Journal of Official Statistics, Vol. 20, No. 2, 2004, pp. 393–410

Identifying and Reducing Response Burdens in Internet Business Surveys

Gustav Haraldsen[1]

Response burden is of great concern in survey data collection activities. Despite extensive efforts on the part of data collection agencies to reduce the burden the problem seems to prevail. In this article we discuss how web surveys can be one way of reducing the burden in business surveys. We offer a perspective on response burden that is different from the one currently used and we include this conceptualisation into a model that links survey design and response quality to response burden. We describe some qualitative tests performed in Norwegian business surveys, which support the usefulness of our conceptual model.

Key words: Data quality; respondent characteristics; visual design; CASIC; business survey reponse process.

1. Introduction

Questionnaires that are presented with the help of computer programs have been with us for a while already. But it was not until questionnaires could be distributed on the Internet and read by ordinary browsers that a cheap and flexible platform for self-administered computer assisted surveys was established. While most individual and household surveys have traditionally been performed with the help of interviewers, self-administered surveys have been the dominating procedure in business surveys. At Statistics Norway interviews were computerized about ten years ago. This transition changed the design of questionnaires used in social statistics. Today the web technology is mainly used to change the questionnaires and data collection procedures in business surveys. Statistics Norway has announced that from July 1 2004 all business respondents should have a choice between a paper and an Internet version of the questionnaires we ask them to complete.

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 respondent burden can be eased. However, none of these effects are realised by the technology itself. In a status report made in the monograph "Computer Assisted Survey Information Collection" in 1998, Mick Couper and William L. Nicholls II stated that:

- Even if the data collection period has been shortened, the planning period tends to increase.
- Cost reductions are only gained when the volume is high or as an effect of reorganisations in the survey institution.

[1] Statistics Norway, PO Box 8131 DEP, N-0033 Oslo, Norway. Email: gha@ssb.no
Acknowledgment: Tests and discussions in the project "Developing Methods for Assessing the Perceived Response Burden" have been an important source for refining the original thinking. I owe thanks to Swedish and British colleagues who took part in this project.

- Some quality improvements are obvious; others only come after a total change of how the questionnaires work and how they are presented.
- Some of the tasks traditionally performed by the data collector have been transferred to the respondent. Normally this leads to a lighter burden for the data collector and a heavier respondent burden (Couper and Nicholls 1998).

In this article we will first of all focus on the last of these conclusions. More specifically we will discuss how web surveys can reduce the response burden in business surveys. Our main contribution to this discussion will be to offer a different perception of response burden than the one used today, and to include this new concept into a model, which links the survey design, the response burden, and the response quality. Then we will refer to some observations that we have made in qualitative tests in business surveys, and which we believe support the usefulness of our conceptual model.

2. From a Macro-economic Calculation to a Personal Perception of Burden and Gratifications

The business organizations have for a long time complained about the burden mandatory governmental questionnaires enforce on them. Even though the statistical questionnaires only represent a small fraction of all the forms that the companies have to fill in, the statistical offices seem to be the favourite target of this criticism. In 2003 business surveys from Statistics Norway only represented 1.4 per cent of the total time businesses spent on governmental questionnaires. One of the reasons for this low figure, however, is that the business surveys are based on samples, while most of the other governmental questionnaires are to be answered by all companies, or by all companies within a certain branch. Hence, even if the total burden is low, the burden felt by individual respondents may still be significant.

The calculation above is based on figures from the Register of Reporting Obligations of Enterprises in Brønnøysund. All governmental institutions which companies oblige themselves to report to, report to the register what kind of information they are asking for and how much time they estimate that the information gathering and reporting will take. The register keeps track of the reporting obligations that the companies have; the main idea of this being to prevent double reporting and other unnecessary information collection. The results from the most recent years indicate that improved co-ordination of data collection activities and questionnaire simplifications have had a positive effect on the response burden, but that this effect is now slowing down. Figure 1 shows the trend that these conclusions are based on.[2]

In his comments on these figures the Minister of Trade and Industry points to the transition from paper to Internet based data collection designs that he expects "will lead to noticeably positive results for businesses in 2004."[3] The initiative he is mainly referring to in this comment is AltInn,[4] which is a common web portal and web application for industry reporting to governmental institutions. The system was launched this spring and

[2] http://www.brreg.no/english/coordination/
[3] http://www.dep.no/nhd/norsk/aktuelt/pressem/024081-070044/index-dok000-b-n-a.html
[4] AltInn = Alternativ Innrapporteringskanal = Alternative Data Submission Channel

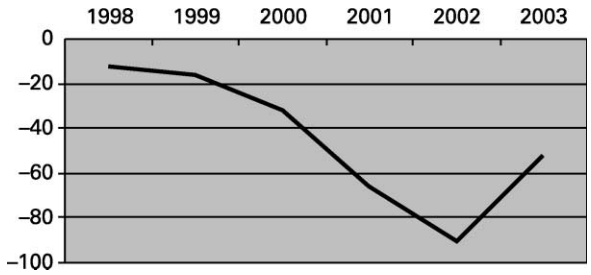

Fig. 1. The reduction of business response burden because of data collection co-ordination and simplification. Man-labour years 1998–2003

Statistics Norway has taken part in the development of this web service (Vik 2004). In addition we have developed our own web business survey system, called IDUN[5] (Sæbø, Gløersen, and Sve 2002). Both these systems are used in order to fulfil the ambition that all statistical business respondents should have a web alternative by July 1 2004.

There are at least two problems regarding the expectation that the transition introduced by these systems will lead to an accelerating reduction of the business response burden. The first one is that we do not yet have any data that supports the notion that the web technology automatically reduces the time it takes to complete the questionnaires. A small empirical comparison made in Sweden suggests rather that it takes more time to complete the web version of a questionnaire than to fill in the paper version (Forsman and Varedian 2002). Observations from the qualitative tests that we will refer to later in this article, point in the same direction. One reason for this is that the paper questionnaires that the web versions are based on are of rather a low quality. This is an observation also made in many other countries (Cox and Chinnappa 1995; Willimack et al. 2002). Hence there is a need for a methodological evaluation and improvement of the paper questionnaires before they are transferred to the Internet. The Office for National Statistics in the UK has initiated such an evaluation (Jones 2003).

The other aspect of web surveys that affects the response burden is of course how well the web application is designed and how it works. When looking into this, it is important to remember that it is not only the questionnaire that is transferred from paper to Internet; the administration of each questionnaire and the internal administration of different questionnaires in the companies will also be based on Internet technology. One of the main problems revealed in the qualitative tests carried out in the AltInn project, for instance, was that respondents were unable to log on to the system and find the test questionnaire. Business respondents are often familiar with the questionnaire from earlier data collection but the web based administrative system is new to them.

One feature that is almost always built into web questionnaires is quality controls that stop the respondent if he/she skips questions or gives illegitimate or illogical answers. The advantage of this is of course that incorrect answers can be corrected on the spot. The disadvantage may be, however, that it adds one more task to the job that the

[5] IDUN = Informasjon og Datautveksling med Næringslivet = Information and Data Exchange with Businesses

respondent has to do. Built-in error checks constitute in fact an outsourcing of revision routines that were formerly carried out by the data collection institution.

In addition to these problems with web questionnaire and survey design, there is a problem with the way we have previously measured the response burden. Seen from a business perspective it seems logical to calculate the time it takes to complete the questionnaires. Time used on questionnaires is an expense that seldom generates income for the company. All cost reductions will therefore be an advantage. From a statistical point of view, however, time reduction is not necessarily an acceptable criterion of success. Even though statistical institutions want to be on good terms with their respondents, their main objective is to collect high-quality data. Therefore a reduction in the response burden is only acceptable if the response quality remains stable or, preferably, improves. An example of a unacceptable time reduction is that the respondents just scan through the questionnaire instead of reading the questions thoroughly. If, on the other hand, different response controls are initiated during the completion of the questionnaire, the response quality will certainly increase, but so will the time spent. While the focus on time and the cost of time is understandable from a business perspective, we believe the quality considerations should be the first priority for statistical institutions. This calls for a response burden concept that focuses more specifically on the balance between response burden and response quality during the response process.

3. A Conceptual Model of Survey Burden and Gratifications

As early as in 1978 Norman Bradburn suggested a definition of response 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)

Only the first, and in some cases the second of these factors can be measured in minutes. In fact time might not be the most obvious measuring scale for any of them. The number of questions is probably a better measure of the respondent burden than the length of the interview or the questionnaire. How often one is contacted should obviously be measured in numbers. Personal effort and stress is generally not measured in a quantitative way at all.

An important point made by Bradburn is that it is necessary to carefully focus on the subjective perceptions of the respondent. Time-consuming questionnaires are demanding, but how demanding they are is not only decided by the time it takes to complete them, but also by how the respondent perceives the time spent on the questionnaire. If the questions are absorbing, time flies fast. If the questionnaire is boring, if it takes longer to complete than expected or if there is a tight deadline, the task probably feels more burdensome. An observation that we made in one of our first qualitative tests of web surveys illustrates this point well. In this test the respondent both filled in the paper version and a web version of the same questionnaire. It was obvious that the test person spent more time completing

the web version than the paper questionnaire. But when we asked him, he strongly argued that the web version was less burdensome to fill in than the paper version. The main reason for this was probably that he thought it was more fun to fill in the questionnaire on the computer. Another reason given by the test person was that the most burdensome part of filling in a paper questionnaire was to remember to mail it. When he used his computer he could send it by e-mail as soon as it was completed.

The main objection to Bradburn's definition of response burden is that it does not distinguish between the causes of burdens and the burdensome feeling itself. The length of the questionnaire, the effort needed to answer the question and the number of times one is asked may not be perceived as a heavy burden as long as the task is interesting, but may otherwise lead to stress. Thus, to our minds, the first three elements of Bradburn's definition are possible causes of burdens, while the last one is an aspect of a burdensome feeling.

Another response burden indicator that has been suggested is unit or item nonresponse (Featherston and Moy 1990). These measurements, however, are more widely used as measurements of data quality. Consequently this indicator confounds the response burden and one possible effect of response burden. Hence, we find two types of confusions in the literature on the subject. One is a mix-up of the causes and the effects these causes have on the perceived response burden. The other is a mix-up of the measurement of response burden and the measurement of response quality. We envisage a basic model where the subjectively felt response burden is an intermediate variable, which may explain why certain modes of data collection, wordings and questionnaire designs lead to higher or lower response quality (see Figure 2). If these three elements are kept apart, the model is testable. If the elements are mixed up, the results will be tautological.

In their response burden model, Fisher and Kydoniefs list 32 elements, which they claim affect the burden of a questionnaire. The elements are grouped under the headings "Respondent Burden," "Design Burden," and "Interaction Burden" (Fisher and Kydoniefs 2001). Respondent burden is defined as personality traits and behavioural or attitudinal attributes of the respondents that affect their ability to complete the questionnaire quite independently of how the survey is designed. 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. These are the components in the survey developer's toolbox. Finally, interaction burdens are products of what happens when respondents with certain characteristics are confronted with a survey that has certain properties. Hence, in Fisher and Kydoniefs' 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: (see Figure 3).

In this version, what we originally called "causes of respondent burden" are divided into two types of causes that coincide with Fisher and Kydoniefs' design and respondent burden. We use different terms in order to emphasize that we consider these elements to be

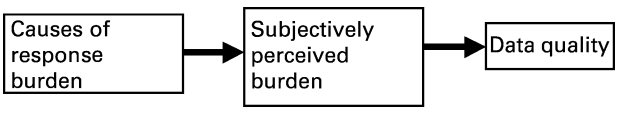

Fig. 2. The relationship between causes of response burden, the subjectively felt burden and possible effects of this feeling

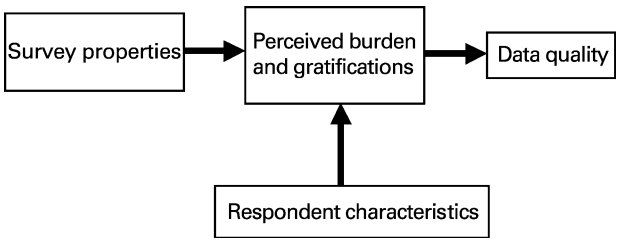

Fig. 3. Response burden and gratifications defined as the result of the encounter between a survey design and a respondent

two types of *causes* of respondent burden, and not the response burden itself. To our minds that term should be restricted to what happens at the interface between the survey instrument and the respondent's ability to respond. Fisher and Kydoniefs list 13 indicators of interaction burden. They divide these into efforts, incentives and prior exposure to instrument features. Here too, we think they tend to mix up what causes response burden and the burden that the respondents experience. 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 confine 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 whether the burden takes presidence over the advantages and other positive aspects of the survey. We have highlighted this point by using the phrase "burden and gratifications" in the model. A term covering both positive and negative reactions to the survey is also *user experience*, which may be defined as "the balance between perceived burden and gratifications associated with the information requests made in questionnaires."

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 the test persons and probably has a positive effect on their attitude towards the statistical office. We think, however, that as web surveys become more and more common, the innovative image will fade and concrete solutions will become more important for the respondents. Hence we think that the most important effects are mediated through the way web surveys affect design features and what is required of the respondent. Therefore the starting point should be to specify the general design properties and personal characteristics that may affect the respondents' perception of a survey, irrespective of what kind of communication technology that is used.

3.1. Survey properties

We have divided the survey properties into instrument features and data collection procedures. Fisher and Kidoniefs make the same kind of distinction. We think that the features that may affect the respondents' perception of the questionnaire can be summarized under these points:

- The number of questions, indicating how much information the respondent will have to provide
- The content of the questionnaire, meaning the combinations of wordings, information tasks and response formats
- The flow of questions and of different elements in each question
- The layout of the questionnaire.

Similarly we suggest that the data collection procedure can be characterised by these elements:

- The contact mode, which includes the communication channel used, the control exercised over the respondent and how long the contact lasts
- The recruiting strategy used, which includes tracing and motivating the respondents
- Administrative tasks before the questionnaire can be answered, during the answering process and after the questionnaire is completed
- Security measures that are designed to protect the respondent and the information he/she reports.

3.2. The web survey toolbox

Internet distribution is one communication channel for surveys. But the way we see it, the Internet technology affects the questionnaire and data collection procedure in more ways than just being another contact mode. It is an external condition that also has consequences for all the other elements of the survey design.

In contrast to the portrait format of paper questionnaires, the pc-screen has a small, landscape format. The screen resolution is also lower. Normally the space for the questionnaire is smaller because we need space to communicate help texts and error messages and to offer navigational information to the respondent. The difference in presentation format and readability needs to be taken into account when one decides the layout of the web questionnaire.

Another important difference between a paper and a computer presentation of a questionnaire is that it is difficult for the computer respondent to have a visual impression of how long the questionnaire is and how it is organised. This may confuse and disturb the respondent. Usually one tries to solve this problem with the help of a progress indicator that indicates how far the respondent has come and how far he/she has left to go. How these indicators are designed and work may later lead to a feeling that the questionnaire was easier or more burdensome than expected.

The main feature of computerized questionnaires is of course that the computer program that presents the questions, can also act as a virtual interviewer. It is possible to control the flow of questions, perform checks and probing and offer help to respondents that find terms or instructions unclear. In a web survey this can even be done by auditory means and with the help of moving pictures. There are both simple and advanced versions of these communication features. Hence one important difference between different web applications concerns how similar to a fixed paper questionnaire or to an interactive interview they are. From a quality perspective it is tempting to include as many pop-up definitions, quality checks and automatic probes as possible. But this kind of feedback may

not have the positive effect on the response quality that is anticipated, because interruptions and error messages also may lead to frustration among the respondents.

In interviewer-mediated surveys, it is part of the interviewer's job to trace the potential respondent and motivate him/her to take part in the survey. After the interview has taken place, the interviewer also sends the completed questionnaire back to the survey organisation. When a paper questionnaire is mailed to the participants, we also transfer several of these administrative tasks to them. The recipient of the questionnaire has to find time to answer the questions and it is his/her responsibility to mail back the completed questionnaire before the deadline. Hence self-administered questionnaires demand more of the respondents than interviewer-mediated surveys. In business surveys the administration tasks are also more difficult than in surveys of individuals. One and the same firm might have to fill in more than one form and one and the same form might have to be completed by several employees. Sometimes different employees might also need to be able to look at the reported information without being able to change any of the answers. Thus different roles and levels of access have to be given to different persons. In web surveys these tasks can be controlled by administrative web questionnaires. How well designed these administrative questionnaires are and how well they work may have great consequences for the burden felt by the users of the computerized data collection system.

Another interesting aspect of the access and authentication procedures in web surveys is what they do to the respondent perception of security. 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 ID numbers, passwords, digital signatures and encryption procedures may or may not reestablish these walls in the minds of the respondents.

3.3. Respondent characteristics

When they specify what they mean by respondent burden, Fisher and Kydoniefs focus on well-established general values and attitudes that are difficult to change. We would rather focus on the more volatile *interest* in the topic of the questionnaire and on the *competence* that the respondent feels that he/she has to answer questions about this topic. Research has shown that these factors are important for what strategy the respondent chooses when answering the questions. If the respondent is poorly motivated or does not feel fully competent to answer the questions properly, he/she might choose a satisficing strategy instead of the step by step reasoning we want him/her to choose (Krosnick 1991). Some of the respondents' interests may be embedded in values and attitudes that are difficult to change, and their competence levels may also be rather fixed. To some extent we can also make things more interesting and make the tasks seem easier than the respondent initially expected. Therefore, personal characteristics like interest and competence are more interesting than those values and attitudes that are rock-solid.

The amount of time and concentration the respondent is willing to invest in the questionnaire is suggested as a third personal characteristic. The keyword used for this characteristic is *availability*.

When a questionnaire is presented on the web, it is no longer only the available time and the motivation and competence to answer the questions that counts. The available

technology, the interest in using the Internet for surveys and the perceived Internet competence are also important factors. In this respect web surveys may add burden to respondents that have no problem with the paper versions.

The characteristics of the respondents may to a certain degree be a part of their personality. But no respondents act in isolation. It also seems obvious that respondents' availability, interests and competence are affected by the external conditions they act within. In business surveys the most important external condition is the business the respondent is working in. Managers decide how much time and effort the respondent is allowed to use on the questionnaire. In most business surveys the respondents need to look up information in documents or to consult other employees before they can fill in the questionnaire. The availability of relevant documents and the competence level in the company are therefore an important condition for the respondent's work. Finally, it also seems obvious that the importance that statistics play in the production or service carried out by the company may influence the respondent's interest in statistical surveys.

Willimack and Nichols (2002) have developed a model for the response process in business surveys that describes it as a combination of decision processes and cognitive processes. According to this model the response process can be divided into eight steps:

1. Encoding information in memory or by record formation
2. Selection and identification of the respondent(s)
3. Assessment of priorities
4. Comprehension of the data request
5. Retrieval of relevant information from memory and/or existing records
6. Judgement of the adequacy of the response
7. Communication of the response
8. Release of the data.

One recognises the classical four-step cognitive process in points 4 to 7 of this list (Tourangeau et al. 2000), while points 1 to 3 and point 8 concern decisions that are often taken by others than the respondent himself. These decisions are external conditions that probably affect the availability, interest and competence of the respondent. In Figure 4 the four-step response process is drawn as an arrow that is divided into four parts, while external decisions that affect the response process are linked to the internal distribution of the survey, the response conditions and the return of the completed questionnaire.

The model draws attention to the fact that business surveys usually are distributed in two steps. First the survey organisation distributes the survey to the company. Secondly it is internally distributed to the respondent(s) selected by the company. That is why we have split the line that links the survey with the respondent into two arrows in Figure 4. The two-step distribution is an important aspect of business surveys that we take into consideration when the distribution is computerised.

4. Observed Burden and Gratifications in Internet Business Surveys

So far, approximately 30 surveys and 100 questionnaires have been implemented in IDUN or AltInn, which are the two web survey applications Statistics Norway use to distribute business surveys on the Internet. The development of the IDUN system started in 2000,

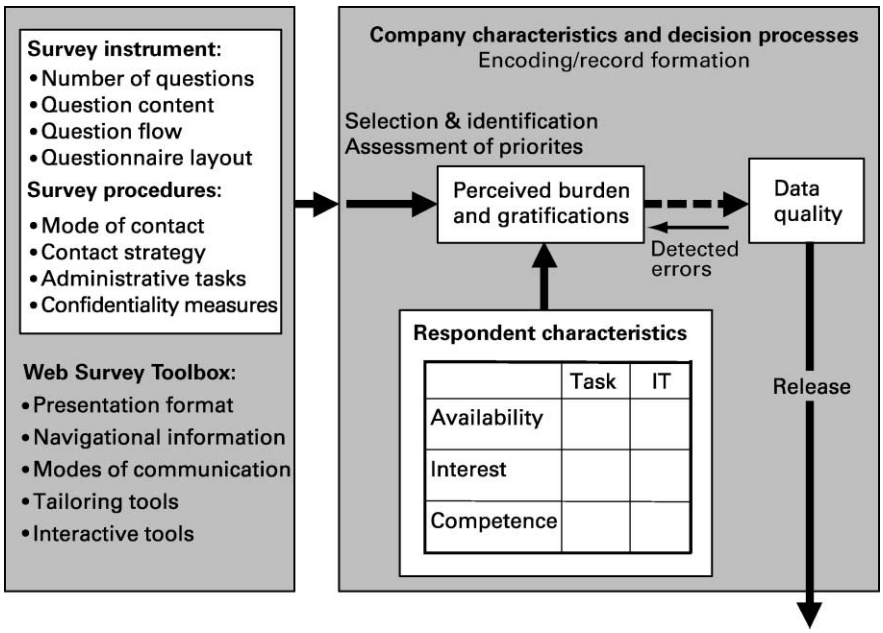

Fig. 4. A conceptual model of survey burden and gratifications in Internet business surveys

while the development of the AltInn system started in 2002 (Sæbø, Gløersen and Sve 2002; Vik 2004).

None of these systems have gone through any extensive or systematic user testing, but some think-aloud sessions, utility tests and pilot studies have been carried out. In the IDUN project, three qualitative tests were carried out in 2001. These tests were done with the Monthly Retail Trade Index Questionnaire and the questionnaire used to collect the quarterly investment statistics. These were the first two business questionnaires that were implemented in the system. The test persons were selected so that two of them had previous experience of the questionnaires, while the third did not have such experience. The first two were also familiar with computers, while the third had never used a computer before. The test persons worked in companies of different size. All tests were run in the companies where the test persons worked, and all tests were videotaped. The test persons were first asked to fill in the paper version of the questionnaires. Then they were asked to fill in the web version. During these two sessions they were encouraged to tell aloud what they read on the screen, how they understood the questions and instructions given and what they did in order to answer one question and move to the next. After they had completed the questionnaire, the moderator posed some follow-up questions.

This test method can be described as a combination of an extended cognitive interviewing and an extended utility testing. Cognitive interviewing is an observation and interviewing method that is designed to map the cognitive process from when a question is posed to when an answer is given (De Maio and Rothgeb 1996). When we use cognitive interviewing in web survey test, however, we also focus on the understanding of metaphors and functions linked to icons and buttons in the computer program. Utility tests

are used to observe how computer programs work. Thus the focus is not on how the respondent thinks, as in cognitive interviewing, but on how he/she acts, and on how well the computer functions initiate the work (Dumas and Redish 1999). When a web survey is tested, however, it is not only the usability of the survey program that is observed, but also the readability of the visual presentation of the questionnaire on the computer screen.

In addition to the qualitative tests of IDUN questionnaires, two small pilot studies were carried out. In the first one 11 respondents who completed the Quarterly Investment Questionnaire were interviewed. Ten questions were asked that focused on the different steps from when the respondent logged on to the IDUN system to when he/she mailed the results back.

The second IDUN study was carried out among respondents who had reported the monthly retail trade on the ordinary paper questionnaire on a certain date in 2001. Thirteen of these, who also had access to the Internet, were invited to test the web version of the questionnaire. Six of the companies took part in the test and gave an evaluation of the system (Bergstrøm et al. 2003).

During the development of the AltInn web application, four rounds of usability tests were carried out. This was done in the period from November 2002 to November 2003. In each round, five users tested the application, except for the fourth round where only three people took part in the test. The three first rounds of usability tests involved only well-experienced computer users. The fourth test comprised inexperienced users only and the last test was carried out with both experienced and inexperienced test persons. All the usability tests followed the same procedure. The test sessions were split into two parts. In the first part the participants were given a practical exercise, and were encouraged to think aloud while they solved the task. In the second part they were asked some follow-up questions. Two of the tests that were conducted in round four were both videotaped and recorded with the aid of a program called Camtasia.[6] This program records the mouse movements and keystroke activities during the test session. When the video and Camtasia recordings are played back it is possible see and listen to the respondent at the same time as one can watch his/her navigations and actions on the screen.

We have no reason to claim that the observations we made during the development of the IDUN and AltInn project can be generalized. Nevertheless we would like to give some examples that we think highlight points that occurred repeatedly in several tests. While we are waiting for more systematic studies of how the business web survey systems work, these are the first pieces of information that could be fitted into our conceptual model.

4.1. Computerized survey procedures

Even though all the tests showed that respondents who both filled in a paper and a web version of the same questionnaire needed more time to complete the web version, they still claimed that they would prefer the web version in the future. One of the test persons said: "The worst part of the paper survey is to remember to mail the questionnaire before the deadline. With a web system I can complete the whole

[6] A demo version of the program can be downloaded from http://www.usabilityfirst.com

task in one operation." We believe that this test person has an important point. The source of frustration is often not the questionnaire itself, but the work that has to be done before and after the questionnaire has been completed. In a web survey system many of these activities are computerized and governed by questionnaires. Before the respondents can access an IDUN or AltInn questionnaire they have to log on to the system and pick the right questionnaire from a list that contains all the questionnaires that the company is obliged to answer. This procedure takes several screens to accomplish. The heart of both the IDUN and the AltInn system is the screen where the respondent picks the questionnaire that he/she wants to answer. In the IDUN system, all users are also forced to update information about their company before they can access the statistical questionnaire. This is in fact a small questionnaire, which we call the Company Fact Sheet. After the questionnaire is completed the respondents are directed to a mailing procedure. In AltInn most of the error checks are also run at this point of the response process.

Many of the problems that were detected in the tests occurred before and after the respondent had filled in the questionnaire. In the AltInn application, many respondents were not able to run the login and authentication procedure properly. Even if they had just updated the Company Fact Sheet, IDUN respondents learned that this screen came up every time they tried to fill in a new questionnaire. There was also a lot of confusion about when the given answers were actually transferred to the recipient. Respondents who updated company information did not understand that the action button "Accept" meant that the pre-printed information was accepted, while they needed to use the "Save" button in order to save their updates.

The AltInn respondents were frustrated because the mailing of the questionnaire was halted by error messages that asked for corrections before the questionnaire was accepted. These examples all illustrate that it was not the web survey instrument that caused most problems, but the computerization of survey procedures.

4.2. Text conventions and visual designs

The different pieces of information in business questionnaires are often presented in an awkward order and without any text convention that clearly separates the question, definition and response areas. Many questionnaires do not have numbered questions at all, while others use multilevel juristic numbering. There is no standard left or right alignment of the response boxes, and different techniques are used to indicate skips. When questionnaires from different subject matter areas are gathered because they are to be presented in a common web application, a variety of text conventions and visual designs are revealed. One may well imagine that respondents who are set to complete several questionnaires for their company have felt it an extra burden that the paper questionnaires are so different. When these questionnaires are computerized it is necessary to use one standard for their visual design. Web survey projects like AltInn and IDUN may therefore play an important role in streamlining the visual design of business surveys.

More uniform questionnaire designs will probably make the work easier for those who fill in many questionnaires. But the effect will be even greater if the conventions and

standards followed also are good. The recordings of mouse movements and keystrokes that were done in two of the AltInn tests address this issue. Unfortunately the impression these recordings gave was that the AltInn application had too many visual elements and options on the screen. This obviously confused the test respondents. The mouse often scanned the screen aimlessly.

Figure 5 shows a skeleton diagram and an example of what an AltInn screen looks like. As one can see from the skeleton diagram, the questionnaire is surrounded by different functions. One can switch between navigational information and enclosures on the left hand side of the screen. The titles listed in the navigation area are also hyperlinks to different questionnaire screens. Help and error messages are shown on the right hand side of the screen. Help messages can either be initiated by the survey program or by the

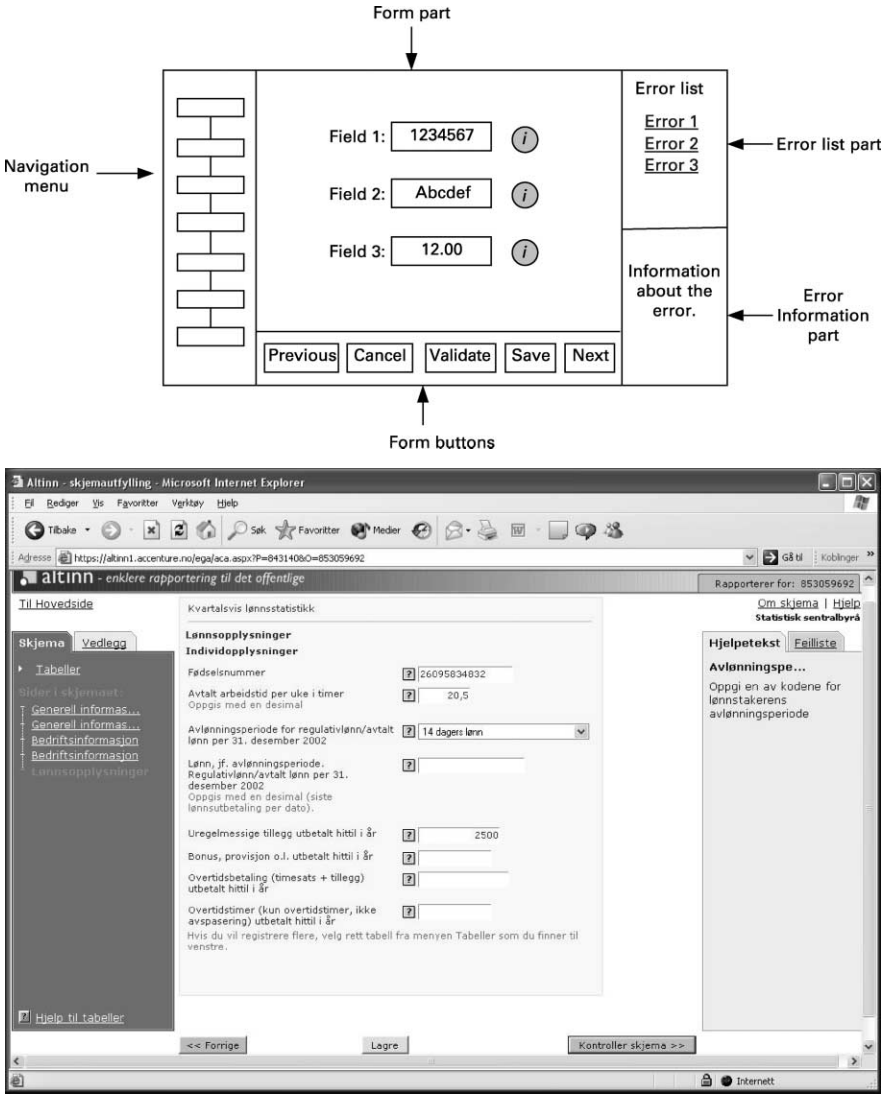

Fig. 5. Skeleton diagram and facsimile of an AltInn screen

respondent if he/she clicks on the question marks at the end of the questions. Note, too, that the third question in the example has a drop box. The action buttons used in the application are placed at the bottom of the screen.

Most of the functions packed into this screen can probably be useful. But if the screen layout becomes so complicated that it confuses the respondent, what were meant to be useful options may become obstacles to an effective completion of the questionnaire.

4.3. Navigational control

The cognitive interviews with respondents who reported their retail trade were carried out with two very different test persons. The first one was the owner and head of a letting firm. He had already reported his retail trade many times on paper. In addition he was an experienced Internet user. The second test person ran a small leather shop together with his wife. He had neither any prior respondent experience nor any computer experience.

As soon as test person 1 had opened the page introducing the web questionnaire, he clicked on every hyperlink he could find on the screen in order to see where they would lead him. The result was that he was taken directly to the Company Fact Sheet instead of the questionnaire that he was expected to fill in. Here he corrected some of the background information about the company, but used the back button on the web browser menu to move back to the system's homepage. Because of this, his corrections were not recorded. When he eventually entered the retail trade questionnaire, he was first brought to the Company Fact Sheet for the second time, only to find than none of his corrections from the previous visit had been saved.

After this test person had completed the survey, a receipt page with two options appeared. One was a link back to the home page of the system, while the other was a link to Statistics Norway's home page. Contrary to what we wanted, our test person chose the last option. From here he found a link to Statistics Canada. The survey system that he was expected to work in was now forgotten.

Test person 2, the novice and amateur, read thoroughly through the introductory 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 supposed to do next." We think this probably was the most telling comment given during all the tests.

Even though 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 a simple instruction that gave them only one option on how to move forward.

4.4. Handling error messages

In the IDUN application error messages were initiated when the respondent moved from one screen to the next. He did this by clicking an action button in the right corner at the

bottom of the screen. The error messages, however, were presented at the top of the screen. Because the screen did not change before the error was corrected, and the respondents focused on the bottom right of the screen, neither of the test persons saw the error messages. They wondered rather why the survey program did not move to the next screen. Thus, as long as the respondent did not see the error messages, he was stuck with a program that apparently had stopped working.

The test respondents also often complained that the error messages were difficult to understand and handle. This was particularly true in the tests of the AltInn application. In this application the respondents could initiate error checks at any point during the answering session, but this option was never used. At the end of the response session an automatic check was initiated before the respondent was allowed to send the completed questionnaire. The result was that when the respondent expected that the job was finished, he was instead presented with a list of errors that had to be corrected before he was allowed to leave the application. The next problem occurred when he tried to move back to the point in the questionnaire where an error was detected. In order to move back to this point, the respondent was supposed to click on the error message, which functioned as a hyperlink. But since he had learned that it was the action buttons that brought him forward and backward in the questionnaire, he tended to use an action button named "Back to the questionnaire." But this button only led him one screen back.

Many of the problems that were observed with the error messages can probably be solved by better presentations, or will be solved by the respondents themselves when they get more used to the applications. Handling error messages is a new and unfamiliar task for the respondents, and it causes them extra work. Error checks prevent low quality, but the ambition should be that the questionnaire is so good that none of the checks are activated.

4.5. Anticipated automation

In the Monthly Retail Sales Questionnaire the respondent is asked to report three figures, the net worth of sales excluding VAT, the VAT and the gross worth of sales including VAT. The reason why the statistical agency wants 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 thus appear to be two questions too many.

This feeling of unnecessary questions was accentuated in the web version. All test persons were surprised at and disappointed with a computerized questionnaire that apparently was unable to calculate and multiply one figure by a VAT percentage. One solution that can be easily implemented in a computerized questionnaire is to first ask about the composition of goods that have different VAT rates. If this is known, the respondents only need to report one sales figure. The other can be calculated automatically.

This example tells us that the respondents expect that web surveys will take over some of the tasks that they previously had to do themselves in the paper version. When the

application does not meet this expectation it leads to disappointment. A very common response format in business surveys are matrixes where the figures given in the cells should be added up to a total. The argument for this has been that this is an extra check that the figures are correct. If this procedure is maintained in the web survey, we think that the respondent will find the web application to be old fashioned.

4.6. 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). Since most business surveys are mandatory, however, motivation strategies have not traditionally been a big issue in business survey design. This situation changes with web surveys. Even if it is still mandatory to answer the questionnaires, the respondent is free to choose between the paper and the web version. Response rates from the business surveys that have already been introduced on the Internet indicate that the web option does not sell itself. The surveys typically start with response rates between 10 and 15 per cent for the web version.

In the IDUN application we have carried out some experiments with tailored statistical reports that are generated at the end of each response session. The reports compare the information given by the respondent with equivalent figures for his/her kind of industry as a whole. As an example, respondents who report sales figures may be able to compare their sales volumes with the average volume in other companies. When figures are reported over time, the reports can also give a picture of trends.

The results from the tests indicate that such reports can serve as an incentive for some of the companies. Respondents in small firms were particularly pleased with this statistical product. In bigger firms, however, they often have their own marketing departments, which have access to more appropriate statistics than the standardized reports we can produce in a web application. None of the test persons from bigger firms worked in the marketing department. Consequently, even if they were sympathetic to the idea, the statistics were of no great interest to them.

4.7. What the tests tell and do not tell

Even though very few qualitative tests have been carried out, we think that they have effectively revealed important problems in the Norwegian web survey applications. If one looks back at the conceptual model developed in this article, however, it is apparent that the results only shed light on the left part of the model. The tests have given us valuable knowledge about how web survey tools that are used to design the data collection instrument and procedure are received and handled by respondents. We have also observed how response problems and frustration may affect the response quality. But what we have not learned very much about yet, is how the respondents' perceptions and attitudes towards the questionnaires are coloured by their initial interest and competence, and by the time and technology they have available. Even less is known about the effects that company characteristics and decision processes have on the respondents' available time and technology, interest in statistical surveys and response competence. This is the part of the conceptual model that we would like to focus more on in the future.

5. References

Bergstrøm, Y., Wang, J.H., Bakke, S., and Haraldsen, G. (2003). Documentation and Instruction for Implementing Web Surveys into the Web Portal of Statistics Norway. Internal memo 2003/5. Statistics Norway, Oslo, 33–48. [In Norwegian]

Bradburn, N. (1978). Respondent Burden. Health Survey Research Methods, DHEW Publication No. (PHS) 79-3207, 49–53.

Couper, M. and Nicholls II, W.L. (1998). The History and Development of Computer Assisted Survey Information Collection Methods. In Computer Assisted Survey Information Collection, M. Couper et al. (eds). New York: Wiley and Sons, Inc., 1–22.

Cox, B.G. and Chinnappa, B.N. (1995). Unique Features of Business Surveys. In Business Survey Methods, B.G. Cox et al. (eds). New York: Wiley and Sons, Inc., 1–17.

De Maio, T. and Rothgeb, J. (1996). Cognitive Interviewing Techniques: In the Lab and in the Field. In Answering Questions. N. Schwarz and S. Sudman (eds). San Fransisco: Jossey-Bass Publishers, 177–195.

Dumas, J. and Redish, J.C. (1999). A Practical Guide to Usability Testing. Bristol, Intellect Books.

Featherston, F. and Moy, L. (1990). Item Nonresponse in Mail Surveys. Paper presented at the International Conference on Measurement Errors in Surveys, Tucson.

Fisher, J. and Kydoniefs, L. (2001). Using a Theoretical Model of Respondent 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), 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 Current Internet Science - Trends, Techniques, Results, Aktuelle Online Forschung - Trends, Techniken, Ergebnisse. U.D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, and A. Werner (eds./Hrsg.). Zurich: Online Press, Available at URL: http://dgof.de/tband99/.

Jones, J. (2003). A Framework for Reviewing Data Collection Instruments in Business Surveys. Survey Methodology Bulletin, 52, 4–9.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. Applied Cognitive Psychology, 5, 213–236.

Sæbø, H.V., Gløersen, R., and Sve, D. (2002). Electronic Data Collection in Statistic Norway. Paper presented at The International Association for Official Statistics conference on Official Statistics and the New Economy, 27-29 August. Office for National Statistics, London.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). The Psychology of Survey Response. Cambridge: Cambridge University Press, 3–23.

Vik, T. (2004). The AllInn Portal. A Common Portal for Reporting Data from Private Industry to Governmental institutions. Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems, Geneva, 17-19 May.

Willimack, D.K., Martin, J., Whitridge, P., Japec, L., and Lyberg, L. (2002). Current Practices in Questionnaire Development, Evaluation and Testing for Establishment

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

Willimack, D.K. and Nichols, E. (2002). Building an Alternative Response Process Model for Business Surveys, Memo, U.S. Census Bureau.

Received February 2003
Revised April 2004