

Analyses of Web Survey Data

Vesa Kuusela, Social Survey Unit, Statistics Finland

1. Introduction

Data capture process of a self-administered questionnaire is inherently different from that of an interview. That has many implications on the settings in which data capture takes place. In a web interview (CAWI), the whole response process differs from that of an interviewer-administered interview. In CAWI respondent answers alone without any assistance or control, and sending the filled forms requires only a click of a button. In addition, the different implementations of web based data collection change the settings of responses process: in a static web questionnaire only a limited amount of checking can be implemented and most of checks are run only when the form is received in the web server. In an interactive questionnaire there is a lot more possibilities for checks and checks are run after a page has been sent and data is stored after each page. In an interactive questionnaire also question order can be defined which is not possible in a static form. The settings of the web based data collection produces such errors in data that do not exist in other modes.

In an interviewer-administered survey the interviewer controls the situation and what is entered in the form. If a respondent agrees to be interviewed he or she usually contemplates the questions and answers. Very rarely the answers given in an interview are completely nonsense. In any case, the interviewer is able to assess the feasibility and reliability of the answers. Breaking a started interview does not happen very often. If a respondent has agreed to be interviewed in most cases he or she does not interrupt answering because of minor discomforts. Consequently, the data obtained from an interviewer-administered data capture – using a well-defined instrument – will be ready for editing and analysis after minor checking. In addition, unit non-response is coded adequately and may be processed using available methods. There may occur item non-response, though. Usually it remains in a moderate level, however.

A web interview differs also from that of a postal survey even though both are self-administered. In a postal survey, like in a web survey, the data capture process cannot be controlled. However, in a postal survey, the forms have to be returned to the agency or company that is conducting the survey and that requires some effort. Sending a static web questionnaire is usually a very simple task and when using an interactive web questionnaire the data is stored all the time requiring no special effort. In a postal survey the returned forms are usually checked before data entry. The validity of the data is assessed (and screened) during checking and also during the data entry from the forms. Therefore in a postal survey, the data to be analysed mostly is composed of answers, which can be regarded as given seriously and with consideration. In a web survey one cannot be as sure of that.

The different response situation has given rise to new types of errors in survey data and that calls for new types of data cleaning methods. The new problem cases should be filtered out or otherwise be taken into account in the data analysis of the survey. Theoretically interrupting the data capture leads to item non-response. It is slightly different from conventional item non-response where values are missing here and there. In an interrupted filling of questionnaire all the answers after the “break point” are

missing. Some respondents are interested in what is asked but are not interested in giving well-considered answers to them. A form that is filled in only to browse through the questions is more problematic. Basically it should be classified as unit non-response but it may require sophisticated methods to single them out from “serious” answers.

The new problems of the data quality have not received much attention, yet. For example Dillman (Dillman, 2007) or Reynolds et al. (Reynolds et al., 2007) don't recognise nor discuss them. It is difficult to know whether there has been any research touching this problem because there doesn't exist many reports. The aim of this study is to find out whether breaking and browsing exists and if yes, to what extent.

2. Material and Methods

The material for this study comes from a survey conducted by an afternoon paper in Finland. On its Internet site was general invitation to readers to fill in a questionnaire on its site. The questionnaire consisted of 20 statements comprising a psychological profile of respondent's emotional intelligence. Statements dealt with the traits of character and respondent was asked to answer whether the trait did not describe him or her at all, did not describe well, described fairly well, described well, or described him or her very well. Many of the statements were contradictory in such a manner selecting the first option (say) for each statement could not be true for one person. For example “I usually plan my activities well ahead” and “I rather improvise than plan carefully”; “I often take risks only basing on my feelings” and “I rely more on facts than feelings”; or “I often show sympathy to other people ” and “I refrain from showing emotions that might impede my performance or my human relations”. However, the formulation of statements was not done to facilitate the screening of inconsistent responses.

The questionnaire was a simple page based form and there was no consistency checking for given answers. The statements were given in blocks (or pages) of four. The “survey” had a general invitation without any identification of respondents. That means that there was no control of who responded or how many times a response was given. Neither the time nor duration of answering was recorded. The only background variables were respondent's gender and age class.

The questionnaire was open for four weeks. A total of 202 411 responses were recorded in that time. Women answered slightly more frequently than men (53.9% vs. 46.1 %). Young respondents, under 25 years, were the largest group (43.4%) and second largest group were young adults between 25 and 44 years (39.2%) and middle aged or older readers responded most infrequently (17.4% of responses). Probably, the demographics are typical to this kind of a CAWI survey but do not represent the population neither the readers of the paper.

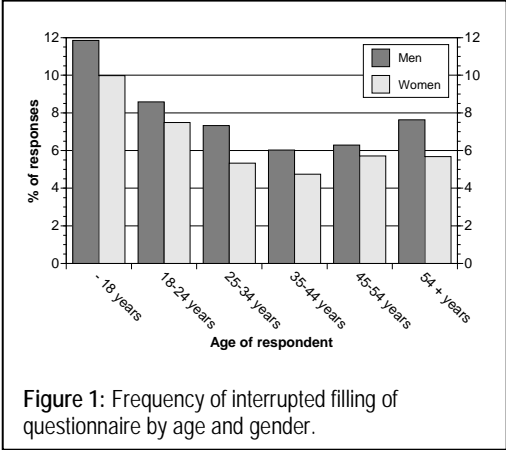
3. Results

The analysis is focussed only on interruption of filling and finding out potential browsing of the questions. This was done using only fairly simple criteria. Very small variation between answers was assumed to indicate that the answering was not based on deliberation. High variation also may show browsing but it is more sensitive to erroneous classification. Also the use of the extreme answer categories only may be an indication of browsing. The variation was analysed using standard deviation, s , as a

criteria. Also “too” small or large mean value, \bar{x} , may indicate browsing as well as the values of the coefficient of variation, $100s/\bar{x}$. More sophisticated criteria could be applied but they were not tested in this study.

3.1. Interruption of interview

An answer to all statements were given in 92.55% of the responses, and in 7.45% of cases the filling was discontinued at some point. Most often the interruption took place after the first block of four statements (4.05% of all responses). As the questionnaire was arranged in blocks of four statements, in most cases the interruption took place after a block or after receiving a new page.

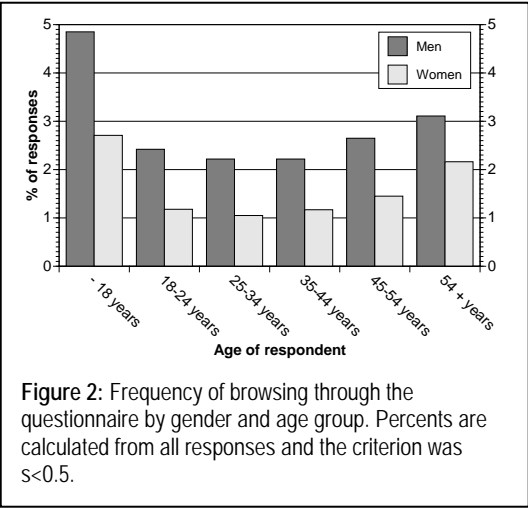


The male respondents interrupted filling of the form more frequently than females (8.25% vs. 6.75%). Also among the young respondents interruptions occurred more frequently than among the older ones (see figure 1). In each age group the interruption frequency of men was higher than that of women. Probably this was partly related to the topic of the questionnaire that can be guessed to interest more women than men.

3.2. Browsing through the questionnaire

Browsing means that the respondent does not answer to the questions after reflecting them or does not indicate his or her true reaction to a statement. Instead, they enter some answer only to move ahead in the questionnaire. Obviously there are many different styles to give answers to reach this goal. For example, selecting always the first alternative or the same alternatives that are vertically (or horizontally) lined up. The other possibility is to select in turns the first and the last option (e.g.). Certainly there are also other styles, but these two styles may be identified using simple statistical criteria: the first one minimises variation and the other style maximises it.

In 1.13% of all responses there was no variation between answers (standard deviation was 0) indicating that all the answers had the same value. When only the complete answers (answer in each of the 20 statements) were analysed, 0.97% of the cases had no variation. This indicates that those who are not really answering to the statements are more inclined also to interrupt the filling. In most of the cases without any variation the selected answer category was the



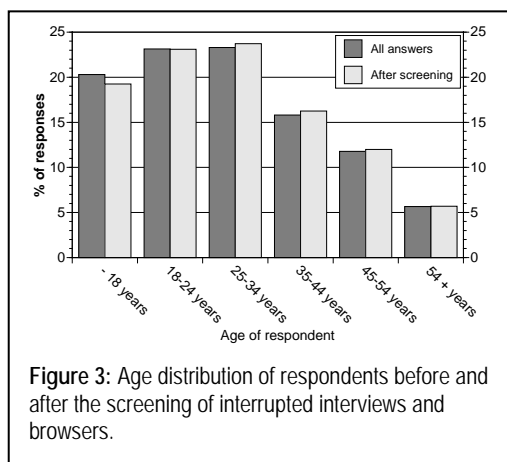
one in the middle. Also cases where the only selected answer category was the first one were common. Logically or psychologically neither of them can be true.

Using slightly wider criterion ($s < 0.5$), obtained when one or two of the answers were different, led to exclusion of 2.17% of all cases and to exclusion of 2.01% of completed cases. This criterion may have led to omission of few “real” answers, but it is not probable. The browsing showed similar age and gender pattern as interruptions. Men were browsing much more frequently than women and young respondents were browsing more often than older ones. In figure 2 is shown the frequency of browsing. Percents were calculated from all respondents using the wider criterion for variation (i.e. $s < 0.5$). However, the pattern was similar if percents were calculated only from completed cases or using the stricter criterion.

Using “too” large standard deviation as a criterion seemed to single out also valid answers. With slightly different formulation of statements had enabled also this criterion. In this case it had required adding some logical criteria to screen only implausible answers. On the other hand, coefficient of variation (CV) screened some illogical cases: by using criterion $CV < 15$ for coefficient, 2.23% of all responses were screened and 1.77% of complete responses. Nearly all of the screened cases were such that all given answers were close to the centre of the scale. Some of these cases would have been screened also by using the standard deviation criterion, but not all.

The mean value of answers could also be used as a criterion: too small or large values indicate the greatest part of answers is close to either end of the scale. However, it proved that mean is influenced by extreme values too much to be reliable.

The different criteria, which were tested here, overlapped to some extent. Using the strict criteria for screening browsers and including only complete cases left 89.67% of the cases. It is probable that very little, maybe none, of truthful answers were screened out. It is more probable that all invalid cases were not screened out. Using slightly wider criteria decreased the number of accepted responses considerably but at the same time the number false negatives seemed to increase relatively fast.



The applied screening criteria left false positives in the data but it is not known how much. In order to minimize the probability of false positives and false negatives more sophisticated screening methods had been needed.

In Figure 3 is shown the influence of the strict screening on the age distribution of respondents. After screening the proportion of young respondents decreased and the proportion of middle-aged respondents increased. Additionally, the proportion of women increased from 53.87% to 54.61%.

4. Conclusions

In a web survey, only partly filled forms, or interrupted interviews, is a new phenomenon that occurs only rarely in interviewer-administered surveys or in postal surveys. An interview, once it has been started, is rarely interrupted, and in a postal survey partly filled mail questionnaires usually are not sent or the data is not entered. Web based data capture takes place in different circumstances, which result in interruptions in the filling in the questionnaire. With a static web questionnaire this is a lesser problem because there is a separate “Send” button, which has to be clicked. In an interactive web questionnaire interruptions may become a problem because answers are stored all the time more or less automatically.

The aim of this study was to find out how difficult problems breaking an interview or browsing questions may pose, but the results of this small study should not be generalized too quickly. The analyzed web survey was an unrestricted self-selected survey, close to an entertainment poll, with a general invitation. In a CAWI survey carried out by research agencies and with a more “serious” topic, respondents both open the survey with more seriousness and also pertain more to thought when they answer the questions. However, both of these phenomena probably are present in all web surveys in smaller scale. Based on this study, it seems that young respondents and men are more inclined to interrupt a web interview.

The main reasons leading to interrupted interviews are loss of interest and frustration. They originate partly from technical factors but more probably they originate from the questionnaire itself: Unpleasant layout, badly formulated questions, unfitting answer categories, difficult structure, etc. most probably frustrate some respondents to a point where they stop.

The response process in a web survey is more sensitive than in an interview. A real interview once started is not interrupted because of minor discomforts but that danger always exists with self-administered questionnaire. Careful planning of instruments is of great importance in avoiding interruptions. In a web survey it is even more important than in a postal survey because in a postal survey the forms do not end up in the data file. A problem with interrupted interviews may be that they may remain unnoticed by the checking procedure because there is a record stored although it is incomplete.

Answers, which have been stored after browsing through the questionnaire, are more problematic than interruptions because they are more difficult to locate. In addition, they may cause bias in the results if they are not found out. The magnitude of bias depends on how much browsing has occurred. Basically, responses obtained after browsing through a questionnaire should be coded as unit non-response.

In designing the web questionnaire it is possible to make “traps” that would catch browsers. If it is possible the answer alternatives should be selected in such a style that browsing through a questionnaire leads to inconsistencies between answers. This way they could be identified in data checking or even during data capture. However, it is not certain how well the traps would work out in real situation. In addition, duration of the interview may provide information on how serious the respondent was.

Breaking the interview and browsing through the questionnaire are not completely separate phenomena. A respondent only browsing the questions will stop when he or she notices that the questions are not interesting enough.

It is quite obvious that the data collected via the web needs a thorough checking, much more than interview data and more than data from a postal survey. In addition, the checking procedure has to be designed on different basis than when checking data collected in other modes.

5. References

Dillman Don, A.: Mail and Internet Surveys. The Tailored Design Method. 2nd edition, 2007, Wiley.

Reynolds Rodney, A., Woods Robert, Baker Jason D.: Handbook of Research on Electronic Surveys and Measurements. 2007, Idea Group Reference.