

Developing CASI standards at Statistics Netherlands

Frans Kerssemakers, Statistics Netherlands

1. Introduction

Apart from the advantages computer assistance can offer it is above all the expected increase of efficiency in the processing of data that makes CASI (Computer Assisted Self-Interviewing) or, more general, the use of CASQ (Computerized Self-Administered Questionnaires) one of the spearheads of Statistics Netherlands policy. As a consequence there is a rapidly increasing demand for CASI questionnaires. This is particularly true for business surveys where self-administering has always been the common mode of dealing with questionnaires.

One could argue that CASI is not a very precise term. For generally there is little interviewing or interaction involved. However, the focus in this article will be on (Blaise) questionnaires as more or less standardized user-interfaces, which are filled out by a respondent who still has some help desk or fieldworkers' assistance at his disposal. Therefore a term like CASQ would be too broad.

This article will focus on recent CASI developments in business surveys. Some experiments were carried out with persons and households, and although CASI is considered an option here for especially panel studies, there is still a lack of practical experience. Section 2 gives background information on some surveys that represent different types of CASI applications, from simple to complex and from more dedicated to more general. Section 3 takes a closer look at a questionnaire that suffers from space constraints, demanding quite a bit from the layout features of Blaise. Section 4 describes the present efforts to develop standard settings and guidelines on the basis of some kind of form meant to function as a template and test object. Section 5 looks to the future and makes some recommendations concerning CASI for business surveys.

2. Background

Statistics Netherlands has already gained a lot of CASI experience with particular kinds of applications. Most prominent are two business surveys where CASI was taken into large scale production some years ago and has been successfully applied since then (next to the paper forms which are still used by most, especially smaller enterprises).

1. One is the use of e-mail facilities for the so-called short-term production statistics on a monthly or quarterly basis. Here very short question forms are used, with often only one or two questions. Up to six questions are put on a HTML-page which has only limited built-in intelligence (e.g. range checks). The page is sent to enterprises by e-mail (e.g. asking for the turnover rate). Many different versions of these short question forms, both on paper and in browser-readable form, can be generated automatically based on a system called Logiquest.

2. The other is a dedicated application for statistical declarations concerning international trade. Here a CD-ROM is sent by mail. It contains mainly a comprehensive (fully Blaise based) system for the coding of imported and exported goods, which can be used stand-alone or in connection with a traditional DEP-based Blaise questionnaire. The latter is also on the CD-ROM, together with facilities for tailoring data collection to the respondent's own needs (such as creating a file with own product codes linked to the official codes, or options for including default values). Although lengthier the fields in the questionnaire take up only one page. The highly technical nature of the questions and the use of question-by-question help for extensive explanations give the DEP more the appearance of what used to be known as a CADI data-entry machine than what is commonly understood by a questionnaire.

The Traffic and Transport Survey

In the spring of 2002 a CASI version had to be developed for a (road) traffic and transport survey. At first sight this survey seemed to have much in common with the international trade survey. In both cases the respondent's task is of a highly repetitive nature. Analogous to every transaction in international trade, every journey of a carrier during a particular week had to be reported. On closer consideration, however, the information that haulage businesses were to provide, mainly by describing their journeys, much more resembled the kind of information asked from private persons, when describing their travel behaviour for instance. The questions asked for only one journey could cover several pages, including some Blaise tables. Besides, information had to be collected on the enterprise and on every vehicle in the fleet. It was decided that a hierarchical design with several DEP's would be easiest. Each unit on which information had to be provided (i.e. enterprise, vehicle and journey) got its own question form. The three questionnaires were linked in a shell which was built in Maniplus. The best way to prevent the respondent from getting lost in a labyrinth of screens was in our opinion to build a strict hierarchy into the design. First deal with the enterprise, then with the vehicles, and only after a vehicle has been dealt with can the respondent go on with the journeys of that particular vehicle. To continue with journeys of another vehicle first that other vehicle had to be selected. Dialogues were used in Maniplus to offer a clear view of respectively vehicles and journeys already dealt with. From here one could open a new question form or take a look at an old form. Because a reported journey, on its turn, may consist of up to four deliveries of goods, we thought it very important to reduce the number of pages in this largest questionnaire (with one form for each journey). At the same time each page should cover a clear-cut task or recognizable subject in the respondent's eyes. What we needed from Blaise was much flexibility to deal with the inevitable space constraints.

EDR (Electronic Data Reporter)

For installing the necessary software on the user's platform, getting access to the different dialogues and data models and sending back data, a Blaise-based management system called EDR (Electronic Data Reporter) was used and further developed. (For further information on EDR, see den Bolster, 2003.) EDR should serve as the central portal for all the self-administered Blaise questionnaires that are DEP-based, as opposed to WEB-based (designed as websites on the Internet).

A standard model for testing CASI user interfaces

Designing and building the traffic and transport survey was rewarding. Much was learned from laying out the question texts, answer lists and instructions. But it was also a time-consuming process. Developing standards or guidelines was therefore desirable. However, both complexity and space constraints forced to specific, tailor-made solutions. These make the survey less suitable as a standard model or

template for the average multi-page CASI form. It is also not suited as a web survey. Especially the screens for a journey show some closely written pages. Also, the filling process may be interrupted several times. Where should data then be stored in a secure way and for how long?

We therefore needed another survey, with a less extreme questionnaire and yet having enough differing style elements, a survey that could alternatively be conducted as a DEP-based survey or as a web survey. Both modes should preferably be based on the same Blaise data model. For a web survey we now can apply two different *style sheets*, written in XSL (eXtended Stylesheet Language). One is meant to reproduce the style of the DEP and another is specially written for web use and what web users may expect from a web site (with buttons instead of a main menu, for instance). See for further explanation Bethlehem & Hoogendoorn (2003).

An opportunity was offered in March 2003 when there was a demand for a CASI version of the survey on Vacancies and Absentee rates. This survey is made up of both short and more extensive modules on the same subjects. On the paper form is a grid with 13 columns, which is too wide to be displayed as a Blaise table without scrolling. It has to be broken up. There is another table which serves as an auxiliary tool for calculating absentee rates. In Blaise it can be handled as a parallel block and displayed on a tab sheet. So we chose this survey to develop our standard model.

About to start is the next stage, consisting of thoroughly testing the draft standards and guidelines.

3. A DEP-based style when much space is needed

As described in the previous chapter we wanted to reduce the number of pages in what was by far the largest questionnaire in the hierarchically designed Traffic and Transport Survey. Each record represents a particular journey. The main body consists of a repeatedly asked set of questions, up to a maximum of four, about each delivery of goods during the journey (where they were loaded and unloaded, the type of the delivered goods, their weight etcetera). The preceding questions are meant to characterize the journey and to determine the routing (depending in particular on whether the journey was empty or loaded, and if loaded, how many times goods were delivered, and whether the vehicle itself had been transported, for instance, by train).

To give the respondent a basic idea of the structure of the questionnaire it was important to have all preceding questions on one page or screen. For then we would have the simple structure of an opening page, followed by up to four identical blocks about the delivered goods, and finally, if applicable, a table for the just mentioned combined transport and another to list transit countries. Companies providing transport for hire and reward, for instance, had to report details on every journey during the reference week. In exchange they would not be bothered the rest of the year. For some this could amount to hundreds of journeys. Eventually we chose the solution that is shown in figure 1.

Figure 1. The first page of the Traffic and Transport Survey

boxes can be left empty. To get familiar with these and other types of questions we use standard instructions above and below the respective fields. If there is a lookup question the respondent can as a rule still enter a description in the field and have it accepted, after a signal has come up if it does not match with the list. But, except as an escape, this is not what we want. By using a recurring instruction we are hoping for an almost automatic double-click or press on F8 to open the list and start the search. A tick box seems appropriate for the question whether combined transport did occur or not. But in Blaise such a box is reserved for set questions. So we had to specify a question with two categories, the second of which got an empty text and is not displayed, thanks to the option 'hide empty categories' in Blaise (see *Project/ Datamodel Properties/ Types*).

We also agreed on the convention to put an asterisk behind the field label, which we use instead of the field name, indicating that the question can not be skipped. Something has to be entered. In some cases, such as with 'Volume (used) in %' and 'Surface area (used) in %' a choice can be made which of the two to use. Here a check is needed to ascertain that at least one of the two is answered. It may be a good idea to have a symbol also for choices of this kind.

If there is a loaded journey the questions continue on the second page which is shown in figure 2 (on the next page). In the first column the name of the village, town or city has to be coded where the delivered goods were loaded respectively unloaded. All together there are 270,000 place-names in the lookup files for about 50 countries. First the country code has to be known. Usually the two letters of this code are entered directly. But use can also be made of a lookup. Depending on this code a lookup list for a particular country will be opened. This is functioning well. In the other column a huge lookup list is used for the coding of the delivered goods or commodities. This list contains about 37,000 sometimes rather lengthy descriptions. For finding the appropriate code a trigram search is used.

In contrast with the first page of the DEP we did not succeed in getting all the questions about a particular delivery on one page. This is mainly due to the rather long answer lists displayed, even though we did not display the list for the site of loading in the same form for the unloading site. Clicking on the latter field will reveal that there is a drop down box which can be opened. There is also an instruction below the field to look up at the displayed categories in the middle. The order of the questions was changed such that a relatively special subject, namely dangerous goods, came on the next page. It only applies to a modest number of businesses and journeys. And if it applies the classification code will very often be known already by means of a link with the general classification code for goods. It usually only has to be confirmed and can be overwritten.

Figure 2. Subject ‘delivery 1’ on second page of Traffic and Transport Survey

Number plate: STX235 Week: 18 DELIVERY 1 Distance travelled: 1080 km

What are the carriage returns on this delivery?
In whole euros, excluding VAT and extras like toll, transport taxes, ferriage etcetera.
 In case of **combined deliveries of goods** enter the total carriage returns.

enter delivery* Distance delivery in km* 750

Place of loading (goods) delivery 1 ?

Land code (FB)* NL double-click or <FB> to start lookup list
 Nederland

Place-name (FB)* Heerlen Type of goods* Ethyleenchlorbromide

ZIP code digits 6412

Type of site* Gross weight delivered goods in tonnes 6,8

1. production/extraction 5. railway terminal
 2. consumption/processing/retail business 6. airport
 3. seaport 7. bonded warehouse
 4. inland port 8. distribution/wholesale business
 9. other (also truck/lorry stand)

Place of unloading (goods) delivery 1 ?

Land code (FB)* FR double-click or <FB> to start lookup list
 Frankrijk

Place-name (FB)* Orange Type of goods* Ethyleenchlorbromide

ZIP code digits

Type of site* 8 Carriage returns delivery* 560
see 1...9 next to Type of site* above

1. liquid bulk (unpackaged) 6. hanging goods
 2. solid bulk 7. pre-slung goods
 3. large freight containers (20 ft and more) 8. mobile, self-propelled units, including live animals
 4. other freight containers, including rolling containers 9. other mobile units
 5. palletised goods 10. other cargo types

Working with the Modelib Editor

Here are some major experiences regarding the use of *Modelib* for creating a flexible design.

- The *Modelib* offered practically all the flexibility we needed, thanks also to some enhancements introduced during the development of Blaise 4.6 (e.g. enriched text now everywhere possible).
- To fully exploit the possibilities a lot of practical exercise is needed.
- Once mastered, spacing out and tuning the question elements on the formpane remains time-consuming because of the often needed reshuffling, rephrasing and stressing of elements.
- For a flexible design quite a number of *Modelib styles* are needed. If properly ordered, however, they can save a lot of time.
- Although flexibility presupposes freedom of design, the availability of guidelines, standard elements and templates can be a great help.

4. Developing a standard model for testing purposes

Our next aim was to develop a prototype questionnaire that could suit our needs for testing layout and navigational issues. In short, it should make the carefully prepared testing programme pay off. As a starting point we chose to join the colour settings from the website of Statistics Netherlands. We also wanted to use the company logo as an identifying mark in some way. For a versatile design we had to cope with the fact that the length of both question texts and explanations or instructions can vary a lot. A basic assumption was that help should be at hand, whenever possible. While reading down the page the respondent will usually focus on the data-entry fields. So this help should preferably be close to the fields. Too much text, however, may irritate. We therefore decided to use different panes. For a study on how to position explanations and instructions, see Jansen and Steehouder (2001).

A logical solution would be to use the infopane for explanations etcetera and position it on the right side of the screen. For a quick view of the page the respondent can then focus on the middle. Pane. By horizontally aligning with the data field the explanation or instructions for that field are readily available. The text on the infopane will move up and down, depending on the selected field. This can be a help in offering a visual cue to find the active data field. On the other hand, it may look strange if, with certain questions, there is nothing on the infopane.

The example in figure 3 is made with Blaise IS, using the XSL stylesheet for web use, and with some minor manual adjustments. A progress indicator is already available. What is shown on the (turquoise) infopane on the right is the lengthy explanation that belongs to the first question on the screen. It is only meant as an example of a body of text. The last field is automatically changed from 0 to 1 if the routing rules allow the respondent to continue. This OK-question is used to finish a particular subject. It may in due time be replaced by a button. Not all elements are yet fully balanced, but it's a start.

Figure 3. A webstyle based template, with the infopane on the right

The screenshot shows a web browser window titled "Vacancies and Absentee rates - Microsoft Internet Explorer". The page has a header with a logo on the left and the title "Vacancies and Absentee rates" in the center. Below the title is a progress bar. The main content area is divided into three columns:

- Left Column (Navigation):** A vertical list of buttons: "Exit", "Next Question", "Previous Question", "Calc Sheet", "Next Page", "Previous Page", "First Page", "Last Page", and "Info...".
- Center Column (Form):** Four questions with input fields:
 - Question 1: "Total number of vacancies at the end of the quarter*" with a value of 8.
 - Question 2: "Number of vacancies being filled during the quarter*" with a value of 11.
 - Question 3: "Number of vacancies that occurred during the quarter*" with a value of 6.
 - Question 4: "Number of vacancies withdrawn during the quarter from being opened*" with a value of 2.Below these is an "OK: continue with [ENTER]" button with a value of 1.
- Right Column (Infopane):** A section titled "EXPLANATION VACANCIES" with detailed text explaining the questions and providing instructions for counting vacancies.

As far as the designer is concerned, our first impression is that working with two panes next to each other is relatively easy. The division line between infopane and formpane can be shifted from page to page depending on the space needed. Even more space is available if the column with buttons is removed. This could be achieved by choosing a menu-based style like in the Blaise DEP. But if this space is used for question labels, answer lists etcetera they may not fit the narrower middle pane of the webstyle. If using the same datamodel for either style there will be some room left in the DEP on the left side. If only using the DEP it is worth considering to still use a left column. Financial surveys, for instance, often have questionnaires with modules or blocks that can be dealt with in different orders (likt tab sheets). The column could then be used to show an index and offer a view of the filling progress. Or there may still be some buttons if this proves useful for navigational reasons.

A problem in the DEP for which some kind of solution is required concerns the position of the infopane on the screen. If this position is *Right*, for instance, then the size of the pane on the screen is also determined from the right. As a consequence the pane will shrink to the right if a higher screen resolution is chosen. What is on the formpane, however, will shrink to the left. The information on the two panes will therefore move away from each other. This is shown in figure 4. Instead of the minimum resolution (800 by 600) we support and used so far, the page that is presented in figure 4 has a resolution of 1024 by 768.

Figure 4. The template from figure 3 in DEP-style and with higher resolution

GROUP 6	
Total number of vacancies of whom	4
work experience not required*	2
20 or more hours a week*	4
for jobs still manned*	1
shorter then one month looked for*	1
between one and three months	3
put employment advertisement*	4
job centre informed*	3
considered difficult to fill*	0

The questions displayed in figure 4 show the continuation of a Blaise table which was too wide to see all the columns without using the scroll bar. On the paper form the grid has 13 columns. With a minimum resolution there is only place for 5 on one screen. The rows of the table represent groups of vacancies. One row are vacancies for a particular kind of function within a certain municipality. The latter, together with function, required education and the number of people one will be in charge of, are entered in the table. The last question asks for the total number of vacancies in the group. After finishing the table more specific questions are asked per row or group of vacancies, for instance, on the number of vacancies (within the

group) one considers difficult to fill (the active last question in the figure above). As a matter of fact this could well be a much better way of asking these questions.

Finally, a table is presented in figure 5. It is a parallel in the DEP, put on a tab sheet, and meant as an electronic version of a paper sheet on which, as an auxiliary tool, one could calculate the absentee rate. The table can be used arbitrarily, except for obeying the answer types (e.g. date). Its only function is to enable the respondent to write things down to his own needs and have the absence days added up automatically. The parallel still has a second page on which is asked for the average number of employees during the reference period (a quarter). On the basis of the outcomes of the two pages two kind of percentages are calculated. The respondent is free to use these percentages, accepting them, applying a correction factor, or whatever is wanted.

With a wide table the infopane can only be on top or at the bottom of a page. The same is true if there would be added buttons (with a caption). In CASI there seems to be some need for these buttons, especially when using tables. In figure 5 one can see that an extra column is added just to make it possible for the respondent to let the DEP know that the calculations are completed. In the table shown such an explicit button-function, created by an extra column and an extra field, is almost essential. Otherwise the respondent may have used the tab Main Form to leave the table. But then he would not get the percentages. For these he first should have gone to the second page of the parallel. The only escape would have been the use of the next page option in the menu or on the speedbar.

Figure 5. A calc sheet for the absentee rate as a parallel to the main parallel

					Number of days in quarter	Absence through pregnancy/delivery Total number of days	Absence through 'normal' illness Total number of days	If calculation completed enter code 1 at the bottom
period [1]	from	21-12-2002	up to	21-01-2003	21	122	230	
period [2]	from		up to					
period [3]	from	21-01-2003	up to	21-02-2003	32	177	366	
period [4]	from		up to					
period [5]	from	21-02-2003	up to	21-03-2003	29	165	266	
period [6]	from		up to					
period [7]	from	21-03-2003	up to	21-04-2003	11	63	118	
period [8]	from		up to					
period [9]	from		up to			10		
period [10]	from		up to					
period [11]	from		up to					
period [12]	from		up to				11	
period [13]	from		up to					
Totals						537	991	1 - completed

If calculation **completed** enter 1 (= completed).
To return to the main form without completion : click on **tab Main Form** or press **[Ctrl] + [Tab]** .

5. For the future

As a provisional figure, it seems that half of the transport businesses receiving the CASI software are actually using it for providing their data. As a beginning this is more than we had hoped for, given the heavy burden and the technically complex questionnaire. But it is also known that it may well be difficult to raise this initial

figure (once the decision is taken not to use CASI). Against this background it is important to show a feeling of understanding for what is going on in the world of transport and to give something in return whenever practicable.

While pre-visiting a limited number of businesses we noticed a remarkable willingness to cooperate and think along with us on how to improve CASI. They came out with smart ideas, especially on how the questionnaire could be better customized. The journey questionnaire for the transport survey, for instance, would be easier to understand if it did not have to cover such a wide range of different types of transport. A company that provides transport for a chain of supermarkets makes very similar journeys every day. A haulage business providing international transport, on the contrary, often has very differing types of journeys. Making all kinds of dedicated questionnaires may be impracticable with longer and more complex forms. A user created profile of settings, however, could simplify the filling process and enable us to anticipate better the special problems connected with the kind of business activities. It would be much easier to explain what should be considered a journey or a delivery on the basis of a typical example from the world of supermarkets, for instance. Automatic filling with default values, as already practised with the international trade survey, may also help.

One of the things that could be returned to businesses and what they are repeatedly asking for is a facility to print a report of what they just filled out, e.g. a report giving a good view of the fleet.

Among the obstacles for using CASI those occurring because a transport business has different establishments were the most striking. Sometimes a system manager was needed who was at headquarters far away. Regularly also one wanted to have the truck drivers themselves fill out the questions on their own journeys. And although EDR is multi-user, one apparently found it more difficult to organize the filling process than when paper forms could be distributed.

For the kind of business surveys discussed here we do not expect that the use of internet will expand rapidly. There are still problems of security with regard to data transmission and keeping data available on the less protected external network (to continue filling out forms after interruptions). But as it is, the internet can very well offer services such as detailed and up to date classifications together with sophisticated searching mechanisms. The internet is after all already the central mode to consult statistical information in The Netherlands. For experiences of Statistics Netherlands with different forms of using the internet and e-mail facilities for CASI data collection, see Roos (2002).

Although the DEP-style is menu-based, with CASI there seems to be a need for at least some buttons. As respondents get more experienced with the internet, they may also increasingly expect to find some buttons for navigation. For the more distant future it may even be considered to implement a web look-alike in the DEP, especially for reasons of navigation, using buttons together with captions ready at hand for the respondent. For CASI it is also important to have layouts that are not disproportionally distorted by the screen resolution of the user.

References

Bethlehem, J. & Hoogendoorn, A. (2003): Methodological guidelines for Blaise web surveys. Paper presented at the 8th International Blaise Users' Conference, Copenhagen.

Bolster, G.W. de (2003): Electronic Data Reporter. Paper presented at the 8th International Blaise Users' Conference, Copenhagen, 2003.

Jansen C. & Steehouder M. (2001): How research can lead to better government forms. In: Reading and Writing Public Documents, Benjamins, Amsterdam/Philadelphia, pp. 11-36.

Roos, M. (2002): Methods of Internet data collection and implications for recruiting respondents. In: Statistical Journal of the United Nations Economic Commission for Europe, 19 (3), pp. 175-186