

## Validity of Pre-testing Web Surveys

Dirkjan Beukenhorst and Rachel Vis\*

*Statistics Netherlands*

In Statistics Netherlands' contribution for the Washington Quest workshop in 2001 we dealt with the problems connected with testing CASI (Computer Assisted Self Interviewing) questionnaires. We compared different testing methods. On the one hand classic methods: respondents thinking aloud while interviewers observe and ask meta-questions concurrently or retrospectively, and on the other hand 'self-administered meta-questions' inserted in the CASI-questionnaire proper. We also stimulated spontaneous comments in order to find unexpected problems. Our preliminary conclusions were:

Computerized self-administered evaluation yields useful results. Respondents are definitely willing to answer meta-questions in a computerized format. They are also willing to add their own comments into the computer spontaneously. As could be expected, respondents who are more familiar with computers are more inclined to do so. If the functionality of the interview program becomes easier, even less experienced computer users will probably do so.

Both quantity and quality of the problems found by the computerized evaluation depends strongly on the quality of the meta-questions asked. The results of the interviewer probing afterwards show convincingly that not all relevant problems were found with the computerized evaluation tool we used. Especially unexpected problems can be more difficult to detect, as there are no meta-questions asked about it. Also, if meta-questions are too broad, the information they yield is less detailed.

Another interesting finding of our test is that respondents seem to make more mistakes when completing the questionnaire by themselves (without an interviewer present encouraging them to think aloud). **This in itself can be seen as an advantage of the computerized method, since it represents the more realistic situation.** Obviously, this is only advantageous if the mistakes can somehow be detected. This could be accomplished by programming smart control questions, consistency checks and feedback questions (e.g. 'According to the previous questions you used the internet for the following activities. Are there any other activities you would like to add now?').

In 2001 we were obviously primarily interested in the traditional pre-test topic of the cognitive aspects of the questionnaire. The problems we were looking for and wanted to solve were about interpretation, information processing and so on. How the respondent could handle the questionnaire was of no concern: the interviewer started the questionnaire and saved it after the respondent stopped answering the questions. This is not surprising, since at that time we were only thinking of CASI as an interview in which the respondent answers the questions on the computer *while an interviewer is present*. The concept of 'internet questionnaires' was still vague. In that period Statistics Netherlands had no facilities at all for conducting web surveys. So there were no usability-issues to test, yet.

In our present paper we will describe how our attention has been shifting from these cognitive

---

\* The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

preoccupations towards more affective aspects and technical problems concerning the usability of different variants of web surveys. This shift brought along a change in our methodological concerns as well. We will discuss four pre-tests of electronic questionnaires that were conducted between April 2004 and January 2005. These will illustrate this shift quite clearly. In these tests we were primarily interested in the following questions:

What is for SN the best way to administer household surveys electronically?

What is the best way to test whether people can administer the electronic questionnaires by themselves, i.e. without the help of an interviewer?

The first two tests concerned a computer assisted self-completion questionnaire on CD-rom, and the other two tests concerned an on-line questionnaire.

In the first three of the four tests all test respondents were SN colleagues, who were asked to complete the questionnaires at home. They received an advance letter with a short instruction on how to open, fill in and return the questionnaire. In the cases where a CD-rom was used, it was included in the advance letter. Also included in the advance letter was an evaluation form with open-ended questions on what the respondent thought of the questionnaire. After completing the questionnaire they had to fill in the evaluation form and return that (via mail or personally) as well.

Most colleagues were very critical and provided a lot of comments. Besides, many colleagues came to hand over the evaluation form personally and discussed the methodology of the survey with the researchers. Often the short discussions were more informative than the comments written in the evaluation forms.

For the last test we recruited “real” test respondents from outside the bureau, and we were careful to include people with little computer and internet experience. In this test we used two different test methods: Firstly, twenty test respondents were asked to complete the online questionnaire at home. They received an advance letter with a short instruction and an evaluation form they had to return. Secondly, ten test respondents were invited to come to the laboratory to complete the questionnaire, so we could observe them. They also received an advance letter with the short instruction, but the evaluation was done in a cognitive interview after the questionnaire was completed.

We found that the returned evaluation forms were not very informative. Most respondents gave monosyllabic answers to the open-ended questions; “good”, “fine”, “okay”. Looking back, this is not surprising at all and we should have known better.

The advantage of the lab test was that we could observe how the respondents reacted to what happened on the computer screen. We also got a lot of feedback on the questionnaire and also the layout. Yet this test was very obtrusive and it was impossible to recreate a situation “as if the respondent was at home”. For instance we used a laptop for the test, because that was the only way to have a connection to the internet in the laboratory, but some respondents had never worked with a laptop before.

These kinds of tests differ quite evidently from traditional tests in the lab. In these tests we chose explicitly to use a questionnaire that was already pre-tested and used in the field because we did not want to be bothered with comprehension and other cognitive problems.

Concerning the methodology we discovered some new problems which we did not solve as yet. In the Washington paper we were already aware of the ecological validity problems when you test self-completion questionnaires (we concluded that the self-completion pre-test represents a ‘more

realistic situation'). All test respondents are sensitive to the fact that an interviewer/researcher is present (or behind a one-way screen) during a test. This becomes a real problem if you are testing a self-completion questionnaire, especially if you want to investigate motivational aspects. Motivation to complete the questionnaire is one of the most serious problems of web surveys: the drop-out rate can exceed twenty percent! Think of the Hawthorne effect: if your test person knows he is part of a special research project his motivation will already grow.

Now we have reached a dilemma. Only completely unobtrusive methods (such as audit trails or placing cookies) will give ecologically valid data on motivation. However, these methods are considered by many as unethical – especially on the Web.

Another methodological (and very practical!) problem is this: The richest data on cognitive problems in a questionnaire (self-completion or not) will be found in a lab test with an interviewer. The best way to test technical problems of a web survey is by letting respondents test at home without a researcher being present. These two conditions rule out the possibility of a test procedure consisting of only one test round.

Finally, we discovered a rather frustrating side of testing web surveys. If we test a questionnaire and find cognitive problems with question wording or the sequence of questions, we can formulate proposals for improvement ourselves, and in the same test round we can implement our proposals. Yet if we detect problems concerning IT-aspects like the lay-out or navigation, or the handling of the questionnaire, we can only report our findings or even suggest improvements, but we cannot implement the changes and sometimes we don't even know whether an improvement we suggest is feasible. In this we depend on the programming department. This also can be a reason to rule out the possibility of a test procedure consisting of only one test round.