

Usability Testing and Evaluation of Questionnaire Re-design Work

Petri Godenhjelm¹, Jussi Rouhunkoski

¹Statistics Finland, Työpajankatu 13, 00022, Tilastokeskus, Finland

Abstract

Statistics Finland had a two-year project on the evaluation and testing of business web questionnaires, and to redevelop them and give guidelines for future work on questionnaire design and testing. The usability testing was done through cognitive interviews with the concurrent think-aloud method and recording of all interviews. The screen recordings were done with the Dream Screen software, which produced the video clips with audio during the answering process. The video material and usability guidelines guided the re-design work of the web questionnaires afterwards. The results of the actual data collection and response burden measurement were evaluated after re-design work and the focus was in the areas where changes have been made. By using quantitative data parallel with qualitative it is possible to have more holistic view of the effectiveness of development procedures.

This paper discusses the experiences gained from the testing and from re-designing work, which appeared to be the most challenging part of the work. Moreover the first quantitative results of next data collection round in re-designed questionnaires and the results of response burden measurement give a possibility to assess the effects in versatile manner. An Internet data collection method has been available to all over 65 major and permanent data collections from the end of 2006. This means that the web questionnaire designing was done mostly in the early years of the 2000s. In addition, the design work was guided by paper questionnaire development culture in the organization. Since then the web questionnaires in general have been developed in many ways and people's expectations on the web applications have risen. Although the data collections via the Internet is regarded as successful in many ways, continuous evaluation and development of the web questionnaires are still needed to ensure high quality of the collected data and to keep the respondents motivated.

Key Words: usability testing, questionnaire design, response burden

1. Internet data collection and response burden

An Internet data collection method has been available to all over 65 major and permanent data collections from the end of 2006. In 2011, the share of web surveys in enterprise data collection has grown to 73 per cent. As a whole, web collection has been considered as a success in many ways. The average response time of surveys has reduced and the quality of received data has been better. The respondents have also been satisfied.

The EU action plan aims to measure costs arising from EU legislation and to reduce the administrative burden by 25 per cent by 2012. As part of the action plan, Statistics Finland run a programme for developing business data collections in 2007-2011. The purpose of the programme was to reduce the response burden and develop respondent relations and motivation. This included also extensive response burden measurement.

Under the work on response burden measurement, the survey questionnaire on response burden was pre-tested. To be able to test the questionnaire draft, it was essential that the respondents answered to the actual on production questionnaire at hand. The test produced valuable information on developing the web questionnaire on response burden. It was also concluded that there is still some work to be done in designing user friendly and high-quality questionnaires that are in production already. Many web survey elements and technology have also developed from the way the first waves of web questionnaires were designed.

This development work also made it possible to use these qualitative usability tests as material to evaluate the re-design work, which was made between first and second round tests. Together with quantitative response burden measurement, it was possible to use both data as an evaluation framework.

2. Response burden measurement

Statistics Finland and the National Board of Customs together measured response burden of the direct enterprise data inquiries during 2008-2009. On average 250 data suppliers responded response burden inquiry in each data collection. The majority of respondents used a web form but some returned a paper form. Responding to the response burden questionnaire was voluntary. It was possible to reply to the burden measurement survey on a hard copy form or by email, but the companies mainly chose to answer the questionnaire when it was appended to an electronic data collection form. The aim is to repeat response burden measurements regularly. Some individual surveys have already been measured twice. Overall, the burden measurement 2008-2009 consisted 41 surveys and over 13 500 responses. The response burden questionnaire is in appendix 1.

3. Usability testing

The results of the response burden measurement were an important input to the decision process on which questionnaires will be tested. The usability tests were done during visits to the company. The subject area experts were encouraged to take part in testing as this made it possible to get valuable information on subject issues as well. In the answering process, subject and usability issues are more or less intertwined so it is good that the subject area personnel could elaborate these issues at the same time. Yet, the testing protocol was designed so that the respondent should try to figure out the answering task without help and concurrently thinking aloud. Only after the questionnaire module was finished,

elaboration by subject area experts was allowed. In addition, a more holistic elaboration was done at the end of whole questionnaire.

The method applied was a usability test, which has a close resemblance to cognitive interviews where respondents are instructed to think aloud while answering the questionnaire. The aim was to go through the whole questionnaire but in some occasions, this was not possible because all the information or the personnel needed were not available or there were time constrictions. In those situations, a test procedure was arranged so that all parts of the questionnaire were evaluated at least partially. Thus, it was possible to get an overview of the entirety of answering tasks. In the second round of testing, the focus was in the parts that were redesigned.

All tests were done by using the screen recording software Dream Screen. The video clips recorded with the software are uploaded to the server with a password protection from where they can be shared with the others involved in the development. With real response process observations, it was easier to make the subject area experts aware of the problems and to make the development work smoother. Finnish Affiliates Abroad data collection was the first questionnaire where usability tests were applied before and after the re-design work.

The concept of usability consists of three dimensions: effectiveness, efficiency and satisfaction in a specified context of use. ISO standard 9241 on usability recognizes each of them. Usability approach emphasizes the importance of empirical usability tests. It assumes that designers of an application and questionnaires as well are blind to the aspects of novice user and respondent. Still a survey application should support response process as is figured out on the hybrid response process model (Willimack & Nichols 2010). All the analyses were done according to these usability concepts and the hybrid response model framework. Both formed the structure about how these recordings were perceived and interpreted. Even if the web questionnaire and usability issues were in main focus, the other parts of the response process emerged as well in test situations. In the first round of empirical testing, the focus was not on any special part of the questionnaire but in the second round, it was more on aspects that were found and developed between the testing rounds.

The overall objective of testing was to determine how the respondents fulfil the tasks the questionnaire entrusts to them, how do they navigate and how do they understand the questions and response options. One goal was also to find out how the respondents use different kinds of information systems and other records in providing answers. It was examined whether the information was stored in the company in the same way as it was asked in the questionnaire and how the person responded to the questionnaire.

3.1 Re-design

The most burdensome questionnaire also has the most difficult design implementation process. The design process needs different range of expertise on subject area, software, usability testing, and questionnaire design and data collection. The software in use defines more or less what is possible to do. This means also that communication between subject area experts, methodologists (who does usability tests), and software developers and architects plays an important role in a whole process.

All interviews clearly showed how the layout and symbols guided the navigation on the questionnaire and, at the same time, the comprehension of the questions as well. Instructions contain also information that is relevant to answering. It is important that designers consider thoroughly what information is the most important to communicate to ensure that the received answers will be comparable with other respondents. The designer must make a strategic decision about what, how and how much information should be communicated to the respondent. The meaning of concepts is communicated not only in questions, but in definitions and instructions as well.

Intended order, in which answering should progress, caused problems in some of the tested questionnaires. In the web questionnaire design, there are efficient ways to direct the response process but it should be taken care at the same time that you do not design things that will easily make the respondent to get lost in the web questionnaire structure. Good design communicates the status of the respondent's progress and the functionality of different buttons through the form. All these principles were kept in mind during the re-design process.

4. Evaluation of re-design work

The evaluation of the success of the re-design work was of two kinds: second round qualitative usability testing with company visits and open ended questions in response burden measurement, and the quantitative items in the response burden measurement questionnaire. The second response burden measurement (sample size 866, number of responses 274) on the Finnish Affiliates Abroad data collection took place in late 2011.

4.1 qualitative data

The qualitative information from the open-ended question in the response burden questionnaire gave important feedback not only to the data collection itself, but also to other parts of response process that is seen in hybrid response process model by Willimack and Nicholls (2010). For example, record formation in businesses did not always contain the information that was asked in questionnaire. It was also seen how challenging it was to gather all the necessary information to be able to give answers. One common message was that instructions and

definitions were difficult to understand and used classifications induced problems to respondents to find a proper one in a web form with the application.

In the qualitative data of the response burden measurement, there was no such comprehensive information on questionnaire form problems. These were found more clearly in usability testing on the second round when it was possible to focus also to the parts of the questionnaire in which were done some changes. Overall, this qualitative information gave a more profound and versatile picture of the questionnaire in hand. By using this qualitative data in the interpretation of quantitative results of the response burden measurement, it was possible to understand the underlying mechanisms of answering process more nuanced way.

4.2 quantitative data

The overall results indicated clear differences between the first and the second round measurement. The overall time used on acquiring the data and answering to the questionnaire was over 20 per cent below than in the first measurement round. The average time spent on acquiring and processing data prior to their actual entry onto the questionnaire was in the first round 128 minutes and in the second round 108 minutes. The actual average time in answering to the questionnaire decreased from 70 minutes to 49 minutes. The perceived response burden question showed that the share of 46 per cent of respondents found that inquiry was fairly or very easy in the first round compared to 66 per cent in the second round measurement. However, the larger businesses showed little more response burden in every aspects of response burden.

The more detailed response burden issues on Finnish Affiliates Abroad questionnaire are shown in figures 1 and 2 It can be seen that especially in the clearness of definitions and instructions progress was made with the questionnaire. In addition, in difficulty to decide which answer would be correct, favourable development was found.

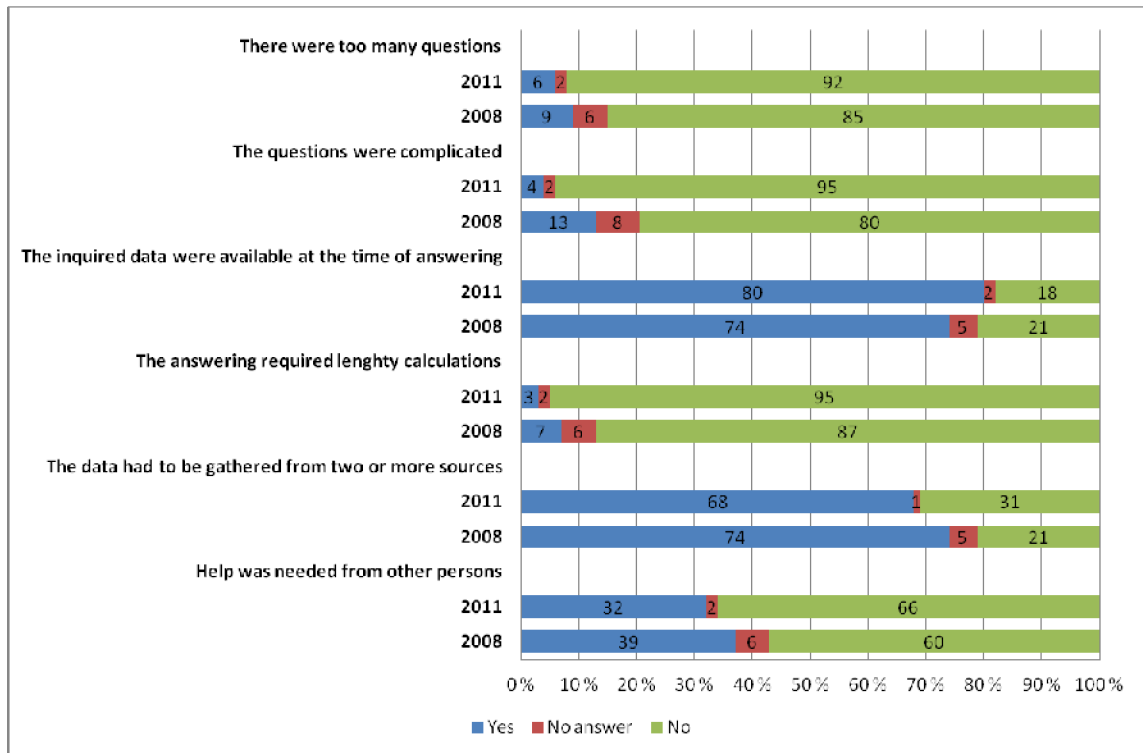


Figure 1: Shares of responses to the different statements of response burden (1)

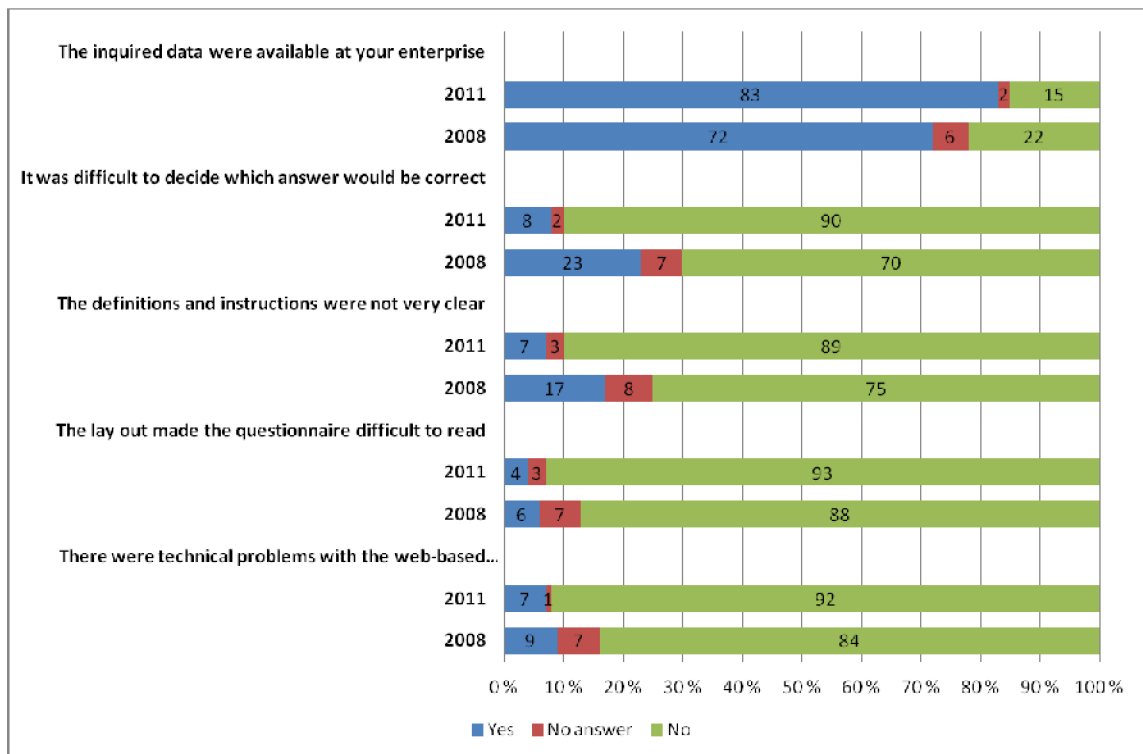


Figure 2: Shares of responses to the different statements of response burden (2)

5. Concluding remarks

The results of the response burden measurement clearly showed that it is possible to have positive impact on actual and perceived response burden. Usability tests and a comprehensive re-design work seem to be an efficient way to decrease different aspects of response burden. The qualitative information produced by second round usability tests gave a possibility to interpret the quantitative results in a more comprehensive way. The results of the response burden measurement provided general information about the perceived functionality of response process and the usability tests focused on the issues that were a possible source for it. In the evaluation of the re-design work, quantitative or qualitative approach alone would have made the evidence on efficiency of re-design actions more distant and too narrow. However, it has to be kept in mind that these re-design and re-measurement has been done only with one questionnaire so far. After a more comprehensive work and with different kinds of questionnaires, it is possible to assess this more broadly.

It is important to underscore that the feedback received from respondents varies according to the method used. Some usability issues cannot be detected with the straight questions from respondents. For example, the respondents in the response measurement reported very little problems arising from lay out or technical problems. These design issues could be appraised only against observed behaviour on response situations. It might be enough for the respondent if she or he had managed successfully go to the end of the questionnaire. A designer of a questionnaire instead could have more goals on how a respondent should navigate in a questionnaire and what tasks should be accomplished. This could be observed only in usability tests because it would be too difficult to explicitly and abstractly describe the navigational process to the respondent and ask questions about it.

The response burden could be lessened by high-quality design and testing of web questionnaires. Good design raises motivation to give good-quality answers and reduces measurement error. Survey practitioners should constantly be aware of the development in the web environment and technology, and the changes it brings to respondents and designers.

References

Willimack, D.K., and Nichols, E. (2010). A Hybrid Response Process Model for Business Surveys. *Journal of Official Statistics*, 26, 3-24.

Inquiry about answering to the survey

Statistics Finland's goal is to make it easier for you to respond to inquiries and to reduce the time it takes to provide the requested data. We hope that you will answer these questions bearing in mind the inquiry you have just answered. Your answers will be treated as confidential. Further information is available from Johanna Leivo, +358 9 1734 3525, or Jussi Heino, +358 1734 3213.

1) How many persons participated in preparing and filling in the answers to the questions included in the inquiry?

Within your enterprise ___ persons, outside your enterprise ___ persons.

2) How much time did you spend on acquiring and processing the inquired data prior to their actual entry onto the questionnaire? (total time for all participants) _____ hours and _____ minutes.

3) How long did it take you to actually fill in the questionnaire? _____ hours and _____ minutes.

4) Have any changes been made to your information systems in the past three years to make it easier for you to answer to this inquiry?

- None
- Minor changes
- Some changes
- Extensive changes
- Don't know

5) Did you find that answering to the inquiry was

- Very easy
- Fairly easy
- Neither easy nor burdensome
- Fairly burdensome
- Very burdensome?

6) How did you find the inquiry you just answered?

- | | | |
|--|------------------------------|-----------------------------|
| There were too many questions | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The questions were complicated | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The inquired data were available at the time of answering | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The answering required lengthy calculations | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The data had to be gathered from two or more sources | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| Help was needed from other persons | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The inquired data were available at your enterprise | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| It was difficult to decide which answer would be correct | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The definitions and instruction were not very clear | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| The layout made the questionnaire difficult to read | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| There were technical problems with the web-based questionnaire | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| Other reasons that increased the burden of answering? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |

7) Feedback and suggestions for improvement concerning the questionnaires, instructions, customer service, data collection application or other matters related to the inquiry:

Many thanks for your assistance!