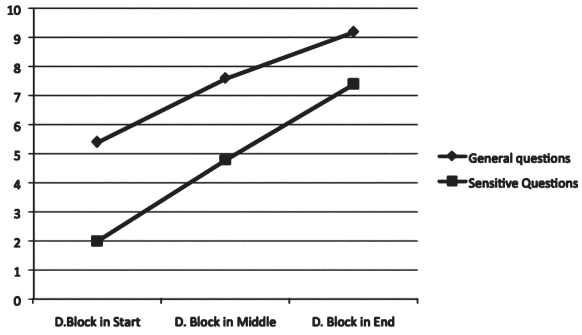

Fig 5.2 Means Plot of Response rate

6. Discussion

The methodology used in this study is innovative in a couple of ways. Bikart and Schmittlen, (1999), illustrate that some respondents display a *survey response prospensity* while nonrespondents may suffer from *survey response fatigue*. Factor analysis is used to tackle these kinds of attitudinal issues. As a sequel to this, ANOVA is carried out to identify whether there is any significant improvement in response rate when the design is altered.

Factor analysis pointed to three dominant factors for nonresponse: *Design, Benefit and Sensitivity*, the first one accounted for about 37% of variance. Fortunately, designing a survey questionnaire is in researcher's hands. Thus nonresponse in a WBS can be controlled to a large extent through a well designed questionnaire. To come up with better design solutions, different versions of questionnaires were prepared by positioning the demographic block at three different points: *start, middle* and the *end*. ANOVA suggests that response rate significantly increases as the block is deferred. A logical question to ask is what if the demographic block is deleted? Will it not improve the response rate further? However, in many surveys this may not be a feasible proposition as demographic variables may be key for the analysis of data received. To understand the effect of demographic block further with respect to the nature of questions, two questionnaires were prepared, one with general questions and the other with sensitive questions. In the presence of sensitive questions too the same type of effect of demographic block positioning is noted.

A classification of respondents into 7 categories was mentioned in the introductory section. These range from total response to total nonresponse. Taking a clue from this, one may devise steps to make nonrespondents move from a "less desirable category" to a "more desirable one": eg from 'lurkers' to 'item nonresponders'. This aspect will depend on the characteristics of respondents and also the corrective steps. Empirical research in this regard is called for in order to arrive at practical guidelines in this regard. Another open question is that of randomness of web samples. This is yet to be addressed satisfactorily.

7. References

Bikart, B., and Schmittlein, D. (1999). The distribution of survey contact and participation in the United states: constructing a survey-based estimate. Journal of Marketing Research, 36(2),286-294.

Bosnjak, M. and Tuten. T.L. (2001). Classifying Response Behaviors in Web-Based Surveys. Journal of Computer-Mediated Communication 6(3)

Couper,M.P.(2000).Web Surveys. A Review of issues and approaches. Public opinion quarterly, 64,464-494.

Couper,M.P.,Blair,J.&Triplett T(1999) A comparision of mail and email for a survey of employees in federal statistical agencies. Journal of Official Statistics, 15, 39-56.

DeLeeuw, E.D. (2005).To mix or not to mix Data collection modes in Surveys. Journal of Official statistics, 21,233-255

Dilman, D.A., Tortora, R.D., Conradt, j., and Bowker, D.(1998). Influence of plain versus Fancy Design on Response Rates for Web Surveys. Paper presented at the annual meeting of the American Statistical Association, Dallas, TX.

Grooves and Couper, M.P (1998). Nonresponse in interview surveys, New York, Wiley series in Survey Methodology

Knapp,F. and Heidingsfelder, M.(2001): " Drop out analysis: Effects of the survey design". Pabst science publishers,pp.221-230.

MacElroy, B. (2000): "Variables influencing dropout rates in Web based surveys". Quirks marketing research review,July/August2000. Paper. http:/www.quirks.com/(November21, 2001)

Pratesi,M.,Lozar.M.(2004). List-based Web Surveys: Quality, Timeliness, and Nonresponse in the Steps of the Participation Flow, Journal of Official Statistics 20,451-465

Schaefer and Dillman 1998, Development of standard email methodology, public opinion quarterly, 62: 378-97

Vera,T.,Marcel,D.& Arthur van Soest(2009) Design of Web Questionnaires: The effect of Lay out in Rating scales, Journal of Official Statistics 25,509-528