DESIGN ISSUES IN WEB SURVEYS

Vasja Vehovar, Katja Lozar Manfreda, University of Ljubljana, Zenel Batagelj, CATI Center
Katja Lozar Manfreda, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia, katja.lozar@uni-lj.si

Key Words: Web surveys, Internet surveys, Web questionnaire design, scroll. vs. screen design, logotypes, survey topic, instructions

INTRODUCTION

In the last five years since research on the use of Web for surveying has been going on, a lot of issues regarding the design of these surveys have been discussed and several research papers are collected on a site dedicated to the Web Survey Methodology at http://www.websm.org (site prepared by authors at the Faculty of Social Sciences, University of Ljubljana). However, there are still unsolved problems which may affect the participants' initial decision to participate in a Web survey or their willingness to proceed with completing the questionnaire. Namely, response rates (respondents among all invited to the survey) and rates of complete respondents (respondents who complete the questionnaire among all who start answering it) are often still low for Web surveys. In this paper the impact of several Web questionnaire design features will be discussed and their impact on unit, partial and item (non)response will be measured.

WEB SURVEY PARTICIPATION

The process of respondent's participation in a Web survey has several stages (inclusion in the sample, solicitation, survey retrieval, questionnaire answering, and transmission of responses) and the non-response process is occurring during all the above mentioned stages. It depends on respondent's characteristics, social and technological environment and survey design features (Vehovar et al. 2001). In this paper we are interested in the later, or more specifically in the features of the Web questionnaire design which can influence respondents' participation during the filling out the questionnaire but have a limited impact on the initial decision to participate in a Web survey. In general, the CASIC modes have a positive effect on item non-response and data quality (Saris 1998). On the other hand, the role of advanced technological features, such as intensive use of graphics, images, animation, cookies, links to other web pages, etc., is much more ambiguous. For example, graphical progress indicator has been shown to increase completion time, but having no impact on survey completion and item nonresponse. (Couper et al. 1999). In this paper the following features of the Web questionnaire design are discussed: one page vs. multiple page design, instructions, topic, and advanced graphics (logotypes).

DATA

The experiments are based on data from two Web surveys, conducted as a part of project RIS (Research on Internet in Slovenia, http://www.ris.org) at the Faculty of Social Sciences, University of Ljubljana. Surveys were widely advertised on frequently visited Slovenian Web sites, on IRC channels, with logon messages of two public ISPs, and in traditional media. In addition, up to three email messages were sent to Internet users whose email was published in the public email directory. For the RIS98 survey, email messages were sent also to respondents who had left their email address in one of the two previous RIS Web surveys (RIS96 and RIS97).

The first survey, the RIS96 survey was conducted in April 1996 and lasted for 30 days. The average length of the interview was 7 minutes. The second survey, the RIS98 survey was conducted from June to August 1998 and lasted for 66 days. The average interviewing time was 22 minutes. The RIS96 survey attracted 2034 and the RIS98 survey 6522 respondents. In comparison to the number of active Internet users in Slovenia, the response was very high: almost 10% of active Internet users participated in the surveys.

The RIS96 survey was relatively short and included questions on Internet use and social-demographics. An experiment regarding the scroll- vs. screen-based design was implemented in this survey.

The RIS98 survey was longer. In addition to the basic block of questions related to the Internet usage and social-demographics, questions about Web site visits and satisfaction with the survey (at the end of the questionnaire) were added for all respondents. In addition, additional modules with different topics were added to randomly selected sub-samples of respondents. Some of these modules were compulsory and some were optional. Also two other experiments were implemented: an experiment regarding the use of logotypes when measuring Web sites' visits and an experiment regarding detailed instruction at the first survey question.

SCROLL VS. SCREEN-BASED DESIGN

In 'simple HTML forms' (Kottler 1997) as in more complex scrolling (Clayton and Werking 1998; Zukerberg et al. 1999) the questions scroll from one to another on a singe Web page. In the alternative 'computer assisted Web interviewing' (CAWI) (Kottler 1997) – also called 'screen-by-screen construction' by Dillman and Bowker (1999) or simply 'pages' by Clayton and Werking (1998) - each question or a set of questions are placed on successive separate Web pages. The combination of both approaches is also possible.

In environment where technological limitations (Java scripts not supported, slow data transmission, etc.) create difficulties, the advantages of either method are technologically driven. In such cases advantages of computer assisted interviewing (automated skip patterns, additional probing for questionable answers or nonresponse, edit checks, etc.) are often not enabled in the scroll-based design, therefore such questionnaires are also called 'flat-file instruments' (Farmer 1998) in contrast to 'interactive instruments' (Farmer 1998) or 'dynamic surveys' (Onyshkevych and McIndoe 1999) for the screen-based design. However, when the technology enables the full data controls in both options, the advantages are related – conditional on a given complexity of the questionnaire - only to data quality, interviewing time, and user's satisfaction.

In the RIS96 an experiment aimed at measuring differences in the effect of the scroll- vs. screen-based design was performed. Only the results on respondents using graphical browsers are reported here. One experimental group (n=644) was answering the questionnaire on one long scrolling page, while the other group (n=672) the questionnaire with each question (or block of questions if there were multiple items with the same scale) on its own page and next page appeared only when the previous was finished.

A surprising and encouraging result is that the percentage of respondents prematurely abandoning the questionnaire was not significantly different in both designs (14.6% for scroll- vs. 16.5% for screen-based design) despite the longer interviewing time in the scroll-based design (466 seconds in comparison to 368 seconds for the scroll version) since several HTML pages had to be downloaded.

In both cases questionnaire was designed so that no skips were needed and no controls for nonresponse were included therefore item nonresponse could be compared across both designs. Scroll-based design resulted in higher item nonresponse; it looks like respondents were more often omitting particular questions if they had all questions on single HTML page. This difference is particularly noticeable for questions in the form of 'grid' (question in table with items presented as the row headings and scales on which these are rated as column headings). Questionnaire included two 'grids' and for both number of skipped items was larger in scroll-based design (40.3% vs. 35.0%, t-test, α=0.013 for 1st 'grid', and 40.2% vs. 35.3%, t-test, α=0.021 for 2nd 'grid'). People were more often ommiting questions in tables when they were placed on one long HTML page together with other questions than when they were placed on a separate HTML page.

Different result regarding the interviewing time was obtained by Zukerberg et al. (1999). They report on an experiment where no significant difference in interviewing time was observed. However, different approach was taken in this case, namely usability testing in a laboratory environment with ideal connection speed. This is very different from our survey where half of respondents mostly accessed the Internet from home with dial-up access. In addition, the above authors did not find any significant difference in the number of missing responses or number of questions skipped incorrectly (skip patterns were not automated) in the two designs.

Another experiment regarding the number of questions/items per screen, however slightly different, was performed by Couper et al. (1999). They compared a grid question on one screen (multiple-item-per-screen version) with separately put items on individual screens

	Scroll-based design	Screen-based design	Difference
N	644	673	
Completion rate (complete/started interviews)	85.40%	83.50%	χ^2; α=0.362
Mean length (sec.) for complete respondents	368	466	t-test; α<0.0005
Item nonresponse (% of not answered questions among all questions) for all respondents	16.30%	12.80%	t-test; α=0.035
Item nonresponse (% of not answered questions among all questions) for complete respondents	5.00%	2.30%	t-test; α=0.001
Willing to provide email address among complete respondents	67.50%	70.40%	χ^2; α=0.300

Table 1: Impact of scroll- vs. screen-based design (RIS96)

(single-item-per-screen version). The multiple-item-per-screen version took significantly less time to complete and resulted in smaller item missing data. In contrast to this result, in our case, the scroll version produced higher item nonresponse. However, the two experiments are not completely comparable since in Couper et al. (1999) case the two designs did not differ only in the number of items per screen, but also in the layout of the questions: table vs. no table.

ADVANCED GRAPHICS: LOGOTYPES

Advanced graphics may improve the respondents' motivation and satisfaction, and can generate a valuable feeling of having a "good time" or "fun" while answering a Web questionnaire. However, the trade-off must be taken into account, because of technological limitations, distraction of respondents, biased answers, etc. An experimental study of two questionnaire designs - fancy and plain - (Dillman 2000) provides an important warning about extensive use of graphics, since respondents to the plain version completed more pages, more write-in boxes, were less likely to drop out, spent less time, and were less likely to have to return to the questionnaire in order to complete it.

In the RIS98 Web survey an experiment with the use of logotypes when measuring the knowledge and frequency of visits of Web sites was used. Three sets of Web sites (12, 12 and 10 sites) and six experimental groups (n= 751, 743, 537, 528, 582 and 517) were included in the experiment. Each set of Web sites appeared on one questionnaire screen for two of the experimental groups: one experimental group was assigned a screen where sites' logotypes were presented together with their names, the other experimental group was assigned a screen where only names of the sites were writen.

Due to the use of logotypes slower data transer was expected resulting in more respondents abandoning the survey if asnwering the questions with logotypes. In addition, since logotypes offer a visual stimulation, differences in responses and smaller item non-response was expected.

The results confirm the hypothesis. The overall interviewing time did not differ significantly between those having and those not having logotypes. Nevertheless, the use of logotypes resulted in significantly more respondents abandoning the survey when logotypes were used. In the three groups with logotypes 4.2% of all respondents abandoned the questionnaire excactly when they were answering these questions, while in the three groups with no logotypes this share was only 0.4%. There was no difference if respondents were answering the questionnaire using a dial-up access or no, however there was difference in the type of browser used. They especially more often abandoned the survey when logoyptes were used if they were using Internet Explorer 3.

While the above result shows negative consequences of the use of logotypes, their effect as a visual stimuli was positive. First, item nonresponse was slightly lower when logotypes were used (11.8 vs. 13.5, t-test, α=0.007). Second, there were differences in results. Statisticaly significant differences (t-test, α<0.05) in results occured at 22 out of 34 Web sites (65%).

For each Web site an absolute difference for six of possible answers was calculated. Then, an average across all sites for six of possible answers was calculated and is presented in Figure 1. On average, percentage of »don't know« answers was smaller and percentage of »visit occasionally« answers was larger when no logotypes were used. When respondents saw

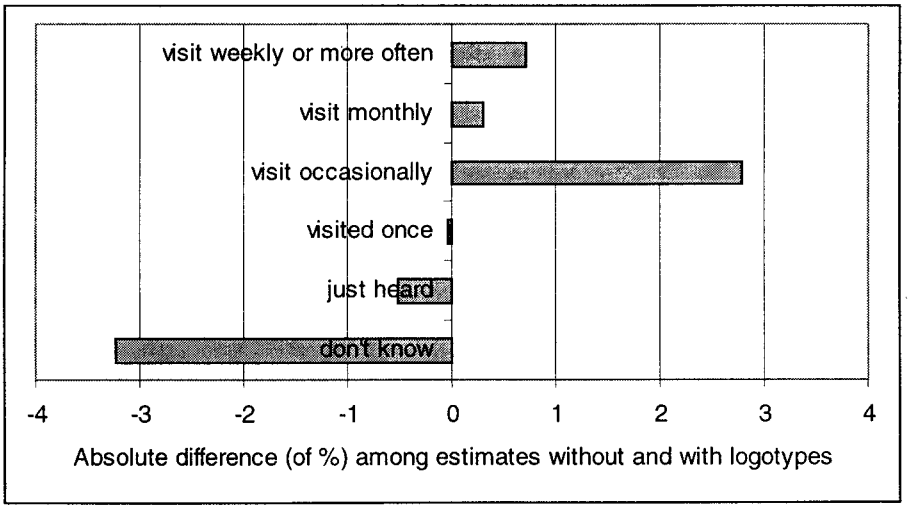

Figure 1: Impact of logotypes (RIS98)

the logotype they were therefore more often inclined to admit that they do not know the site, while on the other hand, when they were not confronted with the logotype they more often answered that they visit it occasionally.

We can conclude that respondents are more committed to answer all the questions and more often admit that they do not know the site when logotypes are used. However, they also more often abandon the survey prematurely if logotypes are used.

SURVEY TOPIC

In general, the respondent's decision to participate in a survey depends on the survey topic, its salience, respondent's knowledge and involvement in the topic (Groves and Couper 1998; Kojetin et al. 1993; Martin 1994).

In the RIS98 Web survey an experiment involving different survey topics was included (Figure 2). After a basic set of questions on the Internet use, demography and Web sites respondents were randomly assigned one of ten additional sets of questions. After answering it 60% of respondents had a chance to select one or more additional sets of questions among the other nine and three additional ones (all together twelve modules). The remaining 40% of respondents received final few questions about the satisfaction with the survey.

Respondents most often chose the modules with questions on computer infrastructure (12.4% of all who had the chance to choose), attitudes toward erotics and pornography (12.2%), activities in spare time (10.2%) and Internet software (9.7%). These were the topics of interest for intensive computer and Internet users and topics which were supposed to be 'funny' and 'amusing'. However, at questions on computer infrastructure they were also abandoning the survey prematurely if this module was compulsory (34% of those who were answering the questions from this module which was assigned to them as compulsory, abandoned the survey at one of these questions). In addition, they were most often abandoning the survey also at compulsory modules on attitudes toward the Internet, on Internet, state and politics and on Internet use.

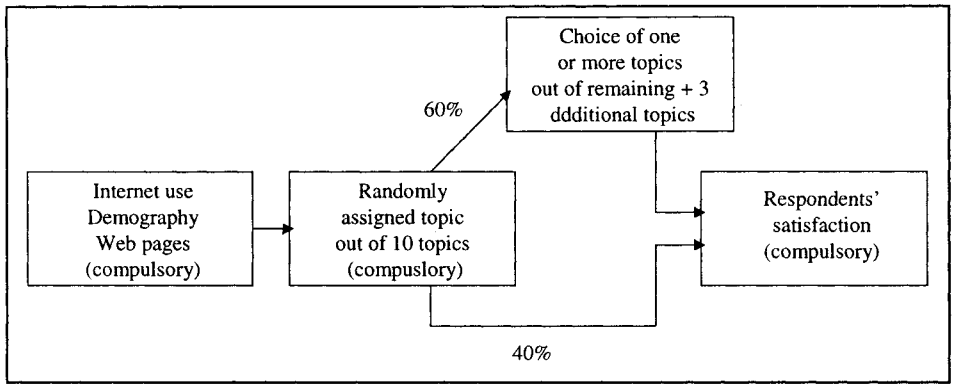

Figure 2: Experiment with survey topics (RIS98)

	% choosing the module	% abandoning (compulsory module)
Activities in spare time	10.2	2
Advertising on Internet	5.2	25
Erotics and pornography on Internet	12.2	/
Attitudes toward Internet	6.2	33
Computer infrastructure	12.4	34
Electronic commerce	6.5	23
Internet and its costs	7.9	/
Internet and media	4.3	23
Internet software	9.7	21
Internet, state, politics	7.4	31
Medicine, pharmacy and Internet	3.2	25
More about Internet use	8.3	28
Pop culture	7.9	/

Table 2. Response at different topics (RIS98) (/ is written for modules which were not compulsory)

INSTRUCTIONS

In self-administered interviewing respondent's burden is larger in comparison to interviewer administered surveys. In addition, the respondent's burden is even larger in CSAQs such as Web surveys. Namely, he/she needs to know two things: how questionnaires are answered and how computers and Internet are used. When answering a specific question respondents need to understand whether one or more answers are possible. This can be additionally stressed by the use of graphics (e.g. 'radio buttons' where only one answers is possible and 'check boxes' when several

answers are possible, suggested by Jenkins and Dillman (1996)). If such a pattern is used consistently through the questionnaire the respondents learn how to answer the questionnaire easier and faster, therefore his/her burden is smaller.

Sometimes however, explicit instructions are needed, as computer literacy varies. In the RIS98 Web survey an experiment regarding the instructions was implemented at the first survey question. One experimental group (n=5173) was assigned a question with detailed instruction how to answer this question and the following ones. The other experimental group (n=5221) was assinged a question with only basic instruction that several answers are possible. The results of this experiment are reported for all users accessing the questionnaire, also for those abandoning the survey soon after the first few questions, which were otherwise counted as nonrespondents.

The hypothesis is that those receiving detailed instructions will check more options. However, results show that there was no impact on number of items marked (on average they marked only one answer). In addition, for the following four questions more people abandoned the survey when detailed instructions were offered. We can conclude that respondents in Web surveys, especially in self-selected surveys as in our case, are usually experienced Internet and computer users and have no difficulties in completing forms. However, additional experiments are needed.

DISCUSSION AND FURTHER RESEARCH

In this paper, four Web survey design features and their impact on partial and item nonresponse were discussed. We showed that partial nonresponse (people abandoning the survey prematurely) was influenced by the use of advanced graphics (in our case logotypes when visits of WWW sites are measured), survey topic, and detailed instructions.

- When logotypes were used for measuring WWW sites visits, people were more often abandoning the survey.
- Regarding the influence of survey topic, questions on computer infrastructure, attitudes toward the Internet and on the Internet, state and politics were those who had most negative influence. However, when discussing the impact of different topic modules on partial nonresponse we should note that differences among them were not only in the topic of the questions, but also in the number and types of questions. It is difficult to establish what was the exact reason for, for example the module 'Attitudes toward the Internet' to have such negative impact. It might be that topic was not really interesting, but also that questions were too difficult, there were too much questions or maybe some other reason.
- Detailed instructions also had surprisingly negative impact: people were more often abandoning the survey if detailed instructions were offered at the first survey question.
- Design where each question was put on its own HTML screen did not stimulate larger partial nonresponse in the RIS96 survey. However, the questionnaire was relatively short in this case and extensive negativne comments from the RIS98 survey where only screen-based design was used, suggest that additional experiments are needed in order to establish the adequate number of questionnaire screens and number of questions per screen.

Use of logotypes had an impact also on item nonresponse, however positive in this case: item nonresponse was slightly smaller when logotypes were use. Item nonresponse was also smaller in screen-based design. Detailed instructions had no impact on item response: for the question where extensive instructions were offered, the average number of item marked was the same in both experimental groups.

Use of logotypes not only influenced partial and item nonresponse but also respondents' answers. Respondents were more often inclined to admit that they do not know the site when logotypes were used.

We must repeat that the above experiments were performed on two surveys which did not use probability samples of units which would be representative of a larger target population. Respondents were intensive Internet users who self-selected themselves to participate in the survey. In some case, for example regarding the effect of detailed instructions, results might be different if less intensive users were surveyed. Nevertheless, we belive that results are informative.

There are still many design features impact of which has not been sufficiently studied. One of them is already mentioned dillema whether to use scrool- or screen-based design. The advantage of screen-based design is that if a person abandons the survey prematurely, at least answers given up to the particular screen when he/she leaves, are saved. However, especially in longer questionnares, screen-based design is very anoying for respondents. Probably a middle solution is reasonable: splitting the questionnaire on several HTML pages, but having several questions on each page. Kottler (1997) for example suggest to put as many as possible questions where not skip patterns and internal controls are needed on one page. For more complex set of questions, screen-based design is recommended. Similarly Dillman and Bowker (1999) suggest to construct Web questionnaires so that " ... they scroll from question to question unless order effects are a major concern, large numbers of questions

must be skipped, and/or a mixed-mode survey is being done for which telephone interview and web results will be combined".

Another problem that came out at the RIS98 survey by looking at respondents' comments at the end of the questionnaire was the lack of progress indicator; respondents missed an indicator of how many questions are still left. What kind of indicator should be used: text or graphichal? Will the indicator positively or negatively impact the partial nonresponse?

Impact of instructions was also not sufficiently studied in the RIS98 survey. There are several possiblities how to implement instructions in the questionnaire: detailed instructions provided at first few initial questions, hyperlinks to special help screens or two versions of the questionnaires, one for less experienced and one for more experienced respondents (Dillman and Bowker 1999). These modes may differently affect the data quality and their impact should be studied.

Another features are forced reminders and checking controls which in theory should lead to higher quality of data. However, they might be annoying for respondents or their implementation may demand newest versions of browsers not used by all respondents.

With the largest use of panels, it is not important only to study adequate design in order to get high quality of data, but also to increase respondents' satisfaction and therefore their willingness to participate in future Web surveys. All the above discussed design features may be relevant also when studying the respondents' satisfaction.

REFERENCES

Clayton, R.L., & Werking, G.S. (1998). Business surveys of the future: The World Wide Web as a data collection methodology. In M.P. Couper et al. (Eds.), *Computer assisted survey information collection* (pp., 543-562). New York, NY: John Wiley & Sons.

Couper, M.P., Traugott, M., & Lamias, M. (1999). *Effective Survey Administration on the Web*. Paper presented at the 1999 MAPOR Conference, Chicago, Illinois.

Dillman, D.A. (2000). *Tailored design of mail and other self-administered surveys*. New York, NY: Wiley-Interscience.

Dillman, D.A., & Bowker, D. (1999). *Principles for the design of Web surveys: A review of current practices and the need for change*. Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida.

Farmer, T. (1998). Using the Internet for primary research data collection. *Market Research Library*. Available: http://www.researchinfo.com/library/infot ek /index.shtml.

Groves, R.M., & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York, NY: John Wiley & Sons.

Jenkins, C.R., & Dillman, D.A. (1996). Towards a theory of self-administered questionnaire design. *Working papers in Survey Methodology*. U.S. Census Bureau. Available: http://www.census.gov/srd/www/ byyear.html.

Kojetin, B.A., Borgida, E., & Snyder, M. (1993). Survey topic involvement and nonresponse bias. *1993 Proceeding of the Section on Survey Research Methods. Volume II* (pp., 838-843). Alexandria: American Statistical Association,

Kottler, R.E. (1997). *Web surveys – The professional way*. Paper presented at ARF Conference, New York, NY. Available: http://www.spss.com/spssmr/web_bureau/knowledge/w ebserv.htm.

Martin, C.L. (1994). The Impact of topic interest on mail survey response behaviour. *Journal of the Market Research Society*, 36, 4, 327-338.

Onyshekvych, V., & McIndoe, D. (1999). *Internet technology: Gaining commercial advantage*. Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida. Available: http://www.ronincorp .com/GAININ~1/index.htm.

Saris, W.E. (1998). Ten years of interviewing without interviewers: The Telepanel. In M.P. Couper et al. (Eds.), *Computer assisted survey information collection* (pp., 409-429). New York, NY: John Wiley & Sons.

Vehovar, V., Batagelj, Z., Lozar Manfreda, K., & Zaletel, M. (forthcoming). Nonresponse in Web surveys. In R.M. Groves et al. (Eds.), *Survey nonresponse*. New York, NY: John Wiley and Sons.

Zukerberg, A., Nichols, E., & Tedesco, H. (1999). *Designing surveys for the next millennium: Internet questionnaire design issues*. Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida. Available: http://surveys.over.net/method/ zukerberg.ZIP.