

The Labor Force Survey as a web questionnaire: Results of the German pretest within the ESSnet DCSS project

Karen Blanke & Sabine Sattelberger

Federal Statistical Office of Germany (Destatis)

Abstract

Within the European Statistical System (ESS), data collection for the Labor Force Survey (LFS) has been mainly conducted with interviewers to ensure high data quality. However, due to cost considerations, respondent-administered electronic questionnaires are being increasingly implemented. Switching modes from CAPI/CATI to CAWI is challenging in two ways: Firstly, the respondents can no longer rely on interviewer support to understand the complex concepts of employment. Secondly, they must cope with the technical functionalities, although not being trained in using the tool. So what happens to data quality? In this context, Eurostat initiated a two-year ESSnet project on “Data Collection for Social Surveys using Multiple Modes”. The LFS was selected as an exemplary social survey to assess the consequences of mixing modes and develop a user-friendly web questionnaire. Research was conducted in Finland, Germany, the Netherlands, Norway and the UK. This paper summarizes German pretesting results within the ESSnet DCSS project. The main goal was to improve the design for a web questionnaire collecting LFS data and to electronically integrate the tasks formerly performed by an interviewer. Emphasis was put on three different design elements: navigation through the questionnaire, design of error checks, and layout of instructions. Some critical employment concepts of the LFS have been tested too.

1. Background

To ensure high data quality within the European Statistical System (ESS), the Labor Force Survey (LFS) has been mainly conducted with interviewers so far. However, due to cost considerations, web-based data collection is more and more discussed and implemented for

counterbalancing low response rates, adjusting coverage bias and meeting the expectations of the respondents towards up-to-date official statistics. In this context, Eurostat initiated a two-year ESSnet project on “Data Collection for Social Surveys using Multiple Modes”. The LFS was selected as an exemplary social survey to assess the consequences of mixing modes and develop a user-friendly web questionnaire. Research was conducted in Finland, Germany, the Netherlands, Norway and the UK. The main objectives of the ESSnet DCSS were to stimulate early collaboration among the European countries, to harmonize nationally applied practices, to transfer knowledge and to provide support for countries less experienced with computer-assisted web interviewing (CAWI). In addition to facilitating the introduction of this new mode for social surveys, the project put emphasis on the impact of mixed-mode approaches. This paper will deal with the work package of developing web-based data collection tools.

At the time of the Q2014 conference in June 2014, the project has been almost finished. Findings and recommendations will be available in autumn 2014 [1]. The starting point of the ESSnet DCSS was a query on data collection in social surveys within the ESS and several overseas NSIs in 2013. This query showed that web data collection for the LFS and most other social surveys is still in an early phase of development. However, there is a wide diversity among the different surveys: With regard to the LFS, only two countries (the Netherlands and Denmark) have applied CAWI so far. In contrast, computer-assisted personal interviews (CAPI) and computer-assisted telephone interviews (CATI) are the predominant modes, whereas only some European countries still apply self-administered questionnaires (PAP) for the LFS. When it comes to the Census, however, overall 13 countries of the ESS offer electronic questionnaires as an option for the respondents.

The next step of the ESSnet DCSS was that the above mentioned consortium member states conducted qualitative testing on the design of future web questionnaires for the LFS that should meet the needs of the respondents. In the following, the results of the German research are summarized from the perspective of Destatis. However, the gain of knowledge was intensively exchanged within the consortium so that the presented solutions are highly influenced by the valuable advice of the other participating countries.

[1] For more information, please check <http://www.cros-portal.eu/content/data-collection>.

2. The status quo at Destatis

For several years, Destatis has been discussing how to proceed with and implement web questionnaires for social surveys, as electronic questionnaires had been already successfully in production for business surveys since 2005. After several testing rounds, a prototype tool could be developed for a rather small social survey. These experiences improved the design of the upcoming LFS tool from the beginning. However, due to its household-based structure and its length (more than 150 questions), the development of a LFS online form was much more challenging. A test version of this web questionnaire was programmed in an iterative process of checking and revising (from February to May 2013 and from September to October 2013), using the Java-based in-house software IDEV version 4.

In Germany, data for the LFS are collected within the annual Microcensus, a social survey that involves more than 800,000 individuals. Interviews are conducted with CAPI in most cases (or more rarely by PAP as an alternative option). In the future, PAP shall be subsequently replaced by web questionnaires as an auxiliary mode, whereas CAPI will probably still be the main mode. However, switching modes to CAWI has to be considered as particularly demanding: Firstly, the respondents can no longer rely on interviewer support to understand the complex concepts of employment. Secondly, they must cope with the technical functionalities although not being trained in using the tool. Consequently, usability testing is the essential precondition to safeguard high data quality and reduce response burden.

3. General aims of testing

When thinking about implementing CAWI instruments, one often tends to believe that the IT-features will be so sophisticated that data quality must be high by default due to the integrated error checks and automatic routing. However, this rather unrealistic prospect should be adjusted according to pretesting knowledge gained over the years which has put the assumed strength of electronic forms into perspective. Two exemplary aspects shall be outlined here to illustrate that the development of an online tool is far from being easy [2]:

(a) When it comes to CAWI, completion behavior is comparable to human-computer interaction [3] par excellence – less reflected, very quick and drop-out is only a click away. E.g. reading behavior is similar to screen-reading rather than reading on paper. While being

[2] Couper, M. (2008), *Designing effective web surveys*, Cambridge University Press.

[3] Kaczmarek, L. (2008), *Human-survey interaction: usability and nonresponse in online surveys*, Mannheim.

distributed via the internet, the circumstances of the survey are not always under the control of the researcher, as respondents use different browsers, screens and devices.

(b) Self-completion itself means that there is no help by any interviewers. Consequently, all communication with the respondents can only be transferred via the content and the design of the web questionnaire [4]. Richness on technical functionalities does not always guarantee user-friendliness: "The right functionality – working correctly – is critical, but not sufficient for a product to be successful. A product by itself has no value; it has value, only insofar as it is used" [5]. The respondents have clear expectations how to deliver their information quickly: "Don't make me think" [6]. Unfortunately, this way of proceeding sometimes contradicts the efforts of official statistics to collect reliable and valid data.

The purpose of the ESSnet DCSS pretest at Destatis was to design a functional and user-friendly LFS web questionnaire. The respondents should be able to focus on the content rather than think about how to handle the instrument. Therefore, the flow of the web questionnaire was to be clever and supportive, the tool should assist in minimizing incorrect or incomplete data entries as well as showing only the parts of the form that must be filled in. Overall, high data quality was to safeguard for most respondents, even though some respondents might have preferred another design. Consequently, any decision on layout issues was a trade-off between satisfying the users and achieving quality assurance for official statistics. Pretesting evaluated and supported the efforts of Destatis to make these aims feasible.

4. Results of the ESSnet DCSS pretest at Destatis

The perspective when testing the LFS web questionnaire in Germany [7] was that comprehension problems only became relevant for research when data quality was seen at risk due to critical employment concepts. There were three areas of interest with regard to design when evaluating the CAWI tool: a) the automatic and the user-initiated navigation, b) the type

[4] Snijders, G. et al. (2013), *Designing and conducting business surveys*, Wiley.

[5] Dumas, J. / Redish, J. (1999), *A practical guide to usability testing*, Intellect Books.

[6] Krug, S. (2013), *Don't make me think*, 3rd revised edition, New Riders.

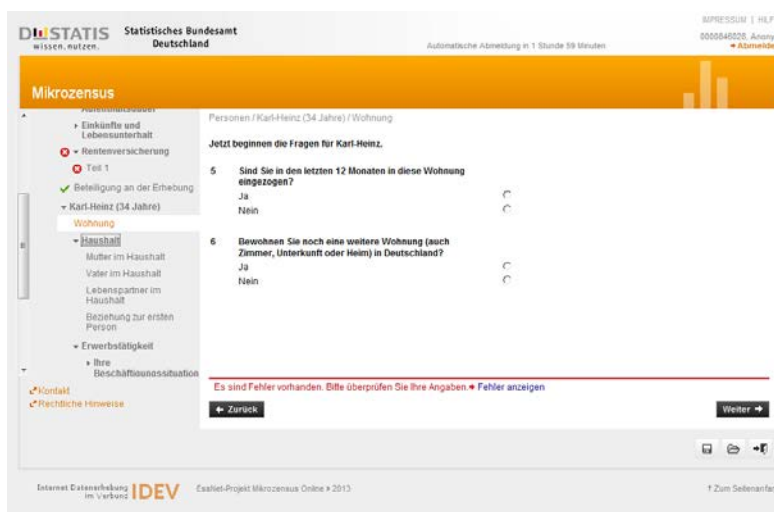
[7] In November 2013, Destatis conducted two waves of qualitative pretesting to capture the respondents perspective. All in all, 19 probands in two wave completed the questionnaire in the in-house pretest laboratory. Given the complex concepts of the LFS, the most important selection criteria were occupation and household size. Since it was predicted that marginal or part-time employment might cause serious comprehension problems, probands in these situations were recruited with priority. With their consent, all probands could be recorded audio-visually while filling in the web questionnaire on their own and in absence of the interviewer. Additionally, a fixed eye-tracker recorded their eye-movements and mouse clicks on screen. For uncovering what respondents might have (mis)understood during self-completion, cognitive interviewing was conducted after self-completion.

and design of error checks, and c) the display of instructions. To sum up the results, the instrument allowed for a correct handling. Need for improvement existed nonetheless with regard to some specific, but also some global aspects of web questionnaire design.

4.1 Navigation, error checks and instructions

Automatic navigation: The navigation in the LFS web questionnaire presented challenges due to its length and the given household approach (see graph 1). All household members were meant to jointly complete one questionnaire. This task had to be guided by a proper design which was not only programmed correctly, but also understood by the respondents. The probands moved through the survey almost exclusively by clicking on the <previous>- and <next>-buttons. Critical was the shift from one household member to the next. During the first testing wave, it was not sufficiently highlighted when the questions for the second household member would start. There was no obvious hint in the perception of the respondents – they overlooked both the header in the breadcrumb and the new name of the section in the navigation tree. In the second testing wave, the switch was indicated by a simple introductory sentence explaining that the following data would refer to the next household member. This approach was effective, as it helped the probands to recognize the break more easily. In contrast to other NSIs (see graphs 2 and 3), Destatis has not deploy an intermediate screen so far (a so called “dashboard”) which would probably announce even more visibly that the data entry for the second person starts.

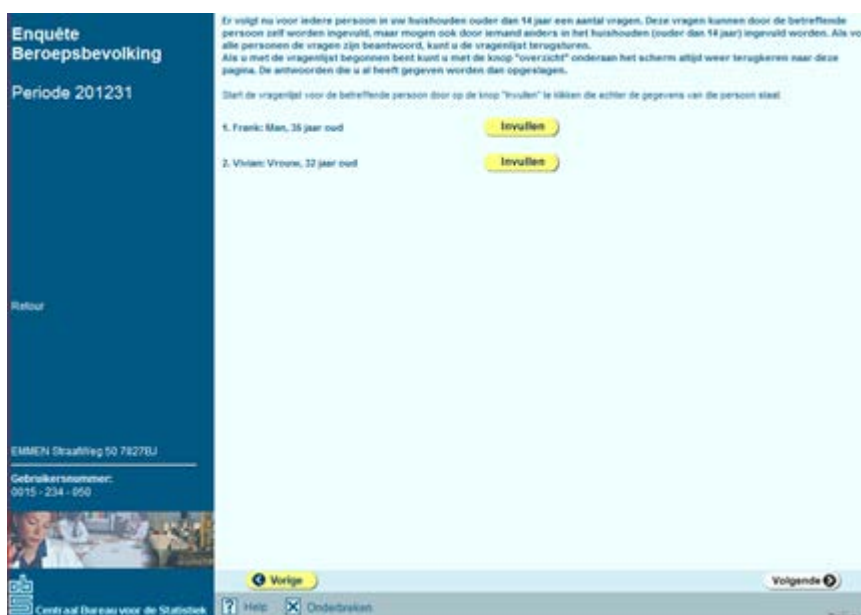
Graph 1: Household structure in the LFS web test questionnaire at Destatis (Germany)



Graph 2: Dashboard in the LFS web test questionnaire at ONS (UK)



Graph 3: Dashboard in the LFS web test questionnaire at CBS (the Netherlands)



User-initiated navigation: Destatis offered the additional option to move through the online form by clicking on a navigation tree on the left part of the screen. During the first testing wave, few probands noticed this functionality at all. When asked for its purpose, most of them regarded the navigation tree as a progress indicator. However, when in need, they also realized to use it in specific cases like skipping respondents or making corrections before sending the web questionnaire. Even though the design was similar to Windows Explorer, most probands felt lost when starting to use the navigation tree, because it was too detailed. With regard to this aspect, more testing is necessary. In future, Destatis will encourage its respondents to use automatic navigation by <previous>- and <next>-buttons because this allows them to move through the questionnaire with little effort. Moreover, the navigation tree is to be promoted as a helpful tool to get quickly to the relevant question when an answer

needs to be corrected or a household member needs to be skipped. Otherwise, too many clicks on the <previous>-button would be required without a dashboard being available.

Type and design of error checks: Automatic data cleaning is an integrated part of web surveys. The respondents are asked to correct implausible or incomplete answers directly during self-completion. They know this functionality from other web applications (e. g. from e-commerce platforms) and also expect assistance by an intelligent IT-system for government forms (e. g. the annual tax declaration) – again it applies: “Don’t make me think” [8]. However, if error checks are implemented too extensively (which can easily happen in complex surveys), the response burden may outbalance the positive effects on data quality, for example, by leading to higher drop-out rates. For subject-matter statisticians, it is yet very difficult to decide on a limited number of error messages. In former pretesting, Destatis has already been engaged in finding a suitable general design for error checks. Two different types could be activated in the LFS web survey: hard and soft checks. Furthermore, errors were indicated on three different parts of the screen (see graph 4):

Graph 4: Error checks in the LFS web test questionnaire at Destatis (Germany)

(1) Above the question, a red text line explained the given error. (2) Within the navigation tree, a red cross additionally showed that an error had occurred. (3) Below the response options, a hyperlink enabled to refer to cross-plausibility checks. During testing, it turned out that the probands did not understand the distinction between hard and soft checks that is self-

[8]Krug, S. (2013), Don’t make me think, 3rd revised edition, New Riders.

evident for official statistics. Moreover, they overlooked the red cross in the navigation tree as well as the hyperlink below the response options. At first glance, these preliminary results might seem rather discouraging, but actually they are not. It is not necessary that respondents fully capture the concept of hard and soft checks. While entering data, they learn that some error checks have to be corrected to be able to continue, others not. Consequently, soft checks are the best compromise to collect complete and high quality data, as it is a smart reminder for the respondents to enter proper data without being too threatening. This might lead to lower drop-out rates. More testing is requested to improve the placement of error messages.

Display of instructions: Comprehensive instructions for difficult technical terms play a major role in self-completion questionnaires. They replace the immediate assistance of an interviewer if comprehension problems occur. During testing the LFS web survey at Destatis, most instructions were written in grey, normal font and placed directly below the bold question. In some cases, further instructions were given in pop-up windows opening after clicks on orange hyperlinks. The labeling of these hyperlinks consisted of a keyword that pointed at the content of the hidden instructions (e. g. “More on the economic sector”). While improving data quality is the main concern of official statistics, respondents try to avoid the effort of additional clicks (once to open, once to close the pop-up) unless they are convinced that it is helpful and time-saving in the end. On the one hand, the combination of directly visible text and hyperlinks proved to be the preferred design in previous testing. On the other hand, it was also shown that respondents seldom read instructions, but rather scan or even ignore them. Official statistics cannot fundamentally change this attitude, but nevertheless instructions should be very visible and easy to access (as few clicks as possible). To sum up, different designs for different purposes are recommendable: If instructions are relevant to almost all respondents, displaying them just below the question is most effective, as they are more often perceived. Less relevant instructions (e. g. for specific subgroups) may be presented with a hyperlink including a meaningful keyword. Overall, sentences need to be cut down because shorter texts are more adequate for the reading behavior on screen. In addition, the presentation in bullet points has proven to be very user-friendly.

4.2 *Critical employment concepts of the LFS*

When it comes to the content of the LFS, pretesting for Destatis showed that respondents in standard employment situations had few difficulties with understanding and completing the web questionnaire in contrast to certain subgroups with a more unusual employment status.

So quantitative testing might reveal that the below described problems are less frequent, but qualitative testing clearly proved that these difficulties were relatively serious in terms of data quality. While trained interviewers were able to assist respondents in answering complicated questions correctly in spite of their atypical employment situations, the respondents face more problems during self-completion. Here are some examples:

Distinction between main status, professional status and status in reference week: The probands were less familiar with the complex and technical concepts of the German LFS. The questions at the beginning dealt with the subjective assessment of their main status (e. g. being employed, in leave, retired, housewife, etc.). Major problems how to allocate to a single response option arose for probands in marginal employment (e. g. apart from being unemployed) or with several small jobs. The next questions focused on any (marginal) employment, the classification of the professional status and economic activities during the reference week. This elaborated sequence of questions would require very careful reading to understand the distinction between the concepts, a behavior respondents usually try to avoid.

Marginal employment and full-/part-time work: Respondents in marginal employment often were in doubt whether their job was to be considered as “employment” at all and were hesitant about how to classify themselves. Classification was facilitated when question and response options clearly required indicating any paid job. Having to choose between working full- or part-time, the marginally employed did not feel like fitting into any category, as they worked very few hours. An additional response option for this subgroup may address their problems.

Job title and economic sector: Open text fields on the job title and economic sector were particularly challenging to fill in for all respondents, but especially for the marginally employed. On the whole, the data entries were low in quality and rarely usable. The comments of the probands during qualitative testing showed some difficulties with regard to the required level of detail. For job changers, it was unclear whether to indicate their trained job or their current occupation. The instructions referred to the latter, but they were not regarded as useful. Paraphrasing the economic sector without any help was even more difficult, as the respondents were not familiar with such a sophisticated classification system and the instructions were too rudimentary.

Registered as being unemployed or seeking a job: Again problems arose when technical terms were used without sufficient instructions. For example, being registered at the German

Employment Agency may mean either to be unemployed and search for a job or to have a job but seek another one. In contrast to the clear definition of public agencies, the difference between these terms is less familiar in everyday language. Without providing a user-friendly designed instruction, respondents might apply their own definition, which may not fit to the official.

Working hours in case of several jobs: In the course of the web questionnaire, respondents with several jobs seldom realized whether questions on their employment referred to all of their jobs or a job in particular. For instance, in case of working hours, some probands added all working hours they had because it was cognitively too complex to keep their occupations separated without having a regular memory aid (e. g. a clearly emphasized reference to the first, second, etc. job). Guidance for this particular group of respondents should be improved.

5. Conclusions

To create a web questionnaire for several household members is a challenge, especially in long questionnaires like the LFS, but feasible after extensive pretesting. Based on the results of the ESSnet DCSS, Destatis will focus its further development of the online tool IDEV on making the different levels within the online form (household versus individual questions) more visible and facilitating navigation. Moreover, the frequency and placement of error checks will be reconsidered as well as the implementation of user-friendly instructions.

The employment concepts of the LFS are rather difficult to present without any interviewer being present in case of comprehension problems. Statistical definitions differ from everyday language. Moreover, respondents do not know how to transfer their personal situation into the rigid grid of the web questionnaire. With good reason, the LFS aims at covering not only standard, but also atypical employment situations like being marginally or self-employed, having several small jobs or working on demand. These deviations from the norm are getting more and more frequent but are hard to capture within a self-completion questionnaire. The types of potential comprehension problems could already be identified by qualitative testing but their frequencies are still difficult to estimate. Quantitative testing would be the next step in improving the LFS web questionnaire.