

Mixed mode in LFS: questionnaire design and mode-effects

Liisa Larja, Senior Statistician, Statistics Finland
Pertti Taskinen, Senior Statistician, Statistics Finland
Contact: liisa.larja@stat.fi

Abstract

In recent years the expectations on NSI's providing electronic questionnaires (CAWI) on the side of other data collection modes have increased. In 2012, Statistics Finland launched a project on designing and testing of web questionnaires, LFS being one of the surveys tested. The goal of the project was to design a user-friendly questionnaire, conduct a pilot survey and test for any arising mode-effects due to the differences in the data collection mode (CATI vs. CAWI). As the LFS indicators are closely monitored in the media, any unexpected mode effects would have harmful consequences on the credibility of the results. The aim was also to gain experience on data collection process to be used in further development of LFS mixed mode design.

In this paper we share our experiences on conducting LFS as CAWI based on qualitative cognitive interviewing and quantitative analysis of pilot survey data. We focus on the analysis of mode effects on working hours and employment status. While designing the questionnaire we were faced with several different options on formulating the CATI questions on the CAWI questionnaire. The results from our *cognitive interviews* showed that people tend to forget official holidays and other absences from work when answering to the question on actual working hours. However, this problem may be of less gravity in CATI where the interviewer can remind the respondent on occurrence of official holidays. Furthermore, we found that in CAWI people tended to invest more effort on recalling the working hours when the hours were asked separately for each day in a grid form.

In the pilot survey we used *experimental design* ("*split-plot*") to test whether the two different CAWI-layouts for the working hour questions produce different kinds of results. Two different types of implementing DK-option was also tested. After comparing the different CAWI question-layouts to each other, we evaluate which of the two layouts produce more comparable data with CATI results (ie. less mode-effects). The results are analysed using standard statistical methods such as chi-square and t-test. In addition, we provide insights on the qualitative feedback received from the CAWI respondents. Finally, we conclude the paper by sharing our experiences on mixed mode LFS and give some recommendations for designing questionnaires for mixed-mode design.

1 Introduction

As the response rates on CATI surveys are declining and the budgets are cut, possibilities on computer assisted web interviewing (CAWI) as explored with great expectations. In 2012, Statistics Finland launched a project on designing and testing of web questionnaires, LFS being one of the surveys tested. As the LFS indicators are closely monitored in the media, any unexpected mode effects would have harmful consequences on the credibility of the results. The goal of the project was to design a user-friendly questionnaire, conduct a pilot survey and test for any arising mode-effects due to the differences in the data collection mode (CATI vs. CAWI). The aim was also to gain experience on data collection process (response rate, time, reminders, motivating the respondents) to be used in further development of LFS mixed mode design.

The sampling frame for the Finnish Labour Force Survey is based on the Population Register on individuals living permanently in Finland. Hence, the sample is based on individuals, not households.

In this paper we share our experiences on conducting LFS as CAWI based on qualitative cognitive interviewing and quantitative analysis of pilot survey data. We focus on the analysis of mode effects on working hours and employment status.

2 Results from the LFS web project

The web-project was started with designing of the LFS questionnaire into CAWI-mode. We chose to use BlaiseIS-software since the CATI-questionnaire is also in Blaise-format and Blaise allows for more complicated routing.

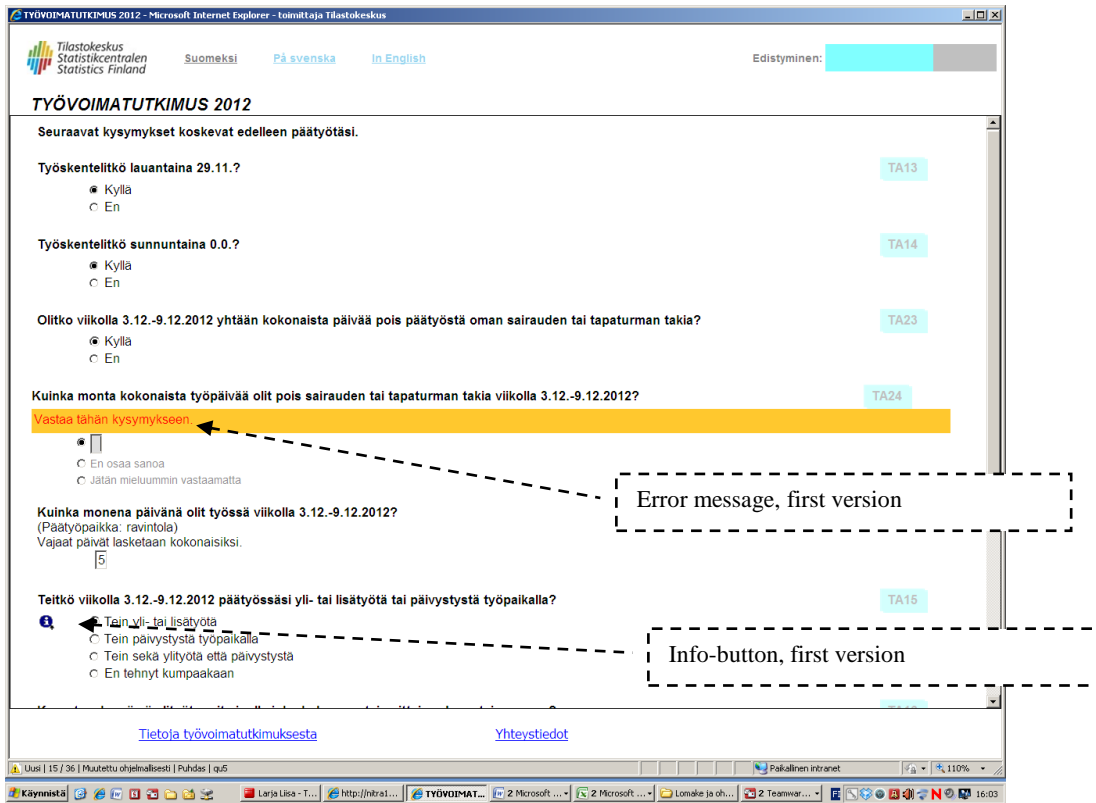
The questionnaire was tested iteratively using *cognitive interviewing* where the respondents filled in the questionnaire in the research laboratory and were at the same time encouraged to “think aloud” on how to proceed in the form and solve any occurring problems. The results of this phase have been reported in detail in elsewhere (Järvensivu, Kallio-Peltoniemi & Larja, forthcoming) but main findings are summarized in the chapter 2.1.

After completing the cognitive interviews and the questionnaire, we conducted a quantitative pilot study with a sample of 8 000 persons. The data was collected during October 2013. In the pilot survey we used *experimental design* (“*split-plot*”) to test whether the two different CAWI-layouts for the working hour questions (see 2.4) produce different kinds of results. Two different types of implementing Don’t know -option was also tested (see chapter 2.1).

2.1 Designing the questionnaire

The respondents were not satisfied with the **visual layout of the questionnaire** (produced with Blaise 4.8) but described it as “barren” or “ugly” (see Picture 1a) and hence we asked our graphic designer to produce a second version to be programmed, in which the respondents were satisfied with (see Picture 1b).

Picture 1a: First version of the LFS CAWI-questionnaire layout



Picture 1b: Final version of the LFS CAWI-questionnaire layout



We observed that if a question has any effect on the following questions, it should be presented on its own page. This is because (when using BlaiseIS) disappearing or appearing of the remaining questions in the page causes a small “flash” on the screen which confuses the respondents. Most respondents interpret the “flash” as a sign of changing the page and started reading the questions again from the start. Furthermore, the font was originally too small, and using bold in questions worked well. The navigation bars were programmed to always appear right below the questions, since some respondents did not find them if they were “hidden” in the bottom of the page. “The continue later” -bar created some problems when the respondents mixed that with “Continue” -bar (resulting in exit from the questionnaire) and the name was changed into “Save and continue later”. However, the paradata results of the quantitative pilot study suggest that this formulation too, has created some confusion and we will probably remove the button altogether.

Error messages were first shown in red font with yellow background (see Picture 1a), but this was changed as the respondents commented them as “clumsy” or “sloppy” and one of them anxiously exclaimed that “now the computer got mad!”. The final layout is displayed in the Picture 2 below with black font and light pink background with black frames.

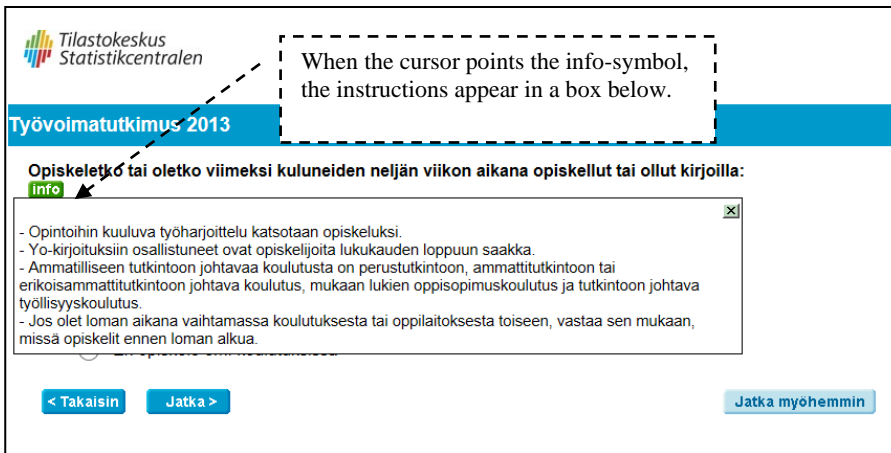
Picture 2: Error message final layout



The screenshot shows a survey interface for 'Työvoimatutkimus 2013'. At the top left is the logo for 'Tilastokeskus Statistiskcentralen'. Below it is a blue header bar with the text 'Työvoimatutkimus 2013'. The main content area has a light pink background and black text. It starts with a paragraph: 'Seuraavat kysymykset koskevat viikkoa 7.10.-13.10.2013 maanantaista sunnuntaihin (viikko nro 41, 7.10.-13.10.2013)'. Below this is a bolded question: 'Olitko viikolla 7.10.-13.10.2013 ansiotyössä vähintään yhden tunnin joko palkansaajana, yrittäjänä, ammatinharjoittajana tai maatilalla?'. Underneath the question is a light pink instruction box: 'Vastaa tähän kysymykseen tai valitse 'En osaa sanoa' -vaihtoehto'. To the left of the question is a green 'info' button. Below the question are three radio button options: 'Kyllä', 'En', and 'En osaa sanoa'. At the bottom of the screen are three buttons: '< Takaisin', 'Jatka >', and 'Jatka myöhemmin'.

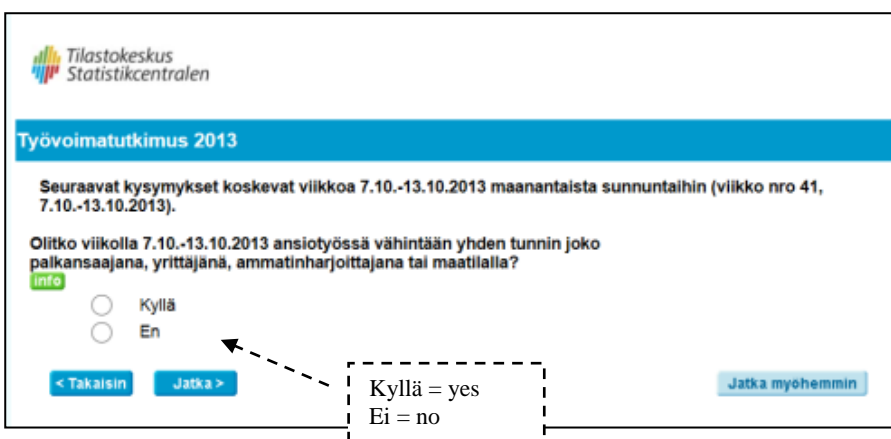
One of the most difficult questions in designing CAWI-questionnaire is **what to do with the instructions**. In CATI, the most frequently needed instructions are displayed in the screen, but the interviewer reads them only if there is confusion on the meaning of the question or the response options. We first tried to retain these same instructions in the CAWI-questionnaire, but this confused the respondents and increased response time. We then replaced the instructions behind an “info”-symbol, but the result was that hardly any of the respondents noticed this symbol (Picture 1a). Not even the instructions in the front page to use info-symbol helped. Finally we made the button bigger and changed the colour into green (see Picture 3) to make it stand out more clearly.

Picture 3: Info-button final layout

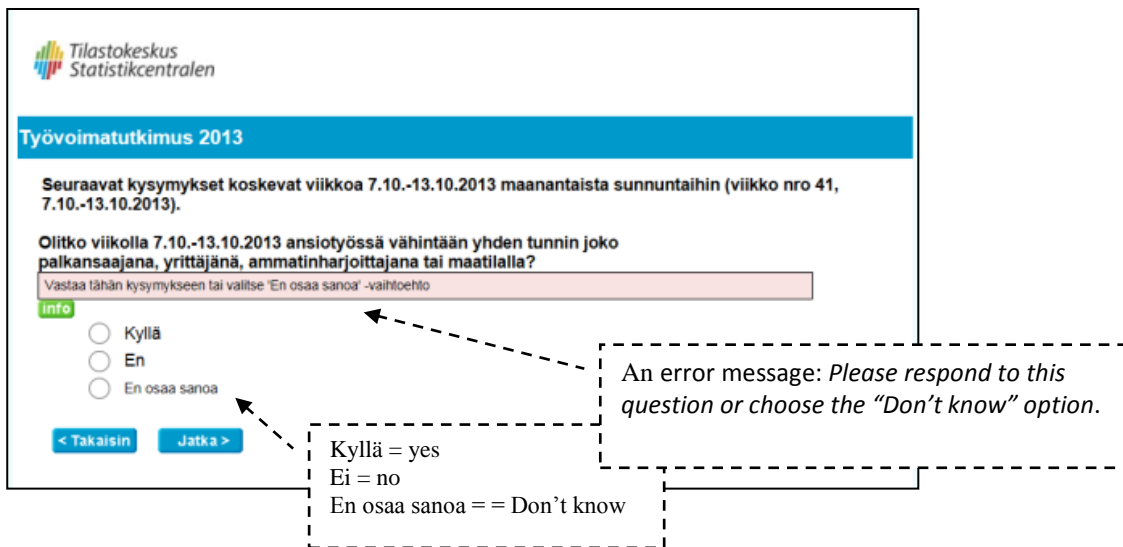


Finally, we were faced with how to proceed with presenting the **don't know -option** (DK). If DK is not provided and skipping of questions is not possible, there is a risk of the respondent quitting the questionnaire. However, if they are always available, we risk encouraging the respondent to satisficing. For the cognitive interviews we created a version where DK was not readily available (see Picture 4a), but only appeared after the respondent tried to skip the question (see Picture 4b). The design parallels the one used in CATI questionnaire, where the don't know -option is never read out, but the interviewer may use this code if the respondent is unable to choose one of the offered options. The solution seemed to work well, but in laboratory circumstances testing was difficult as hardly any of the respondents tried to skip a question or expressed unwillingness to give a response. Hence, we created a split plot design for the pilot survey where half of the sample (N=4000) were allocated a questionnaire with “hidden” DK (as in Picture 4a-b) and the other half (N=4000) received a questionnaire where DK was visible all along (Picture 4c). The results showed that having DK available by default (as in Picture 4c) produced significantly more DK-responses in most questions as compared to the version where DK was hidden (Picture 4a-b).

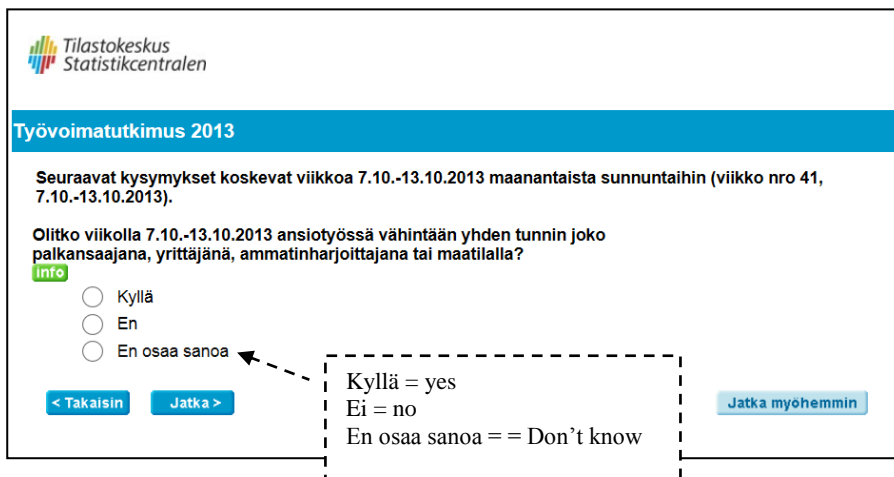
Picture 4a: DK is hidden at first sight (first version of the CAWI questionnaire)



Picture 4b: DK becomes visible as the respondent tries to move to the next page without responding to the first question (first version of the CAWI questionnaire)



Picture 4c: DK is visible by default (second version of the CAWI questionnaire)



2.2 Data collection process

The response rate to the survey was 30 %, which can be described as a rather decent outcome. We estimate, that although managing multi-mode data collection process is more expensive, adding CAWI would reduce total data collection costs in surveys like LFS, where the samples (and hence the amount of interviewing work) are large.

The most active participants were women and older age groups. The worst response rate was among young people aged 15-24, which was disappointing as it was anticipated that the youth would prefer CAWI mode. As compared to the CATI-respondents, it seems that CATI-data is less distorted than CAWI-data as CATI-interviewers seem to reach people of different age and with different educational background more evenly. To reduce the effects of non-response on the results the data was corrected by using weight calibration by area, gender, age and status in the job seeker register, as it is done also in the "normal" LFS in Finland.

One of the most obvious advantages of CAWI is the rate of data accumulation. In our study, too, 40 % of the responses were received within three days after receiving the cover letter. This means, that when designing the multi-mode process, it makes sense to preserve only a few days for web-

responding, after which the CATI-interviewers may start calling to those who have not yet responded.

To increase response rate, we used letters, SMS's and phone calls as reminders. The reminders were targeted to the most problematic groups (young, men, less educated) so it is difficult to conclude which method worked best since the target groups were not randomly selected. However, all reminders increased response rate and phone calls added even 18 % points.

The questionnaire took on average 16 minutes to complete, time increasing with age. Over 40 % of the respondents answered between 5-10 pm and almost 90 % responded from home. Most people used lap top (60 %) or desk top computer (35 %) and the rest tablets or smart phones.

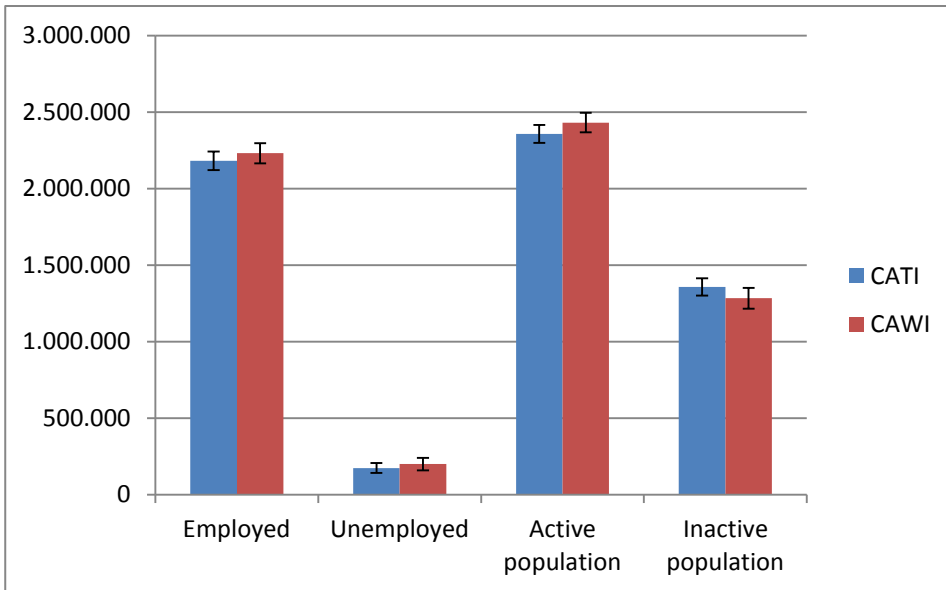
2.3 Mode-effects: Job search and employment status (WSTATOR)

Employment rate and unemployment rate are among the most closely followed economic indicators. Hence, any unexpected mode effects in the results due to the change in data collection mode would cause serious problems. As it can be observed from the Figure 1, **there were no statistically significant mode effects** on employment status between the results from CATI and CAWI, as the confidence intervals do overlap.

However, as these indicators are meticulously monitored, also statistically non-significant effects are of interest, especially because the sample size for the analysis is rather small (CATI = 2 570, CAWI = 2 345) and hence the confidence limits are larger in the pilot study data than in the official LFS data. Analysing the differences more closely it was observed that CAWI data produces higher estimates for the number of employed persons, and yields employment rate of 60,1 % as compared to the 58,7 % calculated from CATI-data. Also, the estimates for unemployed populations are higher based on CAWI-data, which gives unemployment rate of 8,2 % whereas CATI-data produces 7,4 %.

To explain why there are more employed and unemployed person and less inactive in persons CAWI than CATI, we analyzed more closely the response patterns to different questions constructing WSTATOR. It seems, that not-employed respondents in CAWI tend to answer more often that they have been looking for a job as compared to the respondents for the same question in CATI ($\chi^2 = 6,91, p = .01$). The effect is even larger for employed respondents ($\chi^2 = 57,20, p = .0001$). However, as the samples in CATI and CAWI are different, it is difficult to say whether the difference in the results is due to mode-effects or greater selection bias in CAWI. We have plans to analyze this further by adding more variables to weighting frame and trying to increase the similarity of CAWI and CATI data.

Figure 1: Estimates and CI (95 %) for employed, unemployed, active and inactive population in LFS data collected by CATI and CAWI during October 2013



2.4 Mode-effects: Working hours (HWACTUAL)

The results from our *cognitive interviews* showed that people tend to forget official holidays and other absences from work when answering to the question on actual working hours. However, this problem may be of less gravity in CATI where the interviewer can remind the respondent on occurrence of official holidays. Furthermore, we found that in CAWI people tended to invest more effort on recalling the working hours when the hours were asked separately for each day in a grid form. Hence, we constructed an experimental design, where half of the respondents in the CAWI pilot survey were given a grid form -question (see Picture 5b) and the other half was allocated a CATI-style question (see Picture 5a).

Picture 5a: Layout for the “one question” -format HWACTUAL

Tilastokeskus
Statistikcentralen

Työvoimatutkimus 2013

Montako tuntia työskentelit päätyössäsi viikolla 7.10.-13.10.2013 ?

- Jos et muista tarkkaa aikaa, arvio riittää.
- Laske mukaan mahdolliset ylityöt, kotona tehty ansiotyö ja päivystys työpaikalla.
- Älä merkitse ruokataukoja, tilapäisiä tunninkin poissaoloja sekä sairaus- ja lomapäivien tunteja.
- Älä merkitse kotona tapahtuvaa päivystystä tai kotitaloustyötä.

tuntia minuuttia

____ hours ____ minutes

< Takaisin Jatka > Jatka myöhemmin

“How many hours did you work in your main job during the week 7.10.-13.10.2013?”

Picture 5b: Layout for the “grid”-format HWACTUAL

Tilastokeskus
Statistikcentralen

Työvoimatutkimus 2013

Montako tuntia työskentelit viikolla 7.10.-13.10.2013 (7.10.-13.10.2013)? ←

- Kirjoita taulukkoon tekemiesi työtuntien määrä. Jos et muista tarkkaa aikaa, arvio riittää.
- Laski mukaan mahdolliset ylityöt, kotona tehty ansiotyö ja päivystys työpaikalla.
- Älä merkitse ruokataukoja, tilapäisiä tunninkin poissaoloja sekä sairaus- ja lomapäivien tunteja.
- Älä merkitse kotona tapahtuvaa päivystystä tai kotitaloustyötä.

← “How many hours did you work in your main job during the week 7.10.-13.10.2013?”

	Päivämäärä	Tunnit	Minuutit
maanantai	7.10.2013	5	5
tiistai	8.10.2013	6	33
keskiviikko	9.10.2013	5	
torstai	10.10.2013	3	56
perjantai	11.10.2013	4	
lauantai	12.10.2013		
sunnuntai	13.10.2013		
Yhteensä		24	34

←

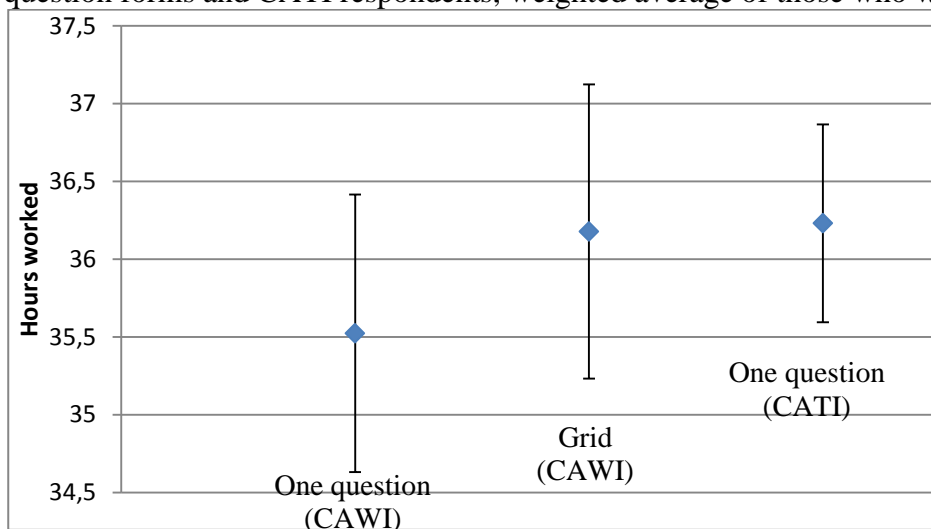
	Date	Hours	Minutes
Monday	7.10.2013	—	—
Tuesday	8.10.2013	—	—
Wednesday	9.10.2013	—	—
Thursday	10.10.2013	—	—
Friday	11.10.2013	—	—
Saturday	12.10.2013	—	—
Sunday	13.10.2013	—	—
Total		—	—

< Takaisin Jatka >

The results from the split plot experiment showed that the CAWI respondents who answered to the traditional, CATI-style question (Picture 5a) reported on average less hours (35,5 hours, $n = 703$) than the respondents faced with grid-form question (Picture 5b) (36,2 hours, $n = 626$). This was surprising, as it was anticipated that grid-form would decrease reported working hours as people would better remember their absences. However, the difference was not statistically significant (see Figure 2). The CATI-respondents who answered a question similar to Picture 5a, had worked on average exactly the same number of hours (36,2 hours, $n = 1316$) as CAWI respondents in grid form condition. The result points towards functional equivalency, meaning that we might need different kind of question format to different modes in order to produce similar results. Although the difference between the average in CATI data and one question -condition in CAWI was not statistically significant (see Figure 2)¹, we decided to opt for the grid form question to ensure better functional equivalency of results from CAWI and CATI. It was observed from the paradata that completing the grid form question took on average three times as long as the one question -condition, but it was interpreted also as a sign of people investing in memorisation and hence better quality.

¹ In addition, we performed the same analysis using larger data set for CATI condition ($N = 4\,706$). The increase of sample size was reflected in the narrowing of confidence intervals and the difference between CATI and one question condition in CAWI was marginally significant ($F = 1.02$, $p = .69$, $t = -1.77$, $p = .08$). Hence, it is probable that with normal size LFS data the difference would be statistically significant.

Figure 2: Hours worked during the reference week among CAWI respondents to two different question forms and CATI respondents, weighted average of those who worked 1-98 hours and CI.



2.5 Qualitative feedback

At the end of the WEB-questionnaire, respondents had a chance to give informal comments on the survey. A lot of comments were given and they were analysed based on the content of the text. No background information was added to them.

First of all, many of the employed respondents were irritated when the exact information on their post was demanded. They wondered where all these items (name, address, municipality) are exactly needed. Also a group of people with untypical employment, e.g. those who are working under many contracts without main or a side line job, or working as a scholarship recipient, told they had difficulties giving some answers.

According the feedback, pensioners often did not clearly understand that the aim of the LFS is to provide labour market data concerning the reference week, not data concerning the situation in life. Some of them experienced that there were too much questions, some of them that questions were too few or questions didn't fit. Correspondingly some entrepreneurs experienced all kind of entrepreneurial activity is not taken into account in the inquiry.

Finally in the feedback, the measurement of the working time was commented occasionally as a difficult item, if working hours vary much or working operations are discontinuous.

We see the direct information from the respondents extremely valuable. It is possible to build up better working instructions at the questionnaire when the feedback from different groups is at disposal.

3 Conclusions and lessons learnt

The experience from the CAWI pilot study encourage us to keep on developing LFS mixed-mode design. Response rates as high as 30 % retain potential for increasing efficiency in data collection and the speed of data accumulation might increase the timeliness of statistics.

The lessons learnt include that visual design of the questionnaire is important in order to assure fluent navigation on the questionnaire. Although more laborious, this was feasible also with BlaiseIS. We learned that fewer questions per page is better than more, the instructions should be hidden and the DK option should appear only after trying to skip a question.

As the CAWI data was more distorted than CATI-data, where non-response was divided more evenly, CATI-interviewers are needed in order to reach the more difficult targets. The qualitative feedback suggests that more personalized cover letters to different groups (young, retired, entrepreneurs) might also be necessary.

Hence, there is still a long way to go before commissioning of mixed mode LFS. Although the analyzed mode effects on employment status and working time were not statistically significant, these should be studied more as LFS results are very closely followed and the limited sample size of test data does not have the same statistical power as normal LFS data. For working hours we recommend grid form question in order to achieve better functional equivalency. The reason for differences in employment status may be due to either mode-effects or selection bias, which should be studied further.

4 References

Järvensivu, M.; Kallio-Peltoniemi, M.; Larja, L. (forthcoming). *The ESSNet project on Data Collection for Social Surveys using Multiple Modes. The pre-testing report of Statistics Finland.* Draft 28 April 2014.