

Identifying and Reducing the Response Burden in Internet Business Surveys.

Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing Methods. Charleston, South Carolina, November 14 - 17 2002

Gustav Haraldsen, Statistics Norway¹

Web surveys may serve different purposes. The day-to-day costs of data collection can be reduced. The data collection period can be shortened. The response quality can be improved. And the burdens we put on the respondent's shoulders can be lightened. However, all of these effects are dependent on how the questionnaires are presented on the Internet. The status report made in 1998 by Mick Couper and William L. Nicholls II in the book "*Computer Assisted Survey Information Collection*" was this:

- Even if the data collection period has gone down, the planning period tends to go up.
- Cost reductions are only gained when the volume is high or as an effect of reorganisations in the survey institution.
- Some quality improvements are obvious; others only come after a total change of how the questionnaires work and are presented.
- Some of the tasks traditionally performed by the data collector have been transferred to the respondent. Normally this leads to lighter burdens on the data collector's shoulders and a heavier burden on the shoulders of the respondent.

(Couper et al 1998)

In this paper we will focus on the last of these conclusions. Reducing the burdens of those who are obliged to fill in governmental questionnaires is the most common argument used to raise money for new computerized data collection solution. This is particularly true in Norwegian business surveys. Most of these surveys are mandatory. The Brønnøysund Register Centre, which co-ordinate information on business and industry that reside in various public registers, has estimated that a typical Norwegian firm spend a little more than 40 hours on governmental questionnaires each year. The total response burden was estimated to be 7358 man-labour years in 2001 (Norwegian Ministry of Trade and Industry 2001). There are several schemes that may lead to lower response burdens in businesses. Firstly laws and regulations that create information needs may be simplified. Hence the need for data collection will also be reduced. Secondly the data collection initiated by governmental offices and public institutions may be better coordinated so that double reporting is avoided. Thirdly, when one needs to collect information, the information needed may be downloaded directly from the registers and databases that are already available in the companies. Finally, for the information that still has to be collected with the help of questionnaires, the forms can be made more user friendly.

¹ Thanks to Dag Roll Hansen who helped me with an earlier version of this paper. Also thanks to Trine Dale who has been my linguistic advisor when the final version was written.

To some extent it is probably because the questionnaires are the most available and the most concrete target for attack that so much public focus has been on the burden of questionnaires in business surveys. To question the need for detailed regulations is a touchy political issue. And for the respondents to complain about the questionnaires may be similar to complaining to the bus driver instead of the bus company or the Ministry of Transport when the services are reduced. But it is also documented that business questionnaires often are of a low quality. Wording and design principles that are established as golden standards in personal questionnaires have not been implemented to the same extent in business surveys (Cox and Chinnappa 2000). It is not difficult to find examples of business questionnaires that demand that businesses respond rather than ask them for information, that confront the respondents with complicated tasks, and/or ask for very detailed responses. These are all examples of elements that contribute to response burden. In our view the information and Internet technology offers new tools to solve questionnaire problems. But if one does not start with an understanding of what the problem consists of, the technology has nothing much to offer. An uncritical computerization of paper questionnaires can even make the task more burdensome for the respondents.

Qualitative development and evaluation of Internet Business Surveys

When we develop new questionnaires we usually start with focus groups or other more informal meetings with the data users. The objective of these meetings is to specify exactly what information that is needed. On this basis a first version of the questionnaire is drafted and tested with potential respondents. The final version is the result of a series of revisions and more tests. Today the most common test method is cognitive interviewing or similar qualitative evaluation methods. However, many of the business surveys that are currently moved to an Internet platform have never undergone such a development process. Therefore, qualitative evaluations based on the existing paper version seems to be the most appropriate first step in the process of moving these questionnaires from a paper and pencil environment to an Internet platform.

In this paper we will report from a series of qualitative tests Statistics Norway have made with the paper and computer versions of two business questionnaires; the questionnaire of the Monthly Retail Trade Index questionnaire and the questionnaire used to collect Quarterly Investment Statistics from companies. Both these questionnaires were presented within a Web based data collection system that consisted of six different kinds of screens:

1. A log-in screen
2. A system's homepage, which consisted of a list of available questionnaires (in this case two), a status report for those questionnaires and links to what we call feedback products. These were tables, which compared the data reported by the test person with statistics for all companies of the same kind. If these statistics were considered to be useful, we believed it could motivate the companies to use the Web version of the questionnaire. The home page was the heart of the system. This was both the gateway into the different questionnaires and the gateway out of the system after the questionnaires were completed.
3. A company fact sheet. This page contained basic information about the company that is used to classify it and to identify the address and a contact person. At each session, the respondent is asked to update this information.

4. Questionnaire pages. In these cases both questionnaires only took one screen. But similar screen designs have also been used with longer questionnaires (Haraldsen et al 2002, Nygård 2001).
5. A receipt page that stated that the questionnaire had been received and accepted by the data collector.

A download of the different screens is shown in appendix 1.

The interview guide used in the tests of this data collection system was divided into the following five parts:

1. An *introduction* about how the test would be performed. The test sessions took part in the companies. Normally one moderator who sat with the test person and one secretary who sat in the background and took notes ran the tests. Since there were no laboratory setting with a one-way mirror, it was not possible to arrange for anybody to watch and react on what was said and done during the test. Instead all tests were videotaped and commented upon after they were finished.

2. Questions about *pc- and Internet competence*. The questions we used in the tests referred to here, were of a rather general character. Later we have constructed a more concrete set of questions that are meant to measure the interest and competence of Internet users. We have found that interest is best measured by a question about how often during the last seven days one has used the Internet, rather than asking how much time one has spent in front of the Internet screen. Internet competence is measured by questions about what Internet services one has used. Presently we use a list with the following ten items:

- Using Internet browsers
- Using electronic mail systems
- Using Web-bank or similar services
- Buying tickets to cultural arrangements, travels or similar services
- Purchasing goods
- Downloading music, pictures or movies
- Downloading computer programs
- Taking part in news- or discussion groups
- Chatting
- Creating one's own Web site or other kinds of Web services

In a follow up study after the 2001 Census, which was offered both on paper and on the Internet, we found that Internet competence was more important than Internet interest for the decision of what mode to use (Haraldsen et al 2002).

3. *Think aloud*. This session was divided into two parts. First the test person was asked to fill in the paper version of the questionnaire. Afterwards he was asked to fill in a computer version. The purpose of this procedure was to try to distinguish between problems with the questions and problems with the computer presentation. During the think-aloud session, the moderator had an observation sheet where he could record how much time different activities took, what kind of problems the test person seemed to have with the questions and what kind of navigation problems he had. We tried to keep this observation sheet simple. The problems with the questions were only divided into conceptual problems, task problems and problems with the response formats. Navigational problems were either observed as the test person

moved from one part of the computerized survey system to another (i.e. from the log-in page to the questionnaire list and from the questionnaire list to a specific questionnaire) or as he moved from question to question within the questionnaire. Even with this short list of problems, however, it was rather demanding for the moderator to take notes at the same time as he watched and encouraged the test person to think aloud. It may be a better solution to let the secretary take this kind of notes.

4. *Follow up questions.* The moderator was instructed to ask for clarifications on the questions the test person had problems with during the think -aloud session. In addition the interview guide contained eight, fixed follow up questions. First the test person was asked to compare the two modes of data collection and explain why he preferred one above the other. Then he was asked how confident he was that he had given the correct answers. If he was not fully confident, he was asked to specify what was the source of uncertainty. The same kind of questions were asked about how confident the test person was about how the computer program recorded and transmitted the information typed in. Finally there were a few questions to check if the test person had noticed important information, like who to contact in case of problems with the computer version.

5. *Summing up.* Two strategies were used in order to help the test person evaluate his own performance from an outside position. First the moderator suggested some conclusions and asked the test person to comment on them. Then he asked the test person to log onto the questionnaire program once more, run quickly through it and tell the moderator what he had learned from the first trial.

This interview guide is almost identical with the guides we normally use in cognitive interviewing. The method focuses on how the respondent processes the information given in the questionnaire. During the development of computer programmes a similar strategy, utility testing, is often used. This method has, however, a slightly different focus. Just like in cognitive interviewing, the strategy is to run the test with only a few users - but to run it in several rounds. In programme development this is done at three stages in the development process; at the exploration stage, during the evaluation of beta-versions and finally at the acceptance stage. In contrast to what is the case with cognitive interviewing, the utility tests are less concerned with the cognitive process of the user and more focused on the functionality of the computer program. We think that this difference in perspective is part of the explanation why utility tests lean more towards quantitative measurement than towards qualitative evaluations. The user may find the program easy to use and may be pleased with its functionality, or he may be frustrated and dissatisfied with it. But the utility tester seems to be more concerned with the lack of program performance than with the cognitive problems of the user. We also believe that the utility tester is more concerned with the visual appearance of the computer screens than on the messages communicated by text and graphics on the screens (Rubin 1994, Nielsen 1993).

In the disposition presented below we have suggested a way of combining these two development techniques into a test programme for computerized questionnaires. Cognitive interviewing may seem to be a small part of this programme, but the cognitive testing and the evaluation test described in point 2 and 3 are in fact the central parts of the testing programme. It is during these tests that standards should be developed to help decide whether or not the computerized questionnaire should be accepted.

Suggestive test programme for computerized questionnaires

1. Exploratory studies based on tests or focus groups discussing:
 - a. Expectations about the computerized version of the questionnaire
 - b. Comments to the basic structure of the questionnaire program
 - c. Terms and ways of speaking about the topic that will be covered in the questionnaire
 - d. Nonresponse problems
2. Cognitive interviews focusing on:
 - a. The understanding of wording, tasks and response formats of the questionnaire
 - b. The understanding of metaphors and functions linked to icons and buttons in the computer program.
3. Evaluations of:
 - a. The readability of the computer screens
 - b. The usability of the programme functions
4. Acceptance tests based on standards developed in point 2 and 3.

We would like to expand the interview guide described earlier so that it covers all the four aspects mentioned in points 2 and 3 in the suggested test program above. For the cognitive part of the test, this implies that the cognitive investigation should not only focus on the textual information but also on how the respondent understand the meanings and the functions behind the icons and buttons used in the computer program. In the utility evaluation we have added observations of how the test person reads the computer screen in addition to observations of how different programme utilities work. In this way the screen presentation is defined as a part of the computer programme just like the flow chart, the quality controls and the help functions built into the programme.

A conceptual model of burdens and gratifications in Internet surveys

The results from the tests we present in this paper are interpreted in light of a conceptual model of burdens and gratifications in Internet surveys. This model describes our understanding of the relationship between the mode of data collection, response burden and response quality. The most common way to measure respondent burden has so far been to register the time it takes to complete the form. It is also customary to include the time needed to collect the relevant information, and the time used after the form is filled in. Unfortunately this measurement method is not based on an analysis of what is perceived as burdensome. Rather the available discussions of the term imply that measurement of time is inadequate as a measurement of respondent burden. As early as in 1978 Norman Bradburn suggested a definition of respondent burden, consisting of four elements:

1. Interview length
2. Required respondent effort
3. Frequency of being interviewed
4. The stress of psychologically disturbing questions which may be asked

(Bradburn 1978)

It is only the first, and in some cases the second of these factors that can be measured in minutes. Empirical studies have also shown that it generally takes more time to complete a

web version of a questionnaire compared to a paper version (i.e. Forsman and Varedian 2002). This was also what we observed in our qualitative tests. Still, even though it took more time to respond with the help of the computer, all of our test persons expressed that the web version was less burdensome than the paper version. To a certain degree this might be what the test persons felt they were expected to answer. If there is some truth to what the test persons told us, however, a time based measurement of the respondent burden would have missed it and come to the opposite conclusion.

A second objection we have against traditional measurements of response burdens is that they do not distinguish between what causes the burdens the respondent feel and the feeling itself. We consider response burden to be a subjective experience that may explain why certain modes of data collection, wordings and questionnaire designs lead to higher or lower quality of the responses. The basic model is like this:



If the personal feeling in this model is mixed up with what causes the feelings, the term "response burden" loses its power as an intermediate variable. This objection is not only a relevant argument against time-based measurements but also against Bradburn's definition.

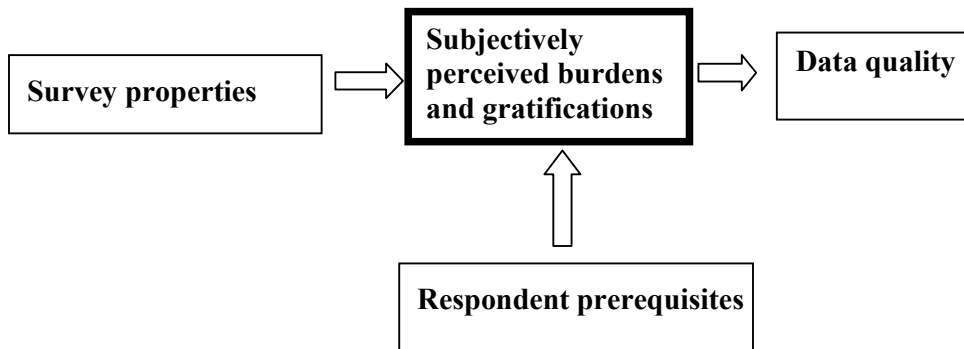
Another kind of measurement that has been suggested is to use unit or item nonresponse as an indicator of response burden (Featherston and Moy 1990). These measurements are, however, more widely used as measurements of data quality. Consequently we believe that this suggestion confounds the response burden and one possible effect of response burden. One example from the tests with web surveys also illustrates that the correlation between response burden and data quality is not at all straightforward. One of the heaviest burdens with the web questionnaire was, as far as we could observe it, that the respondent was arrested for mistakes he made in the questionnaire. Some of the error messages surprised the respondent and some of the faults they made were difficult to correct. Thus, on one hand the error messages often caused burdens. On the other hand they obviously led to higher data quality.

Hence, we find two types of confusions in the literature on the subject. One is a mix-up of the causes and the measurement of respondent burden. The other is the mix-up of the measurement of respondent burden and the measurement of response quality, often defined by partial and item nonresponse. If respondent burden is a relevant concept it should be possible to distinguish between possible causes of respondent burden, perceived burden, and response quality.

Design burdens, Respondent burdens and Interaction Burdens

In a larger model developed by Fisher and Kydoniefs at the US Bureau of Labor Statistics, respondent burden is described as an encounter between the respondent and the form he is to fill in. Under the headlines Respondent Burden, Design Burden and Interaction Burden, Fisher and Kydoniefs list 32 elements they claim affect the respondent burden (Fisher and Kydoniefs (2001)). Respondent burden is defined as the personality of the respondent, plus

behavioural and attitudinal attributes of respondents that impact on the survey completion task and that are unlikely to be moderated by the survey sponsor. The Design burdens are all the burdens that can be linked to the mode of data collection and to the content and presentation of the questionnaire used. It is mainly these components that the survey developer can moderate. Finally, interaction burdens are the result of from the interactive relationship that exists between some aspects of the respondent burden and the design burden. Hence, in Fisher and Kydonieffs model, the perceived burden is influenced both by the respondents ability to answer, by the design of the survey, and by the combination of these elements. In our model, we have distinguished between these three aspects in this way:



In this version, what we originally called "causes of response burden" is split into two kinds of causes that coincide with Fisher and Kydoieffs design and respondent burdens. We use different terms in order to emphasize that we consider these elements to be two kinds of *causes* of response burden, and not burdens in themselves.

We believe that the term response burden should be restricted to what happens in the crossroad between the survey instrument and the respondent's qualifications to respond. In our model the respondent experiences that his task is easy or burdensome. Fisher and Kydoieffs list 13 indicators of interaction burdens, which they divide into efforts, incentives and prior exposure to instrument features. The way we read this list, we think they also tend to mix up what causes response burdens and the perceived burdens that the respondent experience. Even though they use survey experts to evaluate whether the factors they list are relevant for the response burden or not, they do not refer to real-life-tests to find out what is burdensome. And factors like "Oral or written language level of the instrument", "Financial incentives to reward participation" or "Repeated requests" seem to be more like survey features than results of the interaction between the survey and the survey respondent.

As we have already stated several times, we think it is important to reserve the term "response burden" to the respondent's experience. However, the term may still be a bit misleading. Strictly speaking, the issue is not whether the total burden is high or low, but if the burdens are heavier than the advantages and other positive aspects of the survey. We have highlighted this point by using the phrase "burdens and gratifications" in the model. A common term that both contain positive and negative reactions to the survey is also *user experience*.

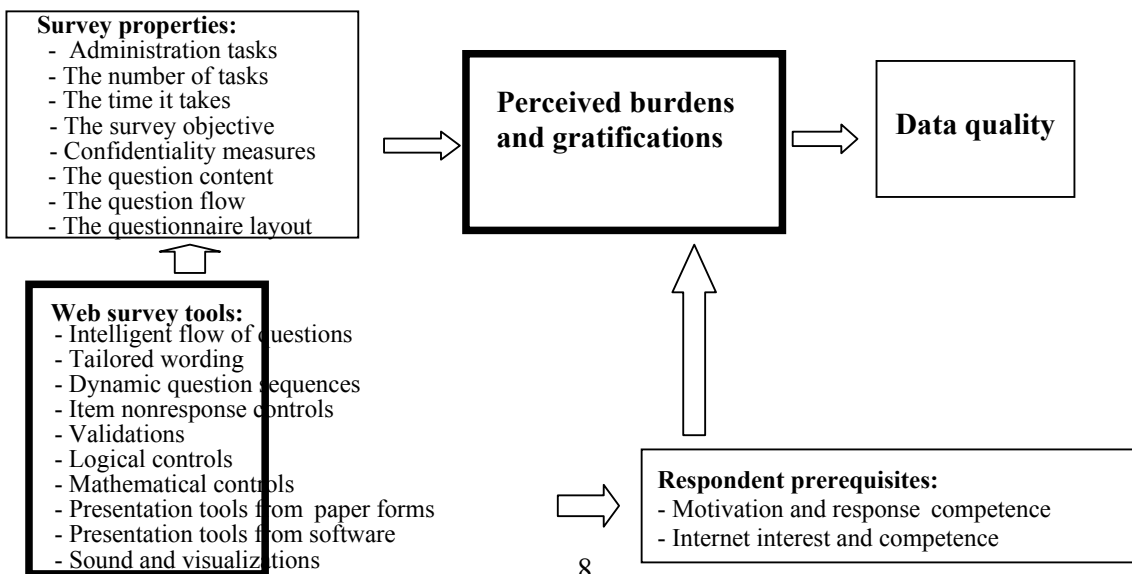
The impact of web surveys on the response burden

We doubt that the computer and Internet technology in itself will have any lasting, positive effect on the user experience of survey respondents. We have certainly observed that a web option for business surveys is welcomed by our test persons and probably has a positive effect on their attitude towards the statistical office. As web surveys become more and more common, we do think, however, that the modern and innovative image will fade and that the more concrete solutions will become more important. Hence, we think that the most important effects are mediated through the way web surveys will affect the survey designs and the respondent's qualifications.

When they specify what they mean with Respondent burden, Fisher and Kydoniefs focus on well-established general values and attitudes that are difficult to change. We would rather like to focus on the interest for the topic of the questionnaire and on the competence that the respondent feels that he has to answer the questions. Research has shown that these factors are important for what strategy the respondent choose to answer the questions. If his motivation is low or if he does not feel fully competent to answer the questions properly, he might choose a satisficing strategy instead of a step by step reasoning as we want him to (Krosnick 1991). Some of the respondents interests may be embedded in values and attitudes that are difficult to change, and their competence level may also be rather fixed. But to some extent we can also make things more interesting and make the tasks seem easier than what the respondent initially expected. Therefore these kind of personal prerequisites are more interesting than those values and attitudes that are rock solid.

We also want to make a second point about the prerequisites of the respondent. When a questionnaire is presented on the web, it is no longer only the motivation and competence to answer the questions that counts, but also the interest in using Internet for surveys and the perceived Internet competence. In this respect web surveys may add response burdens to respondents that had no problems with the earlier paper versions.

In the next version of our model we have specified two main kinds of personal prerequisites and eight survey design properties that we believe can affect the burdens and gratifications of the respondent. We have also opened the web survey toolbox and have named ten means by which the technology can affect the response burden.



The list of web survey tools can be ordered below three headlines. The first three tools mentioned are different kinds of *question controls*. The next four are different kinds of *response controls*. And the last three are three kinds of *presentation tools*. The web survey designer is working with more tools but in a smaller format than what is true for paper questionnaires. In addition to presentation principles known from paper, pc programs have introduced screen conventions and icons and action buttons which have their particular meaning. In addition the computer is not only a carrier of letters, but also of sound, pictures and video clips. On the other hand, the size of the screen a smaller the screen resolution is lower than what is true for a paper page. The space for the questionnaire is also smaller because we might need space to communicate error messages and navigational information.

The list of design properties is divided into three administrative features and five instrument features. Fisher and Kidoniefs use the same distinction, but have a slightly different list. The use of tailored wordings, dynamically constructed questions and an intelligent flow of questions can change the *content* and *structure* of the web questionnaire. The screen is different from the paper format but should hold more information than what is needed in a paper questionnaire. And some of the presentation tools are different. Consequently it is also rather obvious that the *layout* will be different. But also for the other properties listed, the conditions are different in a web survey compared with a mail questionnaire.

Administration tasks

In business surveys the administration tasks are more difficult than in personal surveys. One and the same firm might have to fill in more than one form and one and the same form may have to be completed by different employees. Sometimes people need to be able to look at the information given, but not to change any answers. Thus, different roles and levels of access have to be given to different persons. In a traditional mail survey, we have very little control with these administrative tasks. When a paper questionnaire is transferred to a web platform, however, it is not only the questionnaire that is presented in a different manner. As we described earlier in this paper, the administrative tasks are controlled with the help of a working list where those who are authorized can pick the questionnaire they shall answer. In the tests we will refer from, this was done from what we call the system's homepage. This means that an important part of the response process that previously has not been standardized and controlled by the survey sponsor, now is standardized in a similar manner and controlled by the same kind of tools as is used in the ordinary questionnaires. The same is true for other administration tasks like signing and mailing the questionnaire. Generally one may say that more of the activities that are carried out in surveys are given a questionnaire form.

The number of tasks

In interviewer-mediated surveys, the interviewer first finds the respondent, contacts him about when the interview can take place, ask the questions and take down the answers. After the interview is done, she sends the completed form to the survey organization. When we send a questionnaire to a potential respondent by mail, we also transfer most of the tasks of the interviewer to him. The respondent must...

- Find time to answer the form by himself,
- Read the questions and take down the answers, and
- Put the form in an envelope and send it to the survey organization.

Hence, self-administered questionnaires may be a larger burden to the respondent than interviewer mediated surveys, simply because the respondent is given the tasks of the

interviewer. In good computerized self-administered forms, however, some of these tasks are taken off the shoulders of the respondent and given to "a virtual interviewer" which is the computer program. Among other things, the computer program may simplify the reading of the form by guiding the respondent directly to the relevant questions, and by tailoring the wording so that it becomes consistent with the respondent's situation than in standardized paper forms.

The time it takes

The response burden of the time it takes to respond to a survey may often be very different from the number of minutes it takes to fill in the answers. Even a short moment of time may seem like a lot for a busy respondent. This may be the case because the form has been sent to a particularly busy person, or because the respondent is given a short deadline. Tasks associated with joy are also generally easier to find time for, than tasks that the respondents find boring. Whether or not a task is considered to be time consuming or not will therefore always be subject to the respondent's own judgement. Because a web based data collection system includes questionnaires that govern more activities than just filling in the questionnaire, it may give the respondent a more realistic picture of what activities, which are involved and how long they take. Next this may lead to a different judgement of the burdens than before.

The survey objective

Commercial surveys have lower response rates than surveys run by statistical offices and research institutions. Some topics are considered more sensitive than others. Some objectives are also more motivating than others. If you need a bank loan in order to build a new production hall, it is easy to understand that you have to fill in an application form. In this case it probably also make sense to fill in a form that describes the project to the local authorities before they will permit you to start the work. But if you only wanted to make some exterior modification of an existing building, this task may seem unnecessary. Why you later should describe the changes you made in a statistical questionnaire is probably even more difficult to understand. The last of these three questionnaires described here will probably be the easiest to fill in, while the first probably will be the most detailed and complicated. On the other hand, the first task may seem to be the most important and meaningful, while the last may be considered as a waste of time. The result may well be that easiest questionnaire is considered to be the most burdensome. In the example with the statistical questionnaire the cause of the response burden may not be the questionnaire, but the motivation for completing it. Therefore we should rather look for gratifying rewards than for a more user-friendly form. In the tests that we report from in this paper, we experimented with feedback products in statistical surveys that were meant to have a motivating effect on the respondents.

Access and authentication

In traditional business mail surveys the access and authentication controls are rather weak. Business surveys may be addressed to specific persons in the firm, but are more often only addressed to the company with a request that people with certain qualifications respond to the questions. In some questionnaires the respondent needs to verify the information given by signing the form and in others he may have to identify what kind of position he holds. But again, in most cases the respondents could be anybody in the firm.

There is much more focus on security issues in web surveys. This obviously has to do with the environment in which the respondent finds himself both when he fills in the questionnaire

and when he posts it to the addressee. A traditional mail questionnaire is filled in inside the walls of the company and mailed in a closed envelope. On the Internet the company is outside these walls and transmits information in a different mail system. The walls and the protection of mail has to be re-established with the help of id-numbers, passwords, digital signatures and encryption. This is an extra task in web surveys that may be quite burdensome to complete. The identification number and password are generally posted in separate mail or as sms-messages that the respondent has to have present as he is entering the web system. The entrance and mailing procedures also often consist of several steps with long combinations of letters and numbers that have to be entered correctly (Couper 2002). And even if everything seems to work properly, it may still leave the user with a feeling of insecurity.

The security system can be seen as one administrative task that is given a questionnaire form in a web based survey system, in a similar way as other administrative tasks we have mentioned previously. In this case, however, we think it is more likely that the computerized system will be experienced as an extra burden rather than a more user-friendly solution.

Observations made in tests of Internet Business Surveys

The qualitative tests of the questionnaire of the Monthly Retail Trade Index questionnaire and the questionnaire used to collect Quarterly Investment Statistics were both carried out among test persons who had previous knowledge of the questionnaires and experience in using computers and among people who did not have such prior knowledge or experience. The test persons also worked in companies of different sizes. On the other hand, only a few tests were carried out. Hence, we have no reasons to claim that the observations made can be generalized. When qualitative evaluation and testing methods are used to develop better questionnaires, an iterative strategy with a few tests followed by revisions, followed by new rounds with tests and revisions are recommended. In this paper one may say that we apply the same strategy in a theory building process. The results we report here can be considered as results from the first round with the first draft of our conceptual model.

Common denominators of the amateur and experienced web user

The first two tests with the Monthly Retail Trade Index questionnaire were carried out with two very different test persons. The first one was the owner and head of a firm letting out buildings, but which was also involved in other kinds of businesses. He had already reported his retail trade many times on paper. In addition he was an experienced user of the Internet. He had even taken part in the development of his firm's own homepage. The second test person ran a small leather shop together with his wife. He neither had any prior respondent experience nor any computer experience. He did not even have a computer yet and consequently had to use a laptop which we brought to the test.

In the first part of the think-aloud session, test person 1 jumped directly to the response boxes of the paper questionnaire. He had already read the introduction before, he told us. Therefore he did not need to read it again. This respondent would, not have noticed any new information or definitions of terms written in the introduction letter or in the enclosed instructions.

When he started the web version this test person first looked where the other links on the introduction page would lead him before he logged himself onto the system's homepage.

From here he was able to move directly to the company fact sheet instead of the questionnaire that he was expected to fill in. He did not use the action button presented on the fact sheet, but the back button on the web browser menu to move back to the homepage. Because of this, his corrections were not recorded. The programmers had expected that the users should select a questionnaire from the homepage. To ensure that company background information was updated before the respondents answered any questions, the company fact sheet always came up as the first screen before the selected questionnaire. Consequently the test person was brought to this page for the second time, only to learn that none of his corrections from the first time were recognized.

From the receipt page there were two options. There was one link back to the homepage and one link to the statistical office's homepage. Contrary to what we wanted, our test person chose the last option. From here he linked himself to other web services. Had we not directed him back to the survey system, he would probably never have returned to the working list at the system's homepage as we wanted him to.

Test person 2, the novice and amateur, read thoroughly through the introduction letter and the instructions given on the first screen of the web survey. Then his comment was "This was interesting, but it says nothing about what I am to do next". We think this was the best comment given during all the tests.

Even if these two test persons seemed to react very differently to the questionnaire, their navigational problems were basically the same. There were different ways one could move from one screen to another, but no instructions on the screen indicating what we wanted the respondents to do next and how they should do it. The experienced web user chose the first hyperlink he could find which led him out of the questionnaire system. The amateur had no ideas and consequently did not do anything. Both needed to have some plain navigation instructions brought to their attention.

Another observation illustrates how important it is to present messages on the part of the screen where the respondent is focused. Error messages were initiated as the respondent clicked on the next button at the bottom right corner of the screen. The error message was, however, presented at the top of the present screen. Because the screen did not change and the respondents were focused at the bottom right of the screen, none of them recognized the error messages. They rather wondered if the program had a bug because it did not react when they clicked the next-button. In this case the program would not move forward unless the errors were corrected. Thus, as long as the respondent did not see the error messages, he was stuck with a program that apparently had stopped working.

These observations led us to believe that the activity order and focus of the respondent should be strictly governed in web surveys, while the respondent generally should be able to overrun error messages.

Reducing the response burden with the help of more questionnaires

Even if all the tests showed that it took more time for the respondent to complete the web version than the paper version of the same questionnaire, all the test persons claimed that they would prefer to use the web version in the future. One of them explained why like this: "The worst part of the survey is to remember to post the questionnaire before the deadline. With a

web system I can complete the whole task in one operation". We believe that this test person made an important point. The heaviest response burden is often not linked to the questionnaire itself but to the preparations and supplementary work that has to be done after the questionnaire has been completed. In a web based data collection system most of these activities are brought together and are governed by a kind of questionnaire. It strikes us, however, that the layout of the questionnaires that govern administrative tasks are quite different from each other and quite different from the statistical questionnaires. We believe that the system would be even easier to use if both the administrative and statistical questionnaires were designed according to the same principles. Today, the statistical questionnaires used in web systems are often badly designed, while the questionnaires that govern other activities might not even be recognized as questionnaires at all.

The need for traditional design and explanations of modern terms

The main source of confusion with the questions in the two test questionnaires seemed to stem from design weaknesses in the original paper version. The Monthly Retail Sale questionnaire can serve as an example. In this questionnaire the respondent is asked to report three figures; the net worth of sale excluding VAT, the VAT and the gross worth of sale including VAT. The reasons why the statistical bureau want all three figures is that the VAT rate is different for different kinds of goods. Hence, in some companies the gross worth cannot simply be calculated from the net worth of the goods sold. However, only a minority of the companies operate with different rates for different goods. For the majority of companies three questions appear to be two questions too much. This feeling was even more prominent in the web version. All test persons were surprised and disappointed with a computerized questionnaire that apparently was unable to calculate and multiply one figure with a VAT percentage. The obvious solution in a web survey is of course to split the original sequence into sequences that are governed by the composition of goods and corresponding VAT-rates that exist in the companies.

The different pieces of information were also presented in an awkward order in the Monthly Retail Sale questionnaire. The response box was presented above an instruction of what should be reported. And it is only in the end of this instruction that the respondent is told that the figures should be given in 1000 kroner. Neither is there any common text convention that indicates what are questions, what are definitions and what are response areas.

These kinds of burdensome weaknesses are not introduced by the computerization but are inherited from the paper version of the questionnaire. They point at an often-neglected point in web survey projects, namely that a badly designed questionnaires not only needs to be fitted to a computer version, but also needs to be revised according to traditional standards of user-friendly designs.

We observed the some problems with the terms used in questions and response boxes. The respondents seemed to be quite familiar with the economic terms use in the questionnaires. Only when the unit or time span they were supposed to report from differed from common practices, it sometimes caused confusion. The most frequent conceptual problems, however, were with the terms used to name the different screens in the data collection system or to indicate what effects the different action buttons had. These terms were either invented by the

producers of the questionnaire program or imported from other computer programs. Here are some examples of terms that caused problems:

In the first tests, the screen that we have called Company Fact Sheet in this paper was named Company Profile. The respondents' understanding of this term was something more like the company's business strategy, while the screen asked for addresses and industry codes. Next it was not clear to all that they were supposed to update some of the information on this screen. Also some of the data fields were locked. Respondents who tried to change any of these fields were even more confused. The action buttons on the bottom of this screen were named "Save" and "Accept". Those who did not change anything, but chose "Accept" were not confident that the information was saved. And those who made changes often felt that "Accept" (the changes) was just as an appropriate term as "Save" (the changes). Generally the respondents were also uncertain about what happened when they saved information. Was it saved on the hard disc of the respondent or was it sent to the server, which ran the system?

Sometimes the program produced messages that were meant to explain what the program was doing when nothing actually happened on the screen. This could be a message such as "Report is being generated", "Data is being saved" and so on. In some instances, however, we observed that messages like these had the opposite effect of what was intended. They made the respondent expect that the program was about to produce something and consequently the test persons were confused when nothing came up.

Motivating feedback

It is well known that incentives might improve the response rates in surveys, and that this is also true in web surveys (Frick, Bächtiger and Reips 1999). However, this is not an option in statistical data collection from companies. In these surveys the incentive is rather a negative than a positive one; if you do not respond you will be fined. This strategy surely improves the response rates, but at the cost of a higher response burden we believe. In the web projects we are reporting from here, we also made some experiments with feedbacks from the questionnaire program that were intended to make the information task more meaningful. From both questionnaires we generated a report where the respondent could compare his information with national statistics for the same kind of industry. In this way he could see if his sales volume or investments were higher or lower than what was the situation in the industry as a whole. In this way we also tried to communicate that statistics can be useful.

Respondents in small firms were particularly pleased with this statistical product. In bigger firms, however, they often have their own marketing departments, which have more appropriate statistics than the standardized reports we can produce in a general data collection system. In none of the bigger firms tested, the test persons worked in that department. Consequently, even if they were sympathetic to the idea, the statistics were of no great interest to them.

There was a different kind of feedback, however, that we did not think about, but which was asked for and which is easy to produce. That is a summary of the information given that could serve as documentation for the respondent. The receipt screen

used in this project only stated that the data had been received and gave a reference number in case of future disputes about this. There was no print button on this screen. Yet some of the test persons used the print option on the browser menu to print the receipt. They also found the reference number rather meaningless and would rather prefer a fact sheet containing the main information that they have just given.

Implications for web questionnaire developers

We believe that these observations send important messages to the developers of web-based data collection systems. One important message is that transferring a questionnaire from paper to the Internet is not only a challenge for programmers, but also for the survey sponsors. The reason is that most questions in business questionnaires probably have to undergo revisions so that they conform to traditional standards of user friendliness before they can be implemented in a web system. Secondly we think it is of uttermost importance to recognize that a web-based system for data collection consists of a set of different questionnaires, which govern different tasks linked to the informant role of the respondent. Our main impression is that the user friendliness of the questionnaires before and after the statistical information is given is the most decisive factors for the response burden of the system. One contribution to a more user-friendly web survey system might be that the administrative questionnaires are designed according to the same principles and are given a similar layout as the actual survey questionnaires. Finally we think the tests have demonstrated that a web system introduces both new problems and new possibilities in business surveys. Traditionally business survey sponsors have assumed that the respondents are familiar with the terms used in the questionnaires. With a web system, new and not so familiar terms are introduced. And there is an obvious need for cognitive studies of how these terms are understood. As an example of the new possibilities, we believe that the experiments we made with a statistical product as incentives are a promising path to follow in the efforts to reduce the response burden of the business respondents.

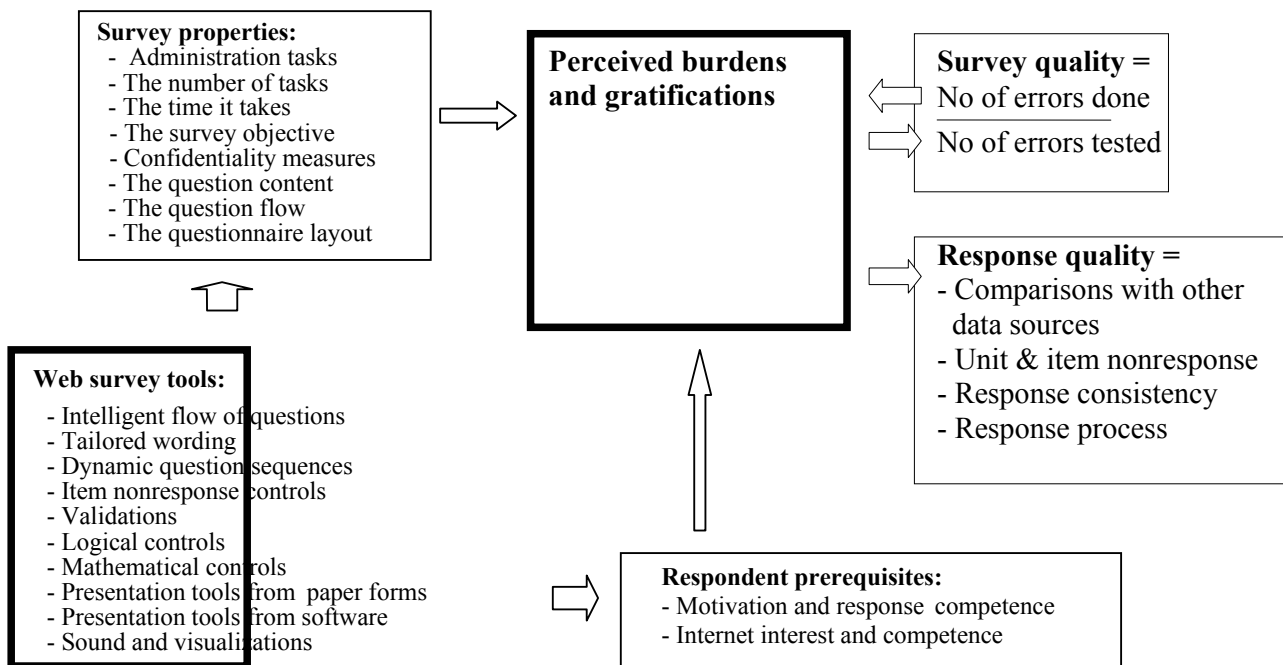
Testing the recommendations from qualitative evaluations

We started this presentation with a simple conceptual model, which distinguished between the causes for response burden, the response burden itself and the effects of response burden. Later we expanded the understanding of what is understood by response burden and specified different causes and different ways in which the Internet technology may influence burdens and gratifications of business surveys. In the last version of the conceptual model, which is shown below, we have also opened the quality box. This is done in order to add a point about how the Internet technology can be used to evaluate the recommendations made on the basis of qualitative development methods.

Traditionally response quality is either measured by comparing the information given with other data sources, by recording unit and item nonresponse or by performing consistency checks of the response pattern. In our model we have added response process indicators. In web surveys it is possible to record the actions performed in addition to the answers given. This kind of information has been called paradata or process data (Heerwegh 2002). When a questionnaire is tested with the help of focus groups, cognitive interviewing or other qualitative methods, we do not only want our test persons to understand and answer the questions correctly. We also test if they proceed through the questionnaire in the way we want

them to. At some point in the testing, we conclude that the questionnaire works according to the standards set. The process data collected in the field may support or contradict that the respondents follow the prescribed path through the system during the actual data collection.

We have also added an indicator of survey quality in our model. Here survey quality is measured as the relation between the number of errors tested for in the system and the number of error controls that were activated. If the questions and instructions are correctly understood, the number of errors and error messages should be low. Consequently the relation between the number of possible and activated error messages could be seen as an indicator of how well the survey questions worked. If the recommendations made on the basis of qualitative development methods were valid, the survey quality indicator should be low and the process data should form the same response pattern as the one observed in the laboratory. In this way these two indicators might be used as an evaluation of the recommendations made during the qualitative development of the web survey system.



References

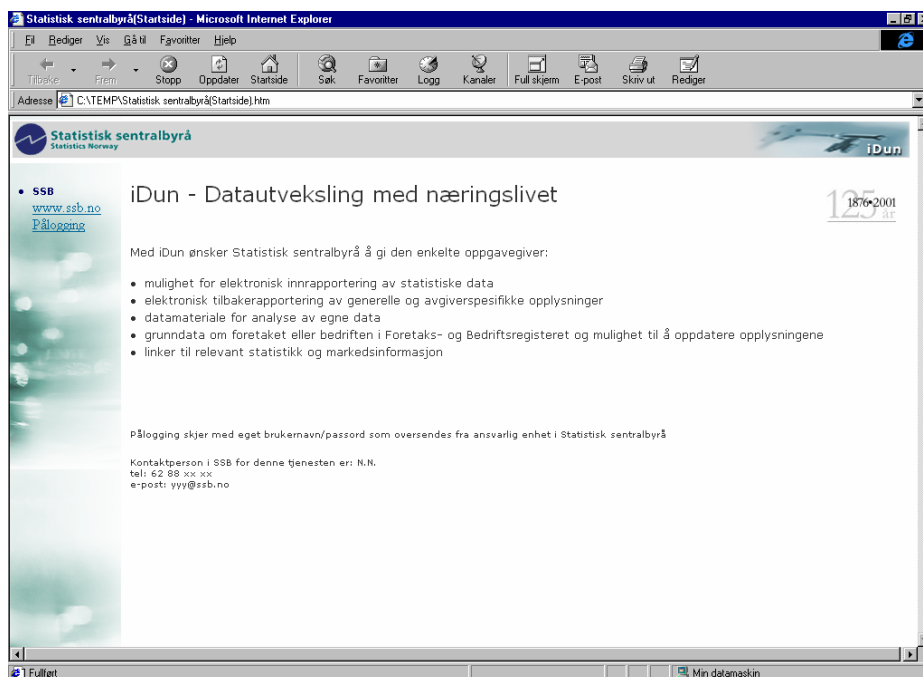
- Bradburn (1978). "Respondent Burden," *Health Survey Research Methods*, DHEW Publication No. (PHS) 79-3207, pp. 49-53.
- Couper, M. (2002): *Web Survey Design*. Survey Research Center, University of Michigan and Joint Program in Survey Methodology, Stockholm September 24 - 25. pp 11
- Couper, M. and Nicholls II, W. L. (1998): *The History and Development of Computer Assisted Survey Information Collection Methods*. In Couper, M et al (eds): *Computer Assisted Survey Information Collection*. Wiley and Sons Inc., NY
- Cox, B. G. and Chinnappa, B.N. (1995): *Unique Features of Business Surveys*. In Cox, B. G. et al (eds) *Business Survey Methods*. Wiley and Sons Inc., NY
- Featherston and Moy (1990). "Item nonresponse in mail surveys", paper presented at the International Conference on Measurement Errors in Surveys.
- Fisher and Kydoniefs (2001). "Using a theoretical model of response burden to identify sources of burden in surveys", paper presented at the 12th International Workshop on Household Survey Nonresponse. Oslo
- Forsman, G. and Varedian, M. (2002). *Mail and Web Surveys: A Cost and Response Rate Comparison in a Study on Students Housing Conditions*. Paper presented at the International Conference on Improving Surveys (ICIS) in Copenhagen August 26- 28.
- Frick, A., Bächtiger, M.T., and Reips, U.-D. (1999): "Financial incentives, personal information and drop-out rate in online studies". In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, & A. Werner (Eds./Hrsg.), *Current Internet science - trends, techniques, results. Aktuelle Online Forschung - Trends, Techniken, Ergebnisse*. Online Press; Zurich.. Available at URL: <http://dgof.de/tband99/>.
- Haraldsen, G. Dale, T., Dalheim E. and Strømme , H (2002): *Mode Effects in a Mail plus Internet Designed Census*. Paper presented at the International Conference on Improving Surveys 2002. Copenhagen August 26- 28.
- Heerwegh, D. (2002). *Describing response behavior in websurveys using client side paradata*. Invited position paper presented at the International Workshop held by ZUMA in Mannheim, October 17-19, 2002.
- Krosnick, J.A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213-236
- Norwegian Ministry of Trade and Industry (2001): *Simplifications and Adoptions for the Businesses. Plan of Action, part I*. Oslo (in Norwegian only).
- Nielsen, J. (1993): *Usability engineering*. Academic Press, NY

Nygård, T (2001): *The questionnaire that was able to listen*. Report about a pilot questionnaire written for the Norwegian Ministry of Trade and Industry, Oslo (in Norwegian only)

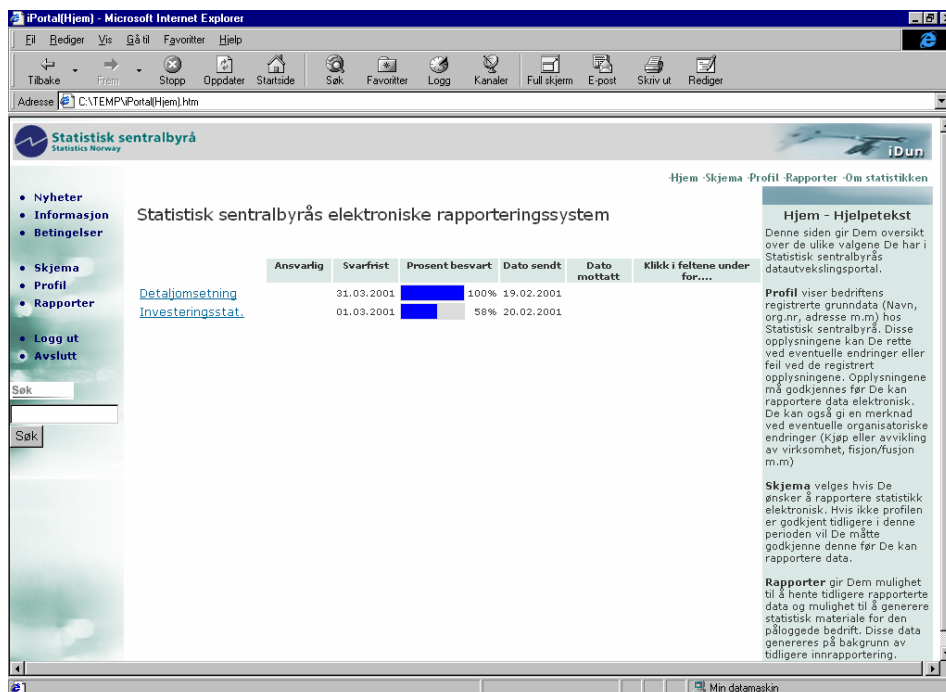
Rubin, J. (1994): *Handbook of usability testing: How to plan, design and conduct effective test*. Wiley and Sons Inc., NY

Appendix 1 Screen downloads from the web system tested with questionnaires for the Monthly Retail Trade Index and the Quarterly Investment Statistics.

The log-in screen



The system's homepage



The company fact sheet

Statistisk sentralbyrå
Statistics Norway

Hjem Skjema Profil Rapporter Om statistikken

Profil - Avgiverinformasjon

Navn: Avdeling: Gyldig dato:

Epost:

Telefon:

Faks:

Orgnr bedrift:

Orgnr foretak:

Postadresse: Postnr: Poststed:

Eielighetsadr: Postnr: Poststed: Komm.:

Næringskode: Trykking ellers:

Merknad:

Kontaktperson: Epost: Telefon:

Profil - Hjelpetekst

Denne siden gir Dem mulighet til å kontrollere bedriftens grunddata og eventuelt rette feilaktige opplysninger. For å kunne fylle ut skjema må profilen godkjennes. Videre må kontaktperson og e-postadresse være registrert før De kan fylle ut skjema. Enhetsregisterloven pålegger SSB å videresende slike endringer til Enhetsregisteret i Brønnøysund. Hvis det er noen uklarheter vil SSB kontakte Dem.

Ved endringer av bedriftens grunddata legges gyldighetsdato inn i tilhørende datofelt. Hvis denne settes blank vil dagens dato bli gjeldende.

Har det vært organisatoriske endringer (kjøp eller avvikling av virksomhet, fusjon/fusjon m.m) kan De rapportere dette i merknadsfeltet. Gi da også opplysninger over organisasjonsnummeret til involverte enheter.

Questionnaire page

Statistisk sentralbyrå
Statistics Norway

Hjem Skjema Profil Rapporter Om statistikken

Navn: web Avdeling: LUTTESTING Org.nr: 900000020

Data registrert: 2001.02.20 10:20:04 Versjon: 44 Lagret som: Foreløpige tall

Hvor stor var omsetningen i Desember 2000 **1000 kr**

1. Eksklusiv merverdiavgift	<input type="text" value="100"/>	?
2. Merverdiavgift	<input type="text" value="23"/>	?
3. Inklusiv merverdiavgift	<input type="text" value="123"/>	?

I spørsmålet ovenfor skal du fylle ut omsetningen for den aktuelle måned. Oppgaven skal gjelde kalendermåneden. Dersom annen registreringsperiode benyttes spesifiser i merknads feltet. Alle tall skal oppgis i tusen kroner. Svarfrist er den 12 i inneværende måned.

Merknad:

Detaljomsetningsindeks

I spørsmålet skal du fylle ut omsetning for den aktuelle måned. Med omsetning mener vi salgsinntekter av varer og tjenester, avgiftspliktige, avgiftsfritt og utenfor avgiftsområdet. Omsetning skal inkludere provisjonsinntekter, leieinntekter, royalties, lisensinntekter o.l.

Ta ikke med finansinntekter

Dersom regnskapstall ikke foreligger på det tidspunkt oppgaven skal gis for, ber vi om anslag på grunnlag av kassapoppler eller andre interne oppgaver.

The receipt page

Portal(Skjema) - Microsoft Internet Explorer

Adresse: <http://obelk.8090/seriet/Page>

Statistisk sentralbyrå
Statistiska Norge

Navn: web Avdeling: UTTESTING Org.nr: 900000020

Hjem · Skjema · Profil · Rapporter · Om statistikken

- Nyheter
- Skjema
 [Detaljomsætning](#)
- Rapporter
- Logg ut
- SSB-kontakt

iDun -Kvittering for skjema

Denne siden er ment å gi informasjon om de registrerte data. Eventuelle kontrollmeldinger er listet ut. De kan være nyttige å ta vare på ved eventuell kontakt fra Statistisk sentralbyrå. De får også et kvitteringsnummer for at data er registrert.

Tabellen med data viser de tall som ble registrert ved innrapporteringen.

Hervil etterhvert en tabell over registrerte data vises

RELATERTE LINKER

Vennligst trykk her for å gå [tilbake til hovedsiden \(Hjem\)](#).

Detaljomsætningsindeks

I spørsmålet skal du fylle ut omsetning for den aktuelle måned. Med omsetning mener vi salgsinntekter av varer og tjenester, avgiftspliktige, avgiftsfritt og utenfor avgiftsområdet. Omsetning skal inkludere provisjonsinntekter, leieinntekter, royalties, lisensinntekter o.l.

Ta ikke med finansinntekter

Dersom regnskapstall ikke foreligger på det tidspunkt oppgaven skal gis for, ber vi om ønske på grunnlag av kassapoppler eller andre interne oppgaver.

[Om statistikken](#)

Søk

Søk

Internettsone