**COOPERATION ON MULTI-MODE DATA COLLECTION (MMDC)**

**MIXED MODE DESIGNS FOR SOCIAL SURVEYS - MIMOD**

GRANT AGREEMENT FOR AN ACTION WITH MULTIPLE BENEFICIARIES

AGREEMENT NUMBER – 07112.2017.010-2017.786

WP5 – Deliverable 4

# Final methodological report presenting results of usability tests on selected ESS surveys and Census

*Smartphone fitness of ESS surveys – case studies on the ICT survey and the LFS*

Date: 20 February 2019

Dag Gravem (SSB)
Vivian Meertens (CBS)
Annemieke Luiten (CBS)
Deirdre Giesen (CBS)
Nina Berg (SSB)
Jeldrik Bakker (CBS)
Barry Schouten (CBS)

WP5: Challenges for phone and tablet respondents within CAWI

*SUMMARY:*

*WP5 of the MIMOD project investigates the employment of mobile devices in ESS surveys. In particular, it explores fitness of ESS surveys for smartphones and it explores the utility of mobile device sensors to replace and/or supplement survey data. In deliverable 1 (Schouten et al. 2018) a set of fitness criteria was proposed and applied to four ESS surveys (EHIS, ICT, LFS, SILC). Two of the surveys, the ICT survey and the Labour Force Survey on the person level were suggested as potentially fit for smartphones with relatively modest revisions. Deliverable 2 consisted of questionnaires for the ICT and LFS, partially optimized for smartphones. In this deliverable 4, we describe and discuss the results of qualitative tests at CBS (the Netherlands) and SSB (Norway) into the fitness of these questionnaires. The tests considered both comprehension and usability. The main findings reflect on the fitness criteria dimensions screen size, touch navigation and duration.*

*Detailed recommendations on questionnaire design for mobile devices are described in chapter 6. Next to these, we recommend two general topics for future discussion within the ESS: mobile device first questionnaire design and questionnaire length of ESS surveys. We view these topics as beyond the scope of the current WP, but they naturally arise from an assessment of fitness for mobile devices, in particular smartphones.*

*We are advocates of a mobile device first questionnaire design, or, at the least, of a rigorous account of the mobile device option in questionnaire design. We have two main reasons for this. The first reason is that smartphones have become a dominant communication channel and cannot be ignored in design. The second reason is that issues with usability and comprehension on smartphones reveal the measurement error prone questions and question blocks. Such a viewpoint, however, has implications for ESS model questionnaires and ESS survey guidelines. Multi-device surveys introduce additional challenges for the questionnaire design.*

*An obstacle that is often put forward to the introduction of smartphones is questionnaire length. As most ESS surveys are longer and, consequently, demanding when filling in on a smartphone, it is imperative to prevent 'speed' and 'stimulate' a relaxed manner leading to better quality and less measurement error. A responsive design should facilitate the respondent filling in on a smartphone screen, however, most questions and answer texts are cognitively demanding due to their specific content or response task, as for example long reference periods. When reflecting on the fitness criteria and the experiences from the test interviews at CBS and SSB it is a harsh job to find a good modus in redesigning, i.e., responsive design and collecting valid and reliable data comparable over devices and modes (also interviewer based). This leads to a question for future discussion: Is it feasible to find possibilities to shorten/redesigning a ESS model questionnaire making it user friendly to fill in on a smartphone?*

*We recommend that these topics are addressed in both general discussion on ESS procedures and in specific working groups for ESS surveys.*

# 1. Introduction

This deliverable is part of WP5 of the MIMOD project, which is concerned with the use of mobile devices, in particular smartphones, for ESS surveys. This fourth WP5 deliverable explores the fitness of two surveys for smartphones based on qualitative tests with test persons at CBS and at SSB.

At the start of MIMOD, in December 2017 and January 2018, a list of ESS social surveys to be investigated in the project was prepared. From it four surveys were selected: the European Health Interview Survey (EHIS), the ICT survey, the Labour Force Survey (LFS), and the European Survey on Income and Living Conditions (SILC). The ESS has other social surveys, but these are either subject of existing innovation projects/task forces (Household Budget Survey and Time Use Survey) or have a specific target population (Adult Education Survey). For the four selected surveys both ESTAT model questionnaires and/or guidelines and country-specific implementations in Germany, Netherlands and Norway were collected for exploration.

Deliverable 1 of WP5 proposes a set of evaluation criteria for smartphone fitness (Schouten et al. 2018). These criteria are linked to three dimensions: screen size, touch navigation and interview duration. The three dimensions are considered to be crucial in the assessment of fitness of questionnaires for smartphones. We must note, however, that many ESS countries are not (yet) using online questionnaires, so that an assessment can have two starting viewpoints. One viewpoint is an assessment of smartphone self-administered questionnaires relative to an interviewer-assisted questionnaire and another viewpoint is an assessment relative to self-administered online questionnaires on traditional devices (desktops and laptops). In our evaluation, we focus on the second viewpoint, i.e. we devote relatively little attention to differences to surveys with interviewers. Obviously, the redesign to a smartphone questionnaire may be much more extensive for surveys that do not yet have a self-administered online implementation. This has to be kept in mind when reading this deliverable. WP4 of the MIMOD project focuses on mixed mode data collection in a broader perspective, including interviewer-administered modes. Deliverable 3 of WP4 of the MIMOD (Gravem & Berg 2019) explicitly discusses strategies for questionnaire development for both interviewer-administered modes and self-administered modes.

The criteria have been evaluated for the four selected surveys. Deliverable 1 presents the scores as well as a discussion and recommendations. The LFS is a household survey, which can become quite long if data have to be collected on several household members. In the MIMOD survey on mixed-mode experiences and practices among ESS National Statistical Institutions (NSIs) (Gravem et al. 2018) the first wave of the LFS was not considered suitable for CAWI data collection by 20 out of 31 participating NSIs.Tthe fitness criteria have been applied to both a household version and a person version of the LFS. The person version is much shorter than the household version and is under consideration at CBS and SSB. SILC has a household component as well. The first wave of SILC is also considered not suitable for CAWI by a majority of NSIs (ibid). SILC however assumes one main reference person and cannot be shortened much by moving to a person version. In deliverable 1, we argue that the person LFS and the ICT survey may be made fit for smartphones with relatively little effort. However, even the effort to migrate ICT and LFS to a smartphone implementation must not be underestimated. In the remainder of the WP5work, these two surveys are selected as case studies. The LFS was already subject of research and pilots at CBS and SSB, and a smartphone option for the ICT survey is explored at SSB. As a consequence, the optimization of the questionnaires for smartphones could be combined with these running projects. Please note: with optimization we mean that the questionnaires have been profoundly adapted to facilitate smartphone friendly data collection. Optimization does not mean we have created the most optimal questionnaire for collecting data on these topics via a smartphone. Questionnaire development was limited by the data requirements and considerations regarding the comparability with other and previous data collections (on other devices and via other modes).

In the next section, we revisit the relevant outcomes of deliverable 1. Section 3 describes the questionnaire designs for the questionnaires of the LFS and ICT as they are tested by CBS and SSB. This section will also discuss the specific choices made in adapting these questionnaires for smartphones. Section 4 describes the methodology and the findings of the tests conducted at SSB. Section 5 describes the methodology and findings of the tests conducted at CBS. Finally, in section 6, the findings are discussed and conclusions and recommendations based on all tests presented.

## 2. Background

In WP5 deliverable 1, the LFS and ICT were scored on 16 criteria. The results are given in tables 1 and 2, for the ICT and LFS, respectively. The ICT survey scores good on both the navigation and duration dimensions for the model questionnaire. The screen size dimension is problematic due to the large number of instructions, introductions and long questions/answers. The LFS turned out to be problematic on the screen size dimension; many questions require long texts. The navigation dimension is somewhat problematic due to open questions. The duration dimension is problematic for the household version of the LFS. On the person level, i.e. persons answering only questions that apply to themselves, the LFS may be doable. It must, however, be made clear that country-specific implementations of the LFS vary widely in length. A person level LFS following the model questionnaire/guidelines is doable in terms of duration. Surveys differ in their general enjoyment-relevance-burden scores to respondents. For online surveys, response rates may vary from 15% to 45% with the exact same data collection strategy in terms of invitation and reminder letters, text messages or emails. Such large differences express the perceived enjoyment-relevance-burden ratio to the general population. For smartphones, surveys that score weaker are at larger risk. Table 3 shows the overall scores of the ICT and LFS on the three dimensions. See also WP 5 deliverable 1 for details (Schouten et al. 2018).

*Table 1: Score on fitness criteria for ICT*

| *Dimension* | *Criterion* | *Operationalization* | *Scores* | | | |
|---|---|---|---|---|---|---|
| | | | *Model* | *CBS* | *DESTATIS* | *SSB* |
| Screen size | Introductions | Number of items with introductions | 5 | 26 | 17 | 12 |
| | Instructions | Number of items with instructions included | 4 | 4 | 16 | 52 |
| | Grids | Number of grids | 1 | 23 | 0 | 0 |
| | #Items per grid | Average number of items per grid | 6 | 2 | NA | NA |
| | Question text | Number of items with > 20 words (excluding introduction text) | 13 | 39 | 26 | 50 |
| | # answer cat’s | Number of items with > 5 answer categories | 7 | 13 | 10 | 0 |
| | Answer text | Number of items with > 10 words in at least one category | 3 | 16 | 5 | 10 |
| Touch navigation | Open question | Number of open questions | 0 | 5 | 9 | 1 |
| | Many answers | Number of items with > 25 answer categories | 0 | 0 | 0 | No |

| *Dimension* | *Criterion* | *Operationalization* | *Scores* | | | |
|---|---|---|---|---|---|---|
| | | | *Model* | *CBS* | *DESTATIS* | *SSB* |
| Duration | # of items | Total number of items | 39 | 140 | 61[1] | 110 |
| | Av duration | Average duration of survey per respondent | NA | 23 min | NA | 12 min |
| | Household | Is survey a household survey? Yes/no | No | No | No | No |
| | Database | Does survey require interaction with a database? Yes/no | No | No | 2 | No |
| | Cognitive burden | Number of (anticipated) items that require calculations by an average respondent, i.e. are cognitively burdensome | 7 | 7 | 1 | 7 |
| | Consultation | Number of (anticipated) items that require consultation of personal documentation by an average respondent | 3 | 4 | 2 | 1 |
| | Enj-Rel-Bur | Response rate to traditional online devices | NA | 33.1% (web)<br>36.9% (CATI) | NA | NA |

*Table 2: Score on fitness criteria for model questionnaire LFS (2016)*

| *Dimension* | *Criterion* | *Scores* | | | | |
|---|---|---|---|---|---|---|
| | | *Model* | *SSB* | | | *CBS* |
| | | | *Employee* | *Unemployed* | *Student* | |
| Screen size | Introductions | 0 | 3 | 0 | 1 | 13 |
| | Instructions | 7 | 14 | 5 | 18 | 42 |
| | Grids | 0 | 0 | 0 | 0 | 1 |
| | #Items per grid | NA | NA | NA | NA | 10 |
| | Question text | 1 | 2 | 1 | 2 | 76 |
| | # answer cat's | 18 | 1 | 4 | 3 | 41 |
| | Answer text | 14 | 1 | 0 | 0 | 4 |
| Touch navigation | Open question | 4 | 5 | 0 | 4 | 70 |
| | Many answers | 0 | 0 | 0 | 0 | 16 |
| Duration | # of items | 85 | 33 | 21 | 48 | 346 |
| | Av duration | NA | NA | NA | NA | 27 min |
| | Household | Yes | Yes | Yes | Yes | Yes |
| | Database | NA | No | No | No | Yes |
| | Cogntv burden | 5 | 0 | 0 | 0 | 21 |
| | Consultation | 2 | 0 | 0 | 0 | 0 |
| | Enj-Rel-Bur | NA | NA | NA | NA | 22% (web)<br>54% (overall) |

[1] DESTATIS also has 8 household items. These are currently not included in the assessment.

*Table 3: Scores on the three dimensions screen size, navigation and duration for each survey. The LFS is also assessed for the person level.*

| *Survey* | *Screen size* | *Touch navigation* | *Duration* |
|---|---|---|---|
| ICT | 🟥 | 🟩 | 🟩 |
| LFS household | 🟥 | 🟨 | 🟥 |
| LFS person | 🟥 | 🟨 | 🟩 |

# 3. Design of responsive questionnaires for the ICT and LFS

For the two selected surveys, ICT and LFS, smartphone questionnaires were implemented, tested and piloted by CBS and SSB. Additionally (related to other projects), at CBS a household roster and a series of grid questions were developed and tested for multi-device data collection.

Adapting a survey questionnaire for smartphones is more than implementing screen size responsiveness. It implies question rewording, revising the use of introductions, breaking up grids of questions, and potentially also shortening the survey as a whole.

In the following two subsections, we give separate accounts of the questionnaires developed at CBS and SSB.

## 3.1 SSB ICT survey

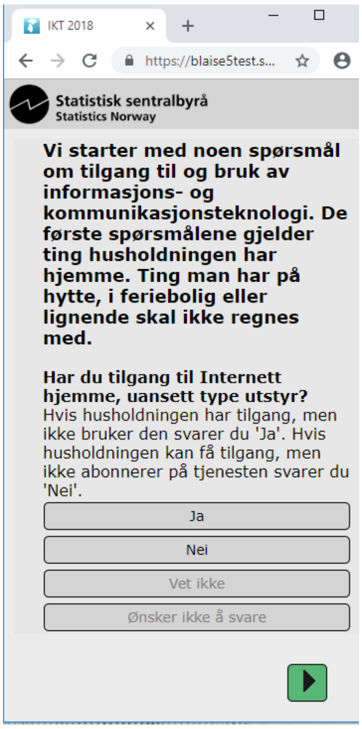
Statics Norway tested a version of the *ICT survey*. This survey is currently CATI only and embedded in a national omnibus survey, but there are plans for conducting a CAWI pilot with the ICT survey as a standalone survey. The test questionnaire is not new, but based on the existing CATI questionnaire, which in its turn is closely based on the ESS model questionnaire. The design was responsive, with layout automatically adjusting to screen size on device (PC/tablets/smartphone).

Some of the wording in the original CATI questionnaire was changed for the CAWI test questionnaire to adapt to self-completion, changing “the respondent” to “you” etc. Introductory texts were presented in larger, bold font, and questions in regular size bold fonts. Clarification and instructions for interviewers were slightly modified and presented in non-bold fonts, as seen in figure 3.1. Response options were presented vertically under each question. In addition to the substantial response options, “Don’t know” and “Do not wish to answer” was available on all questions in a grey font to distinguish them and keep the focus on the substantial response categories.

The main change from the CATI version was that “mark all that applies” questions were converted into batteries of Yes/No questions. On mobile only one question was displayed per screen. In question batteries, the question stem and ending were in bold fonts for first question/screen. For the following questions in the battery the stem question was in non-bold, while the new sub question was in bold (see e.g. figure 4.2 question R2a) .

In the responsive design for mobile, finger-friendly vertical buttons were used. When selected, a thicker black frame was added to the response option. When responding to a question with only one response option, respondents were automatically forwarded to the next question. On questions with possibly more than one answer, of which there was only a handful, the navigation buttons had to be used. If the test person chose “Don’t know” or “Do not wish to answer”, the navigation buttons had to be used regardless of whether one or more than one substantial response was allowed.

*Figure 3.1. Mobile lay-out ICT test questionnaire (A1)*



**Vi starter med noen spørsmål om tilgang til og bruk av informasjons- og kommunikasjonsteknologi. De første spørsmålene gjelder ting husholdningen har hjemme. Ting man har på hytte, i feriebolig eller lignende skal ikke regnes med.**

**Har du tilgang til Internett hjemme, uansett type utstyr?**
Hvis husholdningen har tilgang, men ikke bruker den svarer du 'Ja'. Hvis husholdningen kan få tilgang, men ikke abonnerer på tjenesten svarer du 'Nei'.

The test questionnaire was created in Blaise 5 using smartphone style sheets developed at Statistics Netherlands and cross-platform (Android, iOS) application development. Although some parameters were adjustable, and the questionnaire was designed using Statistics Norway's design principles, the functionality and layout is to a certain extent defined by what Blaise 5 allows for. Blaise 5 change requests must be directed at Statistics Netherlands.

## 3.2 CBS LFS smartphone survey

At CBS, we developed a new designed LFS questionnaire in an online design for an individual sample approach. This LFS questionnaire is under redesign due to a new (not final) regulation of input harmonisation for measuring employment and unemployment. This new LFS model questionnaire is designed with a responsive design meaning implementation of style sheets optimized for a smartphone developed at CBS and built in Blaise 5 for cross-platform (Android, iOS) application development. Several choices have been made in adapting the questionnaire for smartphones:

In general, at CBS the design choice on smartphone is one single question on a screen. Maximum of two (interrelated) questions at one screen if necessary. The formulation of the question and answer question text is shortened as much as possible. Text like 'next, some questions on' or 'the following question is about'  is skipped. Also, due to an individual approach, the proxy element in the question design like referring to a household member (e.g. asking about respondent's mothers/fathers work) is left out from the question texts.

In general minimum level of scrolling is allowed, meaning no horizontal scrolling and sometimes vertically scrolling as with the grid questions, see section 3.5.

As the number of answer categories and the content determines the screen size of the item. On smartphones, items may be split into multiple items by introducing a hierarchy in the answer categories.

Instructions texts are limited and included on the same screen. No buttons with instruction texts or help function are used in the visual design. This is comparable with the original design of the LFS questionnaire at CBS, where in the designed stylesheets for online questionnaires used for household and person surveys, no clickable buttons for additional information or instructions are incorporated. Especially for a smartphone design, the drawback of these buttons is the chance that the instruction text overlaps the original question or answer text which is not user/respondent friendly and might also be a potential risk for measurement errors.

Interaction with classification database: Survey items with many and diverse answer categories, e.g. occupation, educational level or type of economic activity, often employ interaction with a classification database positioned at the server of the survey institute. Such interaction requires internet traffic and for smartphones may slow down the interview speed. Such interaction with a database is not included in the CBS design of the LFS pilot questionnaire. For the question on occupation, in the LFS a smartphone layout is developed considering specific instructions to respondent who has to type text in an open text field and elaborate as much as possible on describing the name of occupation and working activities. In this LFS pilot a research goal is to see whether this yields the same quality of data as the PC internet questionnaire used to code occupations in the Netherlands and international with CASCOT.

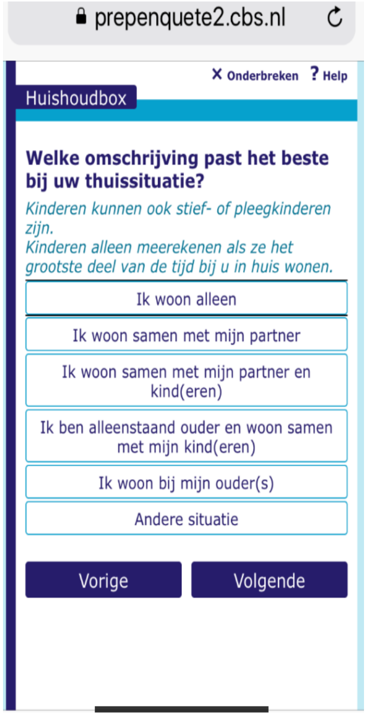
The household approach is not applied as this would make the duration longer as information is needed for each household member (directly or by proxy reporting) and this is not preferable for a smartphone questionnaire. For the LFS pilot the original household sample design switched to an individual sample approach. As this pilot covers the first wave of the LFS, in the second wave household information will be collected using the household roster. This was also designed and optimized for a smartphone and tested, see section 3.4.

The LFS questionnaire developed and tested by CBS includes questions on (un)employment status, job search activities, occupation features and a new developed question on measurement of educational level. In case the respondent is employed, in this questionnaire the focus is on the main job's features like working hours and occupation. In case of second (or more) jobs, questions are not included in this questionnaire for the first wave but in the second wave questionnaire and herewith deviates from length and number of questions of the original LFS questionnaire. Also other variables like questions on overwork and compensation for extra hours are not included for this wave. Consequently, the questionnaire is shortened as some questions are left out in this version for the first wave. This LFS questionnaire was fielded in a sample of the general population aged 15-74 years as a pilot at CBS in November 2018.

## 3.3 CBS household roster

The household roster was developed for mixed mode data collection, including data collection on smartphones. The roster collects information on the features of the household composition and the interrelations between respondent and other household members. Also age and gender is collected. The original version was designed like a table lay-out and needed adaptation for smartphone in visual elements, like using more screens, i.e., splitting up elements and also adaptations with respect to the content (See Figure 3.2 and 3.3). Demographic and societal developments like numbers of people identifying themselves as gender neutral, people living with own and partner's children together and households type like co-parenting initiated a redesign of the household roster.

*Figure 3.2 Responsive design household roster question on household composition*



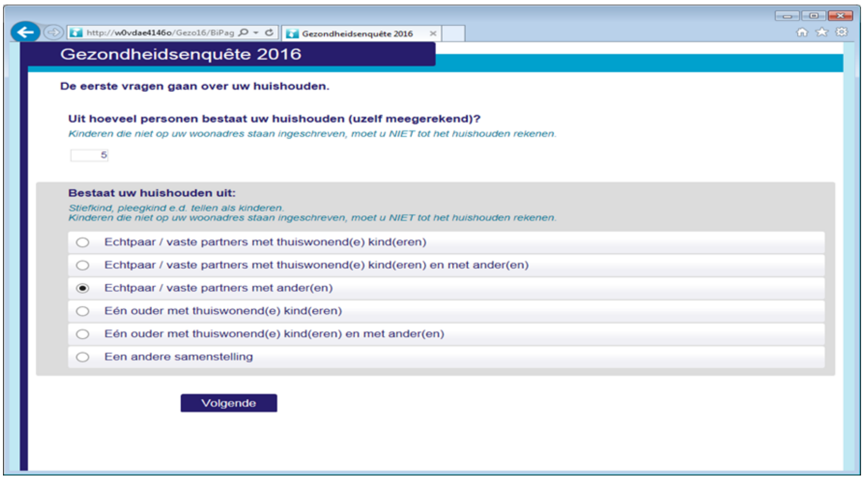
**Welke omschrijving past het beste bij uw thuissituatie?**

*Kinderen kunnen ook stief- of pleegkinderen zijn.*
*Kinderen alleen meerekenen als ze het grootste deel van de tijd bij u in huis wonen.*

- Ik woon alleen
- Ik woon samen met mijn partner
- Ik woon samen met mijn partner en kind(eren)
- Ik ben alleenstaand ouder en woon samen met mijn kind(eren)
- Ik woon bij mijn ouder(s)
- Andere situatie

*Explanation content screen: Question (in dark blue) asks which description fits best to the situation of the respondent. Instruction (in  italic light blue) to include stepchildren and foster children and to only include children  living the household most of the time. Answer options (on buttons): I live alone, I live with my partner, I live with my partner and child(ren), I am a single parent living with my child(ren), I live with my parents, Other.*

*Figure 3.3: Original household roster question (longer answer question texts)*



**De eerste vragen gaan over uw huishouden.**

**Uit hoeveel personen bestaat uw huishouden (uzelf meegerekend)?**
*Kinderen die niet op uw woonadres staan ingeschreven, moet u NIET tot het huishouden rekenen.*

5

**Bestaat uw huishouden uit:**
*Stiefkind, pleegkind e.d. tellen als kinderen.*
*Kinderen die niet op uw woonadres staan ingeschreven, moet u NIET tot het huishouden rekenen.*

○ Echtpaar / vaste partners met thuiswonend(e) kind(eren)
○ Echtpaar / vaste partners met thuiswonend(e) kind(eren) en met ander(en)
● Echtpaar / vaste partners met ander(en)
○ Eén ouder met thuiswonend(e) kind(eren)
○ Eén ouder met thuiswonend(e) kind(eren) en met ander(en)
○ Een andere samenstelling

*Explanation: Question one asking about how many people in household. Question two, conditional on number of people in household, asks about household composition. Answer options presented here: Couple with children living at home, Couple with children living at home and others; Couple with others; Single parent with children living at home; Single parent with children living at home and others; Other household composition. In blue italic: instructions about excluding children not living at home and including step- and foster children.*

This redesigned household roster was designed and built in Blaise 5 and tested for usability and comprehensibility. Some relevant findings with respect to fitness criteria are included in section 5.

### 3.4 CBS grid questions

The classic presentation of grid question as seen on paper and on the larger screens of PCs and tablets is not easily transferable to the smaller smartphone screens. CBS has developed and tested four possibilities for presenting grid questions on smartphones:

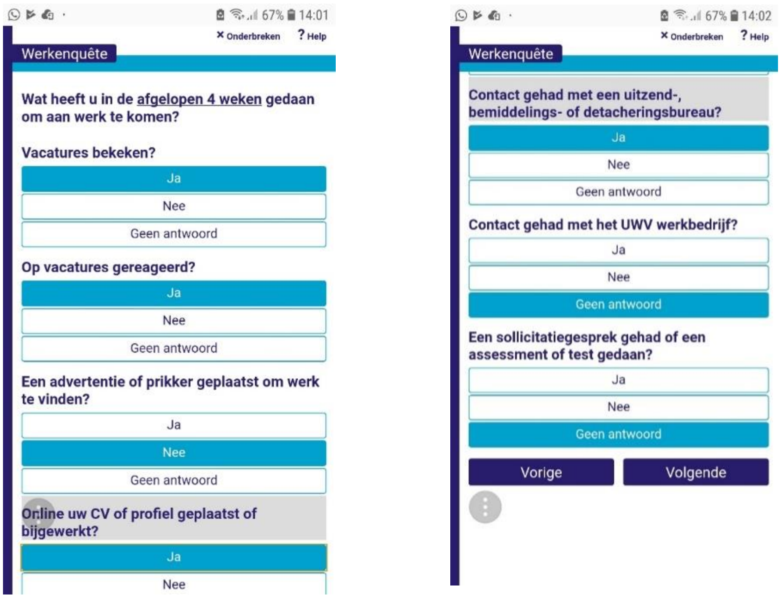
1. Stem of the question and items on one scrollable page – stem not fixed – this was the option as implemented for CBS LFS smartphone version (see figure 3.4).
2. Stem fixed (i.e. always visible), all items on one scrollable page – scroll by respondent – see figure 3.5
3. Stem fixed, all items on one scrollable page - autoscroll; looks the same as the option 2, but after selecting an answer the screen automatically scrolls to the next item. Unfortunately, due to technical problems the scrolling went so quickly it was hardly visible. For this test it was not possible to manipulate the speed of scrolling.
4. Paging - each item presented on a separate page, stem of question repeated on each screen (see figure 3.6). As with all questions, respondents can navigate back to previous items.

These options were compared to the classic presentation of grids on a large screen, as for example shown in figure 3.7.

The idea behind the first three grid-options is that it may be helpful for the response process if respondents can easily see their answers to the other items of the grid. This may also lead to a more comparable context as provided by the classic grid design on a large screen and hence reduce device effects. The problem with option 1 is that the stem of the question, which may contain crucial information, is not always visible when answering items. This may lead to respondents not using all information from the question stem (e.g. a reference period) when answering the question. Format 3, if implemented better, might be more user-friendly as it requires less scrolling. Format 4 is more consistent with the rest of the questionnaire and for that reason may be more user-friendly. Please note: there are other promising ways of presenting grid questions on smartphones, for example the so called “carousel” format. However, within the current technical possibilities and the time frame for this project, it was not possible to develop other options.

In the CBS LFS developed and tested for this project, there is only one grid question, a question about ways to look for work. For the purpose of testing grid questions we added in our test a special “grid questionnaire”. For this questionnaire we selected a variety of grid items. Specifically, we included an item with a very long introduction (which forced us to put the introduction on a separate screen prior to the question for smartphones) and two questions for which it was likely that respondents would anchor their answers relatively to answers given on previous items (two questions about how various actors relate to environmental issues). See table 4 for an overview of all grid questions tested by CBS.

*Figure 3.4: Scrollable version of LFS looking for work question for smartphone.*



Werkenquête

× Onderbreken ? Help

Wat heeft u in de afgelopen 4 weken gedaan om aan werk te komen?

Vacatures bekeken?
[✓ Ja] Nee Geen antwoord

Op vacatures gereageerd?
[✓ Ja] Nee Geen antwoord

Een advertentie of prikker geplaatst om werk te vinden?
Ja [✓ Nee] Geen antwoord

Online uw CV of profiel geplaatst of bijgewerkt?
[✓ Ja] Nee

Werkenquête

× Onderbreken ? Help

Contact gehad met een uitzend-, bemiddelings- of detacheringsbureau?
[✓ Ja] Nee Geen antwoord

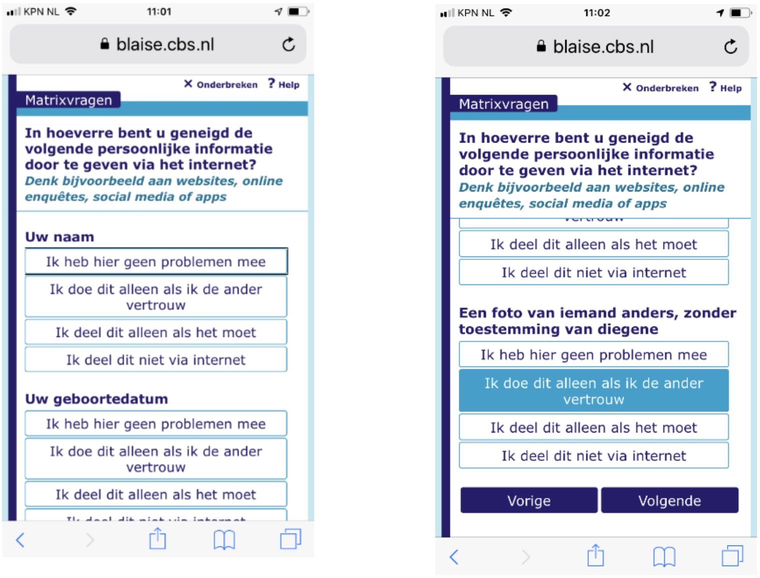
Contact gehad met het UWV werkbedrijf?
Ja Nee [✓ Geen antwoord]

Een sollicitatiegesprek gehad of een assessment of test gedaan?
Ja Nee [✓ Geen antwoord]

Vorige Volgende

*Note: stem of question is not fixed, so not visible after scrolling. If an answer option is selected, the selected option turns blue and the related question gets a grey background. This grey background moves to the next selected question.*

*Figure 3.5: Presentation of question on inclination to share info via internet on smartphone for both option 2 (stemfix-scrolling by respondent) and option 3 (stemfix-autoscroll).On the left screen on opening, on the right screen after scrolling to last item in the list.*



blaise.cbs.nl

× Onderbreken ? Help

Matrixvragen

**In hoeverre bent u geneigd de volgende persoonlijke informatie door te geven via het internet?**
*Denk bijvoorbeeld aan websites, online enquêtes, social media of apps*

**Uw naam**
Ik heb hier geen problemen mee
Ik doe dit alleen als ik de ander vertrouw
Ik deel dit alleen als het moet
Ik deel dit niet via internet

**Uw geboortedatum**
Ik heb hier geen problemen mee
Ik doe dit alleen als ik de ander vertrouw
Ik deel dit alleen als het moet
Ik deel dit niet via internet

blaise.cbs.nl

× Onderbreken ? Help

Matrixvragen

**In hoeverre bent u geneigd de volgende persoonlijke informatie door te geven via het internet?**
*Denk bijvoorbeeld aan websites, online enquêtes, social media of apps*

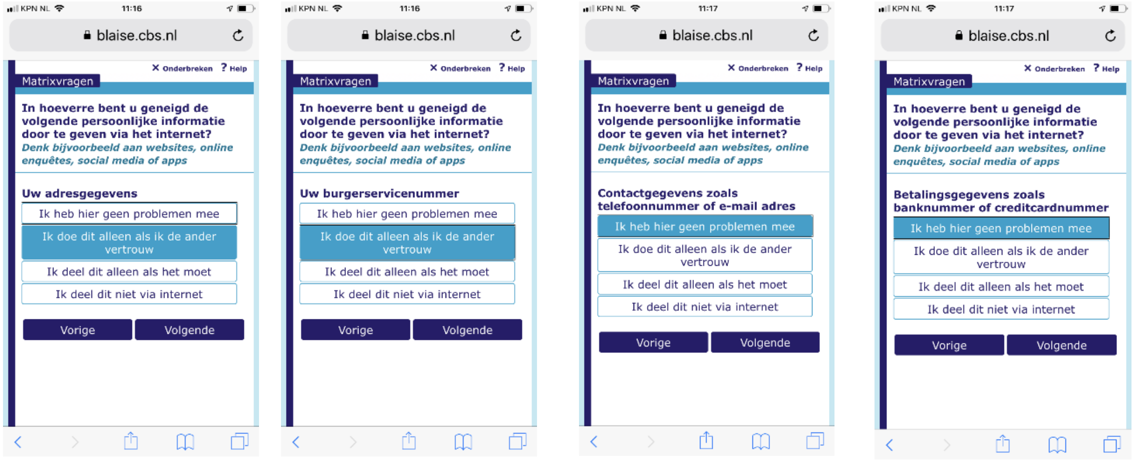
vertrouw
Ik deel dit alleen als het moet
Ik deel dit niet via internet

**Een foto van iemand anders, zonder toestemming van diegene**
Ik heb hier geen problemen mee
[✓ Ik doe dit alleen als ik de ander vertrouw]
Ik deel dit alleen als het moet
Ik deel dit niet via internet

Vorige Volgende

*Note: Question stem is always visible and separated from items by a thin blue line.*

*Figure 3.6: Presentation of same question as in figure 3.4 in paging design (first four items presented only).*



**In hoeverre bent u geneigd de volgende persoonlijke informatie door te geven via het internet?**
*Denk bijvoorbeeld aan websites, online enquêtes, social media of apps*

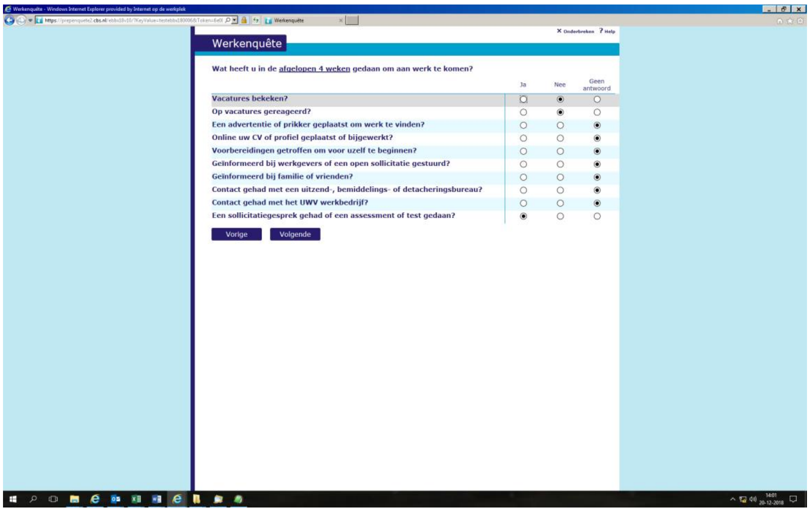
**Uw adresgegevens**
- Ik heb hier geen problemen mee
- Ik doe dit alleen als ik de ander vertrouw
- Ik deel dit alleen als het moet
- Ik deel dit niet via internet

**Uw burgerservicenummer**
- Ik heb hier geen problemen mee
- Ik doe dit alleen als ik de ander vertrouw
- Ik deel dit alleen als het moet
- Ik deel dit niet via internet

**Contactgegevens zoals telefoonnummer of e-mail adres**
- Ik heb hier geen problemen mee
- Ik doe dit alleen als ik de ander vertrouw
- Ik deel dit alleen als het moet
- Ik deel dit niet via internet

**Betalingsgegevens zoals banknummer of creditcardnummer**
- Ik heb hier geen problemen mee
- Ik doe dit alleen als ik de ander vertrouw
- Ik deel dit alleen als het moet
- Ik deel dit niet via internet

*Figure 3.7: Classic grid presentation of LFS looking for work question for tablet and PC.*



**Wat heeft u in de afgelopen 4 weken gedaan om aan werk te komen?**

| | Ja | Nee | Geen antwoord |
|---|---|---|---|
| Vacatures bekeken? | ○ | ● | ○ |
| Op vacatures gereageerd? | ○ | ● | ○ |
| Een advertentie of prikker geplaatst om werk te vinden? | ○ | ○ | ● |
| Online uw CV of profiel geplaatst of bijgewerkt? | ○ | ○ | ● |
| Voorbereidingen getroffen om voor uzelf te beginnen? | ○ | ○ | ● |
| Geïnformeerd bij werkgevers of een open sollicitatie gestuurd? | ○ | ○ | ● |
| Geïnformeerd bij familie of vrienden? | ○ | ○ | ● |
| Contact gehad met een uitzend-, bemiddelings- of detacheringsbureau? | ○ | ○ | ● |
| Contact gehad met het UWV werkbedrijf? | ○ | ○ | ● |
| Een sollicitatiegesprek gehad of een assessment of test gedaan? | ● | ○ | ○ |

*Note: After opening this screen the first question has a grey background (which moves to the next line when an answer is activated on the next line)*

*Table 4: Overview of grid questions tested on multiple devices by CBS*

| Grid Q nr | Question content | # of items | Answer options |
|---|---|---|---|
| 1 | LFS-looking for work | 10 items | yes/no/no answer |
| 2 | Questions on various types of ICT use | 8 items | yes/no |
| 3 | Question about inclination to share various types of info via the internet | 9 items | I don't mind/I only do this if I trust the other party/I only do this if I have to / I do not share this via the internet |
| 4 | Question about using the cloud for various types of documents; long introduction text to explain cloud use | 6 items | yes/no |
| 5 | Statements about environment and environment policies | 5 items | 5 point scale, fully labeled; completely disagree-completely agree |
| 6 | Question about extent of being able to contribute to solving environment problems for various actors (industry, own household, agricultural sector, other households, government) | 5 items | seven point scale, endpoints labeled 1-not at all, 7-to very large extent, midpoints only numbered 2-6 |
| 7 | Question about extent of being willing to contribute to solving environment problems for various actors (industry, own household, agricultural sector, other households, government | 5 items | seven point scale, endpoints labeled 1-not at all, 7-to very large extent, midpoints only numbered 2-6. |

# 4. Methodology and findings of the tests conducted at SSB

## 4.1 Test design SSB

We recruited eight test persons with a similar background in terms of age, 25-35 years. We chose this demographic group under the assumption that it is an age group that is comfortable with using smartphone and web for a variety of purposes. We did this to obtain more robust results, to avoid variations due to differences in familiarity with the technology. To further ensure such familiarity, the CAWI mobile tests were done on the respondents' own devices to make the tests as realistic as possible for the test persons. The drawback of our approach is that the results of our tests cannot necessarily be assumed to be representative of the general population - particularly regarding age and education. However, many problems identified for test persons in this group will likely also be problematic for population groups that are less familiar with modern communication technology.

*Table 5a: Summary of subjects tested*

| Subject ID | Age (approximate) | Gender | Education* | Device features/ Type of smart phone |
|---|---|---|---|---|
| 1 | 30 years | Man | High | Large iPhone |
| 2 | 25 years | Woman | Medium | Large iPhone |
| 3 | 27 years | Woman | High | Medium Samsung, broken screen |
| 4 | 27 Years | Woman | High | Small iPhone (SE) |
| 5 | 26 years | Man | High | Medium Samsung |
| 6 | 30 years | Man | High | Small iPhone (SE), broken screen |
| 7 | 25 years | Man | Medium | Medium iPhone |
| 8 | 25-30 years | Woman | High | Large iPhone |

**Classification of education: High education = Completed 3 years of higher education/university or more. Medium = Completed high school and Low = Completed basic level of education (10 years)*

For the future we suggest running similar tests for a broader spectre of respondent groups, particularly also less proficient smartphones users, to complement our obviously biased sample.

*Recruitment and incentives*

Two of the test persons were recruited from outside Statistics Norway, through adds placed on our Facebook page. They received a gift certificate to the value of € 30 as an incentive. The remaining six were recruited internally at Statistics Norway among new employees with no specific experience with questionnaires or usability. The internal participants did not receive any incentives.

*Test facilities, data security, and consent*

All the tests were done at the Test Lab facilities at Statistics Norway in Oslo. The tests were done using Tobii eye tracking equipment, to observe and record navigation and respondent's behaviour in filling in the questionnaire. The observation gives input to evaluate usability and comprehensibility aspects, and to observe what the test person notices, does not notice and concentrates on in the questionnaire. Using the eye tracking equipment made the test situation to some extent rigid, as the respondents' smartphones had to be placed in an unnatural contraption and the test persons were not free to turn their phones to achieve a landscape view etc. Also, it was more difficult to identify with the eye tracking what the test persons focused on the smaller screen/smartphones.

The recordings from Tobii eye tracking was anonymized to avoid identification of the test subjects. The data files were kept in a secure server with no access by outsiders outside the project. This includes electronic files containing information on interviewees. And files with analysis of tests were anonymized and stored separately from files with information on interviewees. All data files will be deleted three months after the project closes.

Test persons' consent to participation and recording was given aurally and recorded on videotape from test.

*The period when testing was conducted*
The tests took place in Oslo October 15th – November 2nd 2018.

*The level of experience of each interviewer*
We used three moderators for this project. All tests were conducted by senior staff with long experience as qualitative moderators with expertise in questionnaire design. The tests were observed and coded by the colleagues of the same moderator team, supplemented by one coder. The roles of the team members rotated through the project.

*Goals for testing*
We expected the tests to give us a greater insight into how the questions that were identified as poorly fit for mobile performed in practice, as well as how respondents managed and perceived usability/user friendliness in general and retrospectively assessed comprehensibility of terminology and task requested.

*Interview guide/test protocol*
In all the tests, we let the test person complete the questionnaire on their own smartphone and then conducted a retrospective review with them on what was perceived as easy and difficult during completion, as well as probing of certain questions to examine their cognitive comprehension of the questions.

We tested the entire ICT survey as an input for WP5. All tests were done on smartphones of different sizes (not PCs or tablets) and survey link was tested in vertical position, not land scape, as the responsive design was intended for a vertical screen.

We used the fitness criteria from deliverable 1 of WP5 to explore whether the respondents perceive them as difficult, satisficed, misunderstood or otherwise engaged in behaviour likely to result in measurement error or breakoff. On a question level, the relevant criteria (excluding survey-level criteria) was used to identify possibly particularly problematic questions. Table 5b is an attempt at adapting the typology to individual questions.

*Table 5b: WP5 fitness criteria adapted to single questions*

| *Dimension* | *Criterion* | *Operationalization* |
|---|---|---|
| Screen size | Introductions | Item with introduction |
| | Grid questions | Grid question<br>Number of items |
| | Question text | Item with > 20 words (excluding introduction text) |
| | # answer cat's | Item with > 5 answer categories |
| | Answer text | Items with > 10 words in at least one category |
| Touch navigation | Open question | Open question |
| | Many answers | Item with > 25 answer categories |
| Duration | # of items | Duration of test |
| | Database | Question with database interaction |
| | Complexity | Item that required calculations by the respondent |

We also used the Campanelli typology primarily intended for WP4 of MIMOD for identifying generally problematic questions. This typology of question characteristics relevant to measurement error was created for identifying questions where there is a high risk of measurement differences between modes, but it can also be used for identifying measurement risks connected with CAWI mode.

Additionally, the cognitive interview touched on perceived enjoyment, relevance and burden of the whole questionnaire completion experience.

*Reporting and documentation*

All tests were taped and recorded. Based on the screenshots of each question/screen/ a reporting form was filled in using the survey questions and evaluation questions to describe issues on comprehensibility (i.e. findings, observations, remarks from interviewers and quotes from respondents). In a next step the issues were analysed to determine the underlying type of problems assigned with codes (based on Oksenberg, Canell and Kalton 1991):

1) Comprehension; issues on comprehension of the response task, definitions and concepts
2) Access; issues on access to relevant information, memory records, knowledge i.e., in the reference period.
3) Process & report; issues on processing information as intended; issues on accuracy, calculation, guessing, sorting of information to match the response task/options, format/relevant/missing options.
4) Context; issues on the influence of routing and preceding questions

For issues on usability aspects, like problems with navigation, problems with functionalities, problems with open text fields etc., relevant codes were used. These are related to the type of device and possible risk of device effects.

## 4.2 Key issues and findings SSB

*Fitness criteria*

Applying the fitness criteria to the country specific version for mobile in table 6 below, we see it contains many more items than the model questionnaire. It has a larger number of interviewer introductions and instructions visible to the respondents. There were many questions that violated the criteria in respect to length of question text, answer categories and text, but there were no open ends and none with more than three answer categories.

*Table 6: WP5 fitness criteria score for ICT on smartphones*

| *Dimension* | *Criterion* | *Scores* | |
|---|---|---|---|
| | | *Model* | *SSB* |
| Screen size | Introductions | 5 | 12 |
| | Instructions | 4 | 52 |
| | Grids | 1 | 0 |
| | #Items per grid | 6 | NA |
| | Question text (>20 words) | 13 | 50 |
| | # answer cat's (>5 answer cat) | 7 | 0 |
| | Answer text (>10 words one cat) | 3 | 10 |
| Touch | Open questions | 0 | 1 |
| Navigation | Many answers (>25? answer cat) | 0 | NA |
| Duration | # of items | 39 | 110 |
| | Database | NA | No |
| | Complexity | No | No |

The country specific version is in clear violation of the fitness criteria. This was mostly due to a higher number of introductions and instructions, and the “mark all that applies” format that was converted to numerous Yes/No questions. This, however, was not unique to the mobile format: similar feedback also came from test persons who completed the PC and CATI versions of the question (for WP4 deliverable 3: Recommendations for questions and questionnaires).

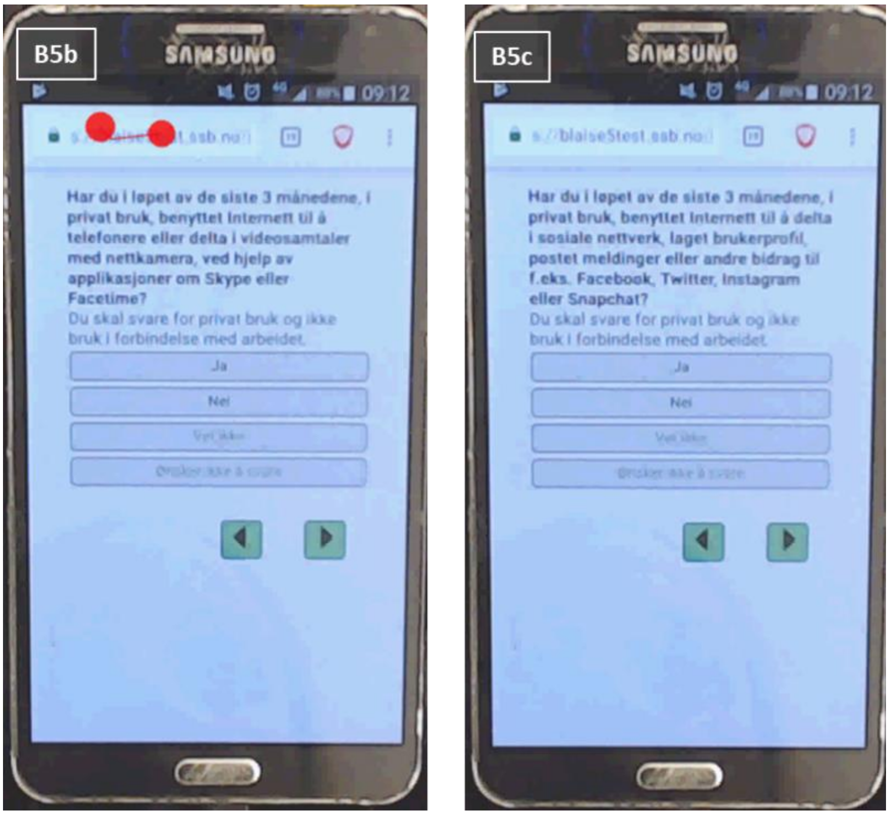
*Test persons’ overall impression*

This group’s overall impression with the survey was good. The link performed well and the length (approx. 12 minutes) was acceptable. In a few of the tests speed/loading time was not optimum and caused in a few cases confusion and possibly irritation.

*Long text and repetition*

Although the question text did fit the screen in almost all cases, the overall amount of text and repetition was perceived as burdensome by the test persons. A major finding was that the respondents had difficulties distinguishing new questions. This applied particularly to the sub questions in the long question batteries, but also questions with long introductions and instruction text. Several test persons noted the monotony of the task and perceived the questionnaire as repetitive. They said that they “looked for what was new” regardless of whether the repeated text was in bold or not, see figure 4.2a and b. One test person also commented on overall impression that the transitional introduction from one topic to another were awkward, as “there really wasn’t a change of topic” – it all had to do with ICT.

*Figure 4.2a. Hard to distinguish “what’s new” in question batteries with long sub question*



**Har du i løpet av de siste 3 månedene, i privat bruk, benyttet Internett til å telefonere eller delta i videosamtaler med nettkamera, ved hjelp av applikasjoner om Skype eller Facetime?**
Du skal svare for privat bruk og ikke bruk i forbindelse med arbeidet.

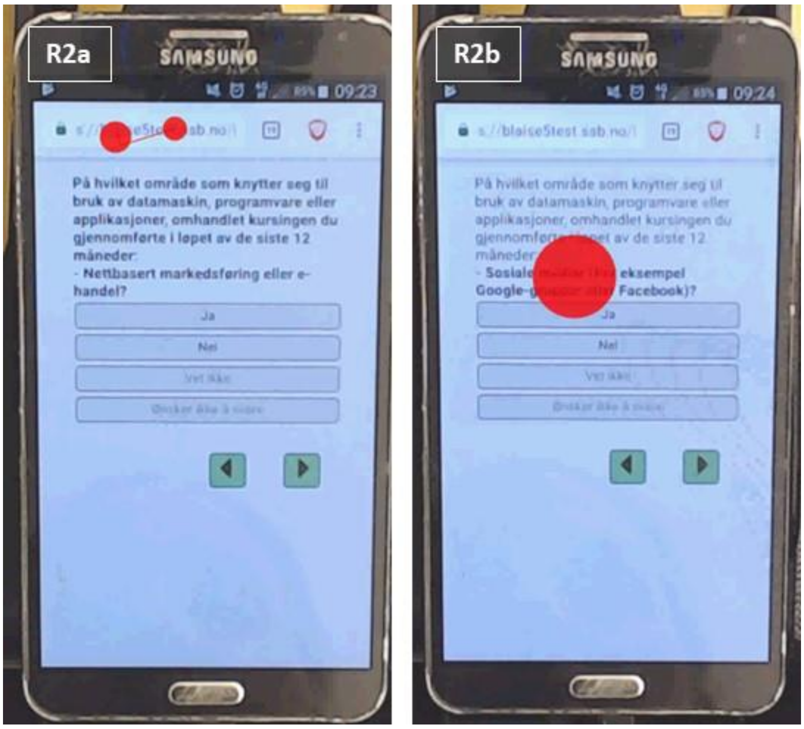
**Har du i løpet av de siste 3 månedene, i privat bruk, benyttet Internett til å delta i sosiale nettverk, laget brukerprofil, postet meldinger eller andre bidrag til f.eks. Facebook, Twitter, Instagram eller Snapchat?**
Du skal svare for privat bruk og ikke bruk i forbindelse med arbeidet.

*Vertical scrolling*

The overall usability was satisfactory, although some test persons with smaller iPhone models experienced minor problems. They had to scroll vertically on the most verbose questions to see all the response categories.

This happened in particular for questions with long intros like we see e.g. in the opening question A1, and/or questions with many answer categories like B1 about internet frequency, see figure 4.2c. Still, this did not lead to any measurement errors. However, in one case, a respondent did not see the “next” button and was not able to navigate to the next page, needing help from the moderator. We hope further development of screen adaptiveness will address this. In further studies we will test splitting the screen in introduction and question and reducing question and answer text/number of questions to reduce the need for scrolling.

*Figure 4.2b. Eyetracker shows focus on “what’s new” in R2b*



R2a:

På hvilket område som knytter seg til bruk av datamaskin, programvare eller applikasjoner, omhandlet kursingen du gjennomførte i løpet av de siste 12 måneder:
- Nettbasert markedsføring eller e-handel?

Ja

Nei

Vet ikke

Ønsker ikke å svare

◀ ▶

R2b:

På hvilket område som knytter seg til bruk av datamaskin, programvare eller applikasjoner, omhandlet kursingen du gjennomførte i løpet av de siste 12 måneder:
- Sosiale medier (for eksempel Google-annonser eller Facebook)?

Ja

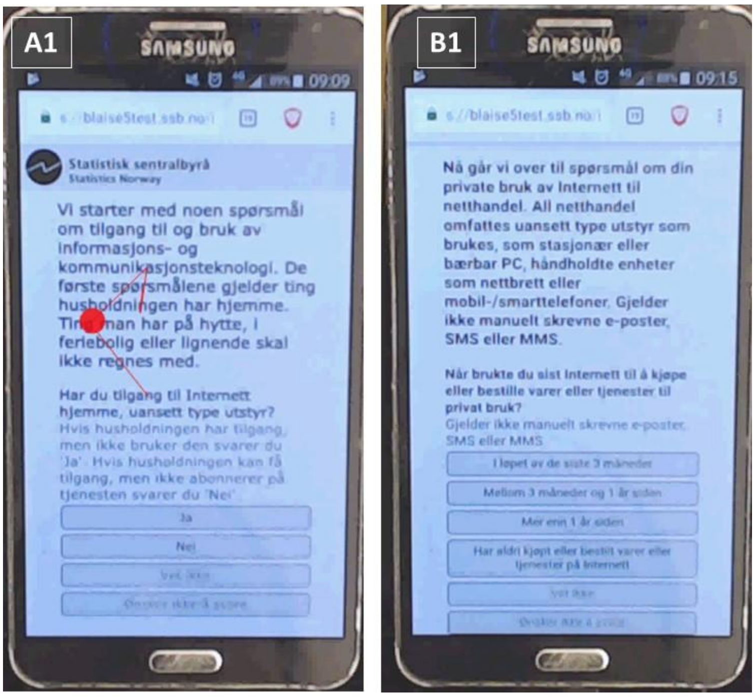
Nei

Vet ikke

Ønsker ikke å svare

◀ ▶

*Figure 4.2c: Need for scrolling to show next-button on small screens*



A1:

Statistisk sentralbyrå
Statistics Norway

Vi starter med noen spørsmål om tilgang til og bruk av informasjons- og kommunikasjonsteknologi. De første spørsmålene gjelder ting husholdningen har hjemme. Ting man har på hytte, i feriebolig eller lignende skal ikke regnes med.

Har du tilgang til Internett hjemme, uansett type utstyr?
Hvis husholdningen har tilgang, men ikke bruker den svarer du 'Ja'. Hvis husholdningen kan få tilgang, men ikke abonnerer på tjenesten svarer du 'Nei'

Ja

Nei

Vet ikke

Ønsker ikke å svare

B1:

Nå går vi over til spørsmål om din private bruk av Internett til netthandel. All netthandel omfattes uansett type utstyr som brukes, som stasjonær eller bærbar PC, håndholdte enheter som nettbrett eller mobil-/smarttelefoner. Gjelder ikke manuelt skrevne e-poster, SMS eller MMS.

Når brukte du sist Internett til å kjøpe eller bestille varer eller tjenester til privat bruk?
Gjelder ikke manuelt skrevne e-poster, SMS eller MMS

I løpet av de siste 3 måneder

Mellom 3 måneder og 1 år siden

Mer enn 1 år siden

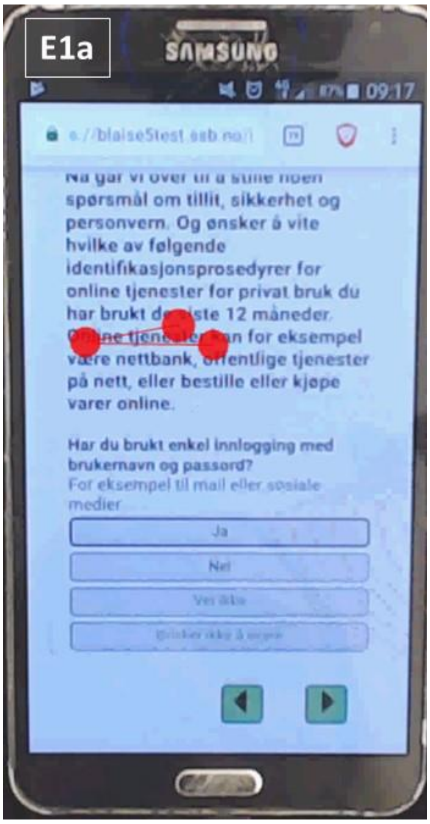
Har aldri kjøpt eller bestilt varer eller tjenester på Internett

Vet ikke

Ønsker ikke å svare

*Loading time and automatic forwarding*
The loading time between pages varied between tests and lead to some navigation confusion, as some users tried to click on the “next” button before the automatic loading of the next page. In one case, a test person was unsure of whether this had led to the skipping of a question.

*Figure 4.2d: Selection of answers button shown with black, bold outline*



Nå går vi over til å stille noen spørsmål om tillit, sikkerhet og personvern. Og ønsker å vite hvilke av følgende identifikasjonsprosedyrer for online tjenester for privat bruk du har brukt de siste 12 måneder. Online tjenester kan for eksempel være nettbank, offentlige tjenester på nett, eller bestille eller kjøpe varer online.

Har du brukt enkel innlogging med brukernavn og passord?
For eksempel til mail eller sosiale medier

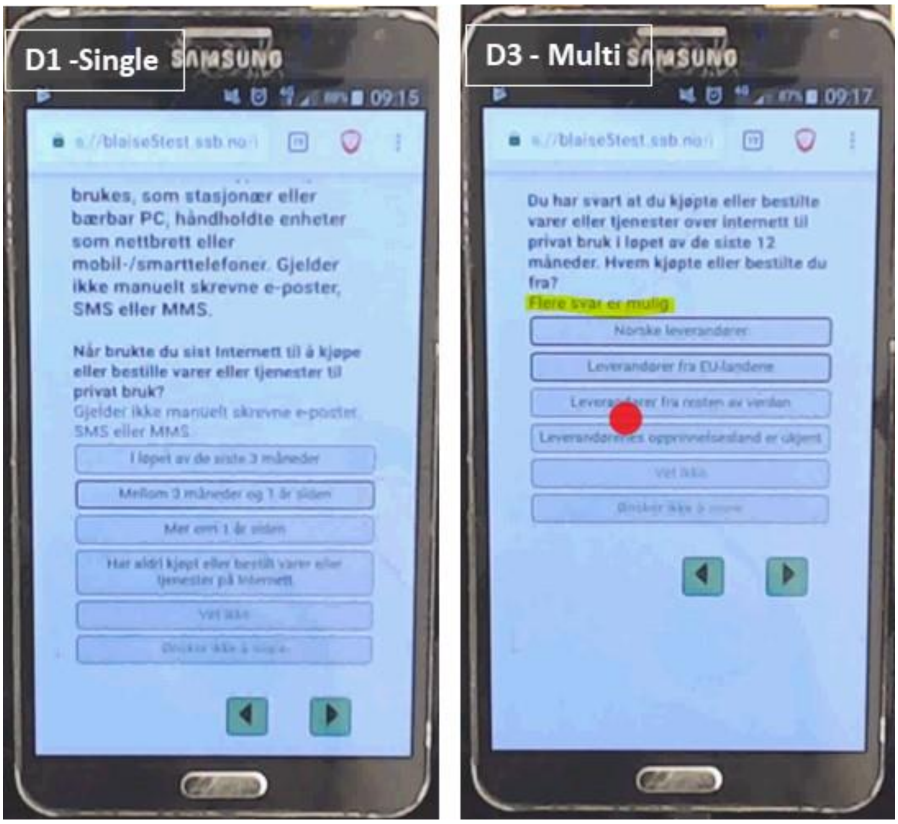
For the future we want to experiment with a stronger black outline around selected answer button to reduce uncertainty. The answer buttons could possibly also be a bit wider and more finger-friendly. Adding a short delay before automatic forwarding to the next page, just enough for the respondents to notify the stronger outline around the answer, should also be tested.

*Distinction between single vs. multiple response questions*
The fact that it was not easy to distinguish the few questions with more than one response option (multi) from the one option (single) questions in the mobile format led to some initial navigation confusion for three of the eight mobile version test persons. They picked one of the options, expecting to continue automatically to the next page. When this did not happen, they all believed they had not actually picked an option: the black frame around the option was not enough to make this clear (see example E1a in figure 4.2d above). After having unintentionally deselected the response, they re-selected the response category and clicked on the “Next” button to proceed. This confusion did not lead to measurement errors, but it is an annoyance we would like to find better design solutions for.

The “Do not wish to answer” button was not used by any of the test persons. The “Don’t know” button was however used on several questions, notably of two types: questions with different technical terms, and the final element of long question batteries asking about “other” uses of/purchases through the Internet.

On question A2 on types of Internet connections used at home in figure 4.2g, three of eight test persons used "Don't know" for one or both narrowband connection types (question c and d), due to unfamiliar and irrelevant terms that showed to be hard to separate. In the telephone interviews, three of six test persons responded "Don't know" to the same questions, even though it was not offered as an explicit response option. Lower share of "Don't know" indicates that data collection with visual stimuli is better assistance than aural for this question, but the level is still too high, indicating that the question is too complicated. In the retrospective interviews we learned that the terms used were too technical and old fashion and should have been updated.

*Figure 4.2e. Question design of single (D1) answer and multi (D3) "mark all that applies" response (in yellow)*



brukes, som stasjonær eller bærbar PC, håndholdte enheter som nettbrett eller mobil-/smarttelefoner. Gjelder ikke manuelt skrevne e-poster, SMS eller MMS.

**Når brukte du sist Internett til å kjøpe eller bestille varer eller tjenester til privat bruk?**
Gjelder ikke manuelt skrevne e-poster, SMS eller MMS

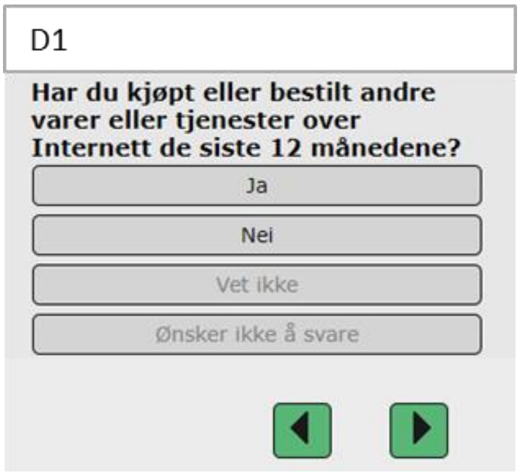
- I løpet av de siste 3 måneder
- Mellom 3 måneder og 1 år siden
- Mer enn 1 år siden
- Har aldri kjøpt eller bestilt varer eller tjenester på Internett
- Vet ikke
- Ønsker ikke å svare

**Du har svart at du kjøpte eller bestilte varer eller tjenester over internett til privat bruk i løpet av de siste 12 måneder. Hvem kjøpte eller bestilte du fra?**
Flere svar er mulig

- Norske leverandører
- Leverandører fra EU-landene
- Leverandører fra resten av verden
- Leverandørenes opprinnelsesland er ukjent
- Vet ikke
- Ønsker ikke å svare

*Use of "Don't know" and "Do not wish to answer" buttons*

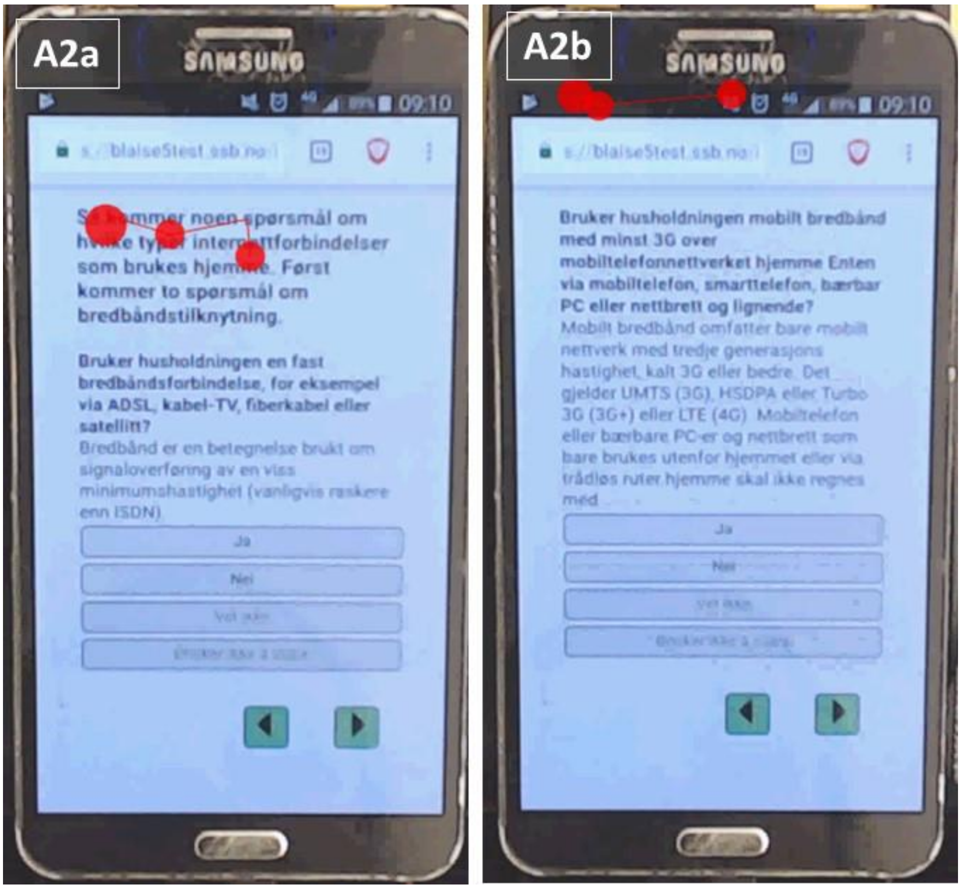
*Figure 4.2f: Greying out design for "Don't know" and "Do not wish to answer" button*



D1

**Har du kjøpt eller bestilt andre varer eller tjenester over Internett de siste 12 månedene?**

*Figure 4.2g. Difficult terminology increases the number of "Don't know" answers*



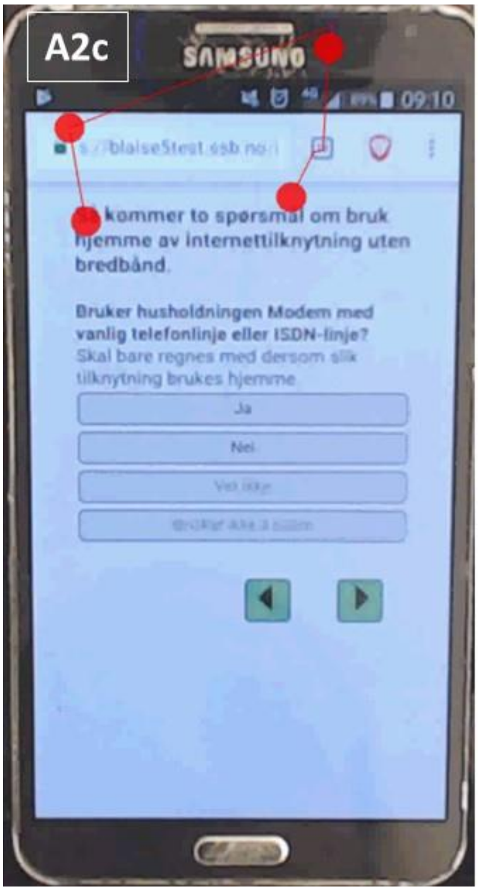
**Så kommer noen spørsmål om hvilke typer internettforbindelser som brukes hjemme. Først kommer to spørsmål om bredbåndstilknytning.**

**Bruker husholdningen en fast bredbåndsforbindelse, for eksempel via ADSL, kabel-TV, fiberkabel eller satellitt?**

Bredbånd er en betegnelse brukt om signaloverføring av en viss minimumshastighet (vanligvis raskere enn ISDN)

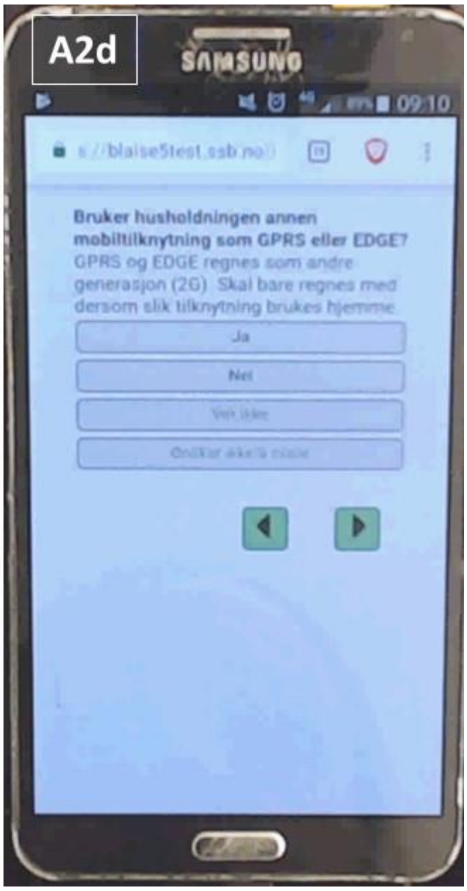
**Bruker husholdningen mobilt bredbånd med minst 3G over mobiltelefonnettverket hjemme Enten via mobiltelefon, smarttelefon, bærbar PC eller nettbrett og lignende?**

Mobilt bredbånd omfatter bare mobilt nettverk med tredje generasjons hastighet, kalt 3G eller bedre. Det gjelder UMTS (3G), HSDPA eller Turbo 3G (3G+) eller LTE (4G). Mobiltelefon eller bærbare PC-er og nettbrett som bare brukes utenfor hjemmet eller via trådløs ruter hjemme skal ikke regnes med



[illegible] kommer to spørsmål om bruk hjemme av internettilknytning uten bredbånd.

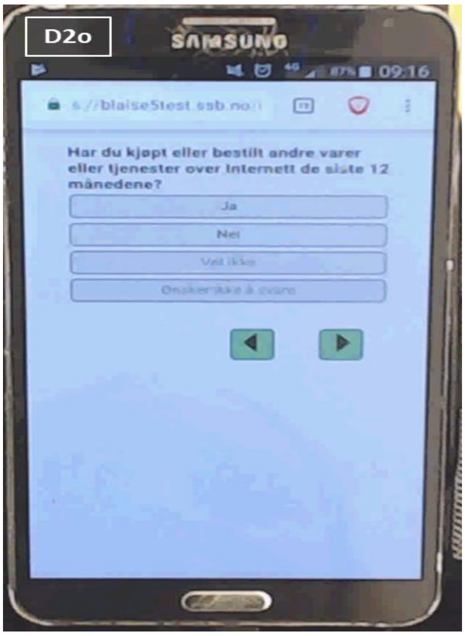
**Bruker husholdningen Modem med vanlig telefonlinje eller ISDN-linje?**
Skal bare regnes med dersom slik tilknytning brukes hjemme



**Bruker husholdningen annen mobiltilknytning som GPRS eller EDGE?**
GPRS og EDGE regnes som andre generasjon (2G). Skal bare regnes med dersom slik tilknytning brukes hjemme.

On question D2 on "goods and services" bought or ordered over the internet in figure 4.2h, four of eight test persons responded "Don't know" to D2o on "other" goods or services. Being the last of 14 different categories of goods and services, the response task of both remembering all the previous categories, among them several difficult and unintuitive ones, and then assessing whether anything else had been procured, is very complicated.

*Figure 4.2h: Use of "other" after long question batteries increases number of "Don't know" answers*



**Har du kjøpt eller bestilt andre varer eller tjenester over Internett de siste 12 månedene?**

In the retrospective interviews, two of the test persons commented that she/he interpreted the greying out as indicating that we (Statistics Norway) would prefer that respondents did not use them. Two others were unsure of whether it was actually possible to select them, although one of them actually did. Another said that "after a while, I just noticed the [non-grey] response options, and kind of forgot about the two last ones." This shows that the greying out design for "Don't know" and "Do not wish to answer" achieves the same as the interviewer instruction "Do not read out answer" does for CATI and it is fair to assume it does not give measurement errors between the two data collection methods.

*Open ended question requiring a number*

In the ICT survey only D5 on amount spent on ordered goods and services was an open question to fill in amount spent in local currency. The test persons did not have any usability or navigation issues with this question, but the findings from the retrospective interviews showed cognitive issues we will address in the next section.

## 4.3 Findings from the retrospective cognitive interviews SSB

The ICT test questionnaire took on average 12 minutes to complete. The overall impression of the test persons was that survey length and the questions were OK. Still, as we already have mentioned under usability findings, the retrospective interviews also support that verbosity, long introductions and repetitions is burdensome for the respondents. Further, the retrospective cognitive interviews also show unclear terminology and recall as challenges that might lead to measurement error.

In respect to comprehension and recall, the most problematic questions were types of internet connections (A2), internet activities (B5) and purchases of goods and services online (D2):

*A2 on types of internet connection*
On question A2 on types of internet connection used at home (see figure 4.2g above) all test persons struggled with the old-fashioned terminology, and we have already seen that this lead to a high level of "Don't knows". In the tests, six test persons responded that they used "Fixed broadband connections" at home. Additionally, two test persons answered that they used "Mobile broadband" at home. In the retrospective cognitive interview, a third test person was unsure of whether to select this option, and a fourth said that she used mobile broadband, but not at home. When the fixed broadband connection is temporarily unavailable, the mobile broadband will be used, mostly on smartphones, but not normally or regularly.

Further; the split in four questions presented on four separate screens also adds to complicate the comprehension of the terminology for internet types and conditions for answer. For future revisions of the ICT survey we urge that terminology to be updated and to test a one screen question design solution.

*B5 on types internet activities*
In the retrospective cognitive interviews, one of the test persons commented that the three-month reference period could be problematic for activities that were rarely carried out. Many of the activities (all in all 13 questions) were carried out daily and were no problem to respond to. Another test person commented that games would be downloaded for the children, and that it was unclear whether it had to be downloaded for own use or not.

*D2 on purchases of goods and services online*
Like B5, question D2 on purchases of goods and services online was also perceived by several respondents as repetitive and boring. The tests showed that many respondents were unsure of whether they had purchased "other" goods or services. This is due to lack of visual context and limited working memory, as it is difficult to remember purchase last 12 months for all the 14 categories, and then retrieve and process the information.

The findings showed that use of such a final “residual” items on these questions (e.g. question D2o “Did you buy any other goods or services [that would not fit in the previous 14 (!) categories]”, led to several test persons responding “Don’t know”. This was also the case for the parallel test on PC and CATI. It is clear that “other” is fruitless at the end of so many “goods and services” listed and should be avoided when possible.

Unclear response categories may also add to cognitive challenges. One of the test persons commented that several of the categories were unclear and lumping different things together, e.g. TV subscriptions, pay-per-view and different phone costs. She also said she sensed that the sequence should have been less random to help her answer. Another test person commented that the “other goods or services” category could be split in two, as she frequently bought cleaning services online.

As mentioned under usability, repetitiveness of long battery questions of Yes/No is considered a monotonous task. It works well when it is not too many questions and the terminology used is clear and intuitive. For future revisions of the ICT questions, both B5 and D2 can have a reduced response burden by cutting down on the number of sub-questions and by testing whether a new internal sequence can improve comprehension. From the retrospective interviews, we will also suggest testing whether a 12-month period is the most suitable time period to account online purchases for.

*D5 (open end) on amount spent on ordered goods and services*
Several test persons stalled when asked to estimate how much they spent on buying or ordering goods or services online the last 3 months in local currency, in an open question requiring a number. From the retrospective interview we learned that this was hard to recall and calculate for test persons who had made many purchases. We also saw that it was hard to evaluate what to include and exclude from the calculation. Several test persons “forgot” to include tickets for bus/public transportation/parking, movies etc, as they did not without consideration/help see this as ordering goods or services or online shopping. As online payment and shopping is taking off, we need to consider testing the need to adjust this question and suitable recall period to fit the “terrain” we want to measure.

# 5 Methodology and findings of the tests conducted at CBS

## 5.1 Test methodology CBS

CBS conducted tests evaluating questionnaires on smartphones, tablets and a laptop with 20 persons in November 2018, see Table 5.1. Various methods and sources were used to recruit respondents: emails and (if a phone number was available) telephone calls to respondents who had applied to participate in a previous test but could not participate then, respondents who had participated in a test before and said they could be contacted again and a group of respondents to a regular CBS survey who said they could be contacted again. Also, flyers were distributed at three nearby supermarkets and at the employment agency. Recruitment aimed to get a variety in age, sex, educational level, both android and iOS devices, and, important for testing the various parts of the LFS, both self-employed, employed and looking for work.

Previous to the test, respondents completed a short questionnaire in which they, among other things, stated the devices they used and the device they prefer for completing questionnaires. Based on this, respondents were prior to the test assigned to specific devices for various parts of the test. We only assigned respondents to devices familiar to them. We asked them to bring their own mobile devices to the test. In some cases CBS phones or tablets had to be used because the devices of the respondents were not available or not working correctly (e.g. internet connection problems). For the PC version of the test a CBS laptop with a mouse was used. Respondents received a gift card of 40 euro's for their participation.

All test were conducted face to face at the building of CBS in Heerlen. Both observation and probing was used. A list of retrospective probes was prepared in the testing protocol, and interviewers also used spontaneous probes (both concurrent and retrospective).

The test consisted of three parts: 1) household roster (to be completed on two devices) 2) the LFS (completed on one device) 3) Grid questions (a set of grid questions on various topics (ICT-usage, Cybersecurity, Environment) presented in a traditional format for PC/tablet or in three alternative formats for smartphones (stem fix and scrolling by respondent, stem fix with auto scroll and paging design). Interviews lasted about 1,5 hours.

In the test interviews we aimed to get greater insight into how respondents manage and navigate screens in various sizes and evaluate user friendliness of the design, i.e. the style sheets with respect to colours, use of buttons, scrolling, font size et cetera. Also, we aimed to get some indication whether there is a risk of device effects, that is, do we see device-related measurement errors.

For questions on occupation and job activities we also explicitly evaluated if the instruction to elaborate and typing in text which contains specific features of the occupation and job activities, was understood and evaluated correctly. We tested two versions, different in type of formulation and instruction direction as well as including a fewer number of examples, i.e., shortened question text length and a different lay-out

Next to usability, another aim was to asses comprehensibility. In the evaluation/debriefing interview afterwards, we checked whether respondents' employment status and educational level can be distracted from their answers correctly. In the case of errors or misinterpretations, we explored if we could get specific input for question improvements and/or adaptations of texts. Also, we explored if any difficulties in the response process could be linked to visual aspects or usability, e.g. the lay-out of the presentation of question and answer texts, the order of questions or the interaction with the device.

Four experienced methodologists conducted the interviews. All interviews were video recorded. Next to the video recording, for completion on the laptop a Camtasia recording was made, registering the screen and usually the voice and the face of the respondent (sometimes the webcam for filming the face was accidently switched of, frequently the Camtasia audio recording made with the laptop microphone was not working properly, in those cases we could use the recording made with the video camera). For completion on a mobile device, the screen and the hand of the respondents were filmed via MrTappy (device with webcam) and this film was recorded in Camtasia, again, usually with the face and voice of the respondent. All recording were made with written permission of the respondents.

For each interview a summary report was written, according to a fixed template and using the recordings of the interview. All reports were merged in one Excel file. This allows comparing findings over topics and within respondents. The interviewers discussed the findings in several meetings.

As can be seen in table 5.1, we tested the LFS 10 times on a smartphone, 5 times on a laptop and 5 times on a tablet.

*Table 5.1: Characteristics of test respondents and devices*

| Res pnr | Age | Gender | Main status: employed, unemployed, student, pensioner or other | Educational level (low=ISCED 0-2; medium=ISCED 3-4; high=ISCED 5-8) | Device used for testing LFS |
|---|---|---|---|---|---|
| 1 | 21 | Male | Student, currently working full-time paid internship and has side job | High | laptop (CBS) |
| 2 | 41 | Female | works as employee | Medium | smartphone Samsung Galaxy 6 edge |
| 3 | 49 | Female | working from an employment agency, lookig for work | High | Iphone 6S+ |
| 4 | 55 | Male | works 9hrs a week as employee, has income from social security (incapacitated) | Low | laptop (CBS) |
| 5 | 40 | Male | Incapacitated | Low | Samsung galaxy S7 |
| 6 | 62 | Male | Incapacitated | High | laptop (CBS) |
| 7 | 54 | Male | works as employee | Medium | Samsung A5 |
| 8 | 56 | Female | Unemployed | High | laptop (CBS) |
| 9 | 16 | Female | student, with small side job | Low | Iphone X |
| 10 | 46 | Female | self-employed | Medium | Iphone 6 (CBS) |
| 11 | 50 | Male | works as employee | High | Ipone XS |
| 12 | 53 | Female | Works as employee (a larger job and a small side job) and is looking for work | Medium | iPad |
| 13 | 52 | Male | Houseman | Medium | iPad |
| 14 | 24 | Female | student with side job | High | Smartphone Sony X2 |
| 15 | 39 | Female | housewife, looking for job, does volunteer work | Low | tablet Galaxy S2 |
| 16 | 54 | Female | Unemployed | Medium | iPhone) |
| 17 | 23 | Male | just finished school, looking for work | High | laptop (CBS) |
| 18 | 24 | Male | works as employee | Low | smartphone Samsung Galaxy |
| 19 | 36 | Male | works as employee | Low | iPad (CBS) |
| 20 | 58 | Male | incapacitated, self-employed | High | iPhone |

As for the grid questions, the looking for work question was tested by eight respondents. These were respondents who either spontaneously came in the routing of looking for work as well as respondents who were currently working or incapacitated, but indicated in the interview they were looking for work or would be looking for work in the near future. As shown in table 5.2, two respondents completed the question on a laptop, two on a tablet and four on a smartphone. For the set of grid questions tested after the LFS, items 2 – 7 almost all respondents completed at least one of the four options. Depending on the time available they also seriously completed or looked at and/or tried a few questions one or more other options. The order in which the respondents tried the options varied more or less.

*Table 5.2: number of respondents per type of grid test*

| Grid questions | completed | | | looked at/ tried a few questions | | | Total |
|---|---|---|---|---|---|---|---|
| | Laptop | Tablet | Smart-phone | Laptop | Tablet | Smart-phone | |
| GQ1 (looking for work) | 2 | 2 | 4 | - | - | - | 8 |
| GQ2-GQ7 (ICT & environment questions) | | | | | | | |
| classic grid | 6 | 8 | - | 2 | 4 | - | 20 |
| stemfix-scroll by respondent | - | - | 8 | - | - | 9 | 17 |
| stemfix-autoscroll | - | - | 4 | - | - | 10 | 14 |
| Paging | - | - | 7 | - | - | 11 | 18 |

## 5.2 Findings tests CBS

Here we present all relevant results and reflect on the fitness criteria issues from WP5. This is also linked to WP4 when it comes to mode/device issues in relation to comprehensibility as we tested in different devices, i.e., smartphone, tablet and PC/laptop.

### *General findings*

Many respondents did not have any big problems filling in the LFS questionnaire on the smartphone. Several commented they found the questionnaire easily doable, but also made comments about specific usability aspects (e.g. typing in a text field) or made general comments about preferring to work on a larger screen and/or with a keyboard and mouse; *"It is easier if you can use the keyboard instead of the little ones on your mobile phone"; "It is a bit more difficult because of the small screen, more afraid to make errors", "Do not like filling in on the smartphone as you cannot type with two hands".*

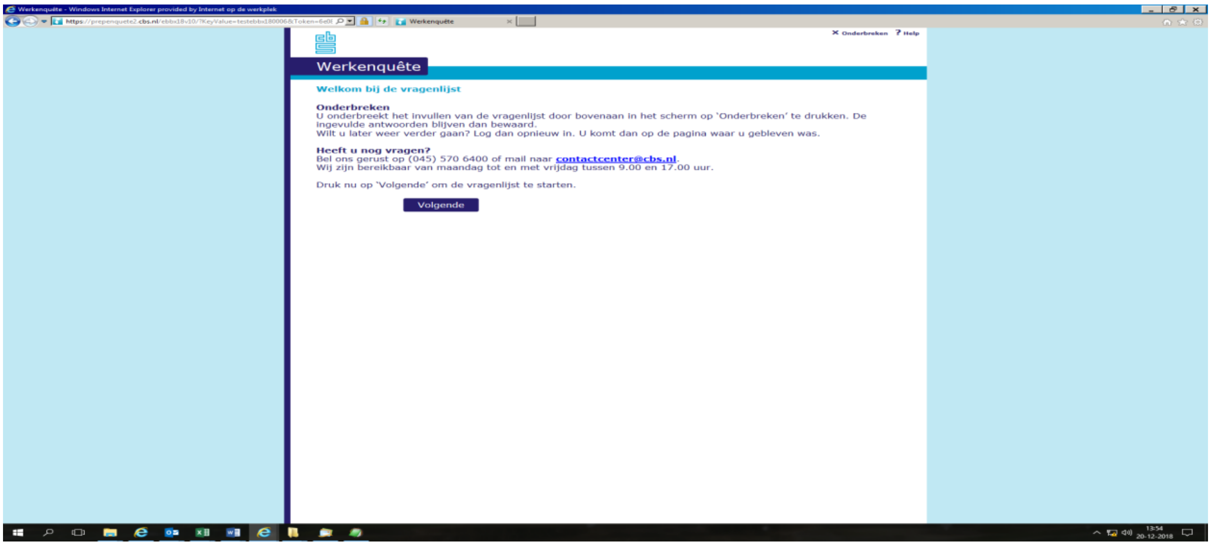
The test indicated a large number of issues that can be improved. Some of these have to do with the style sheets chosen for smartphones by CBS. The more general findings will be touched upon in the section screen size, as they may be informative for others as well. Other issues had to do with the actual contents of the questions and the interaction with usability features that may evoke or enforce issues. This in general is related to the context of questions, the navigation, and speed of filling in. We observed with the respondents in our test, a risk of making unintended mistakes by the way they navigate, use buttons,  or touch screen possibilities or the high speed of filling in. This might evoke  a risk for device specific measurement errors, especially in a smartphone design.  Also, the high amount of text with little space in the question and answer categories on a relative small screen is in general cognitively demanding for respondents (see also Nielsen & Budiu 2013). Finally, we found that the ease with which the questionnaire was filled out varied with the features of  respondent's device. For example for a  Huwaei V20 phone we found that the questionnaire behaved strange when entering a birth date (from the test report: "The screen jumps up and down when entering the date. You don't know where you are when this happens. This is weird and not user friendly.") On some of the smaller smartphones, the questionnaire did not perform to satisfaction (two phones with a 4 inch screen, iPhone 5S and iPhone SE could not correctly display the questionnaire). Some tablet respondents were not used (and in one case it seemed also technically not possible) to switch to landscape mode; the questionnaire did not display correctly on tablets in portrait mode. Finally, devices varied in loading time; a too long loading time can lead to errors as respondents re-select an option (thinking this had not been selected) and may accidently and sometimes without ever realizing it select something else. For example in one of our test reports of a respondent on smartphone: "R accidentally selected an answer to the first question because the finger that was used to start the survey (tap next button) was left on that place a little bit too long. It seems that because of the slow response she tapped that next button again, but in the meantime the page had loaded and she accidentally selected an  answer to the first survey question".

### *Dimension Screen size*

In all cases, the criteria evaluate the size of survey items on a screen and thus the overall visibility of the items and the need to scroll. Partial invisibility of survey items may lead to confusion, underreporting of particular answer categories and respondent fatigue.

In general, at CBS the design choice on smartphone is one question on a screen and the question text is shortened as much as possible. The screen might look full or busy on a smartphone screen as the size and space is smaller, also with lesser amount of texts. In comparison with a tablet or laptop/pc screen size is broader and more suitable for and in general easier to collect response tasks and information.
See figures 5.1a on the information of the welcome page

*Figure 5.1a: Welcome page LFS survey on tablet/laptop PC version*



Werkenquête

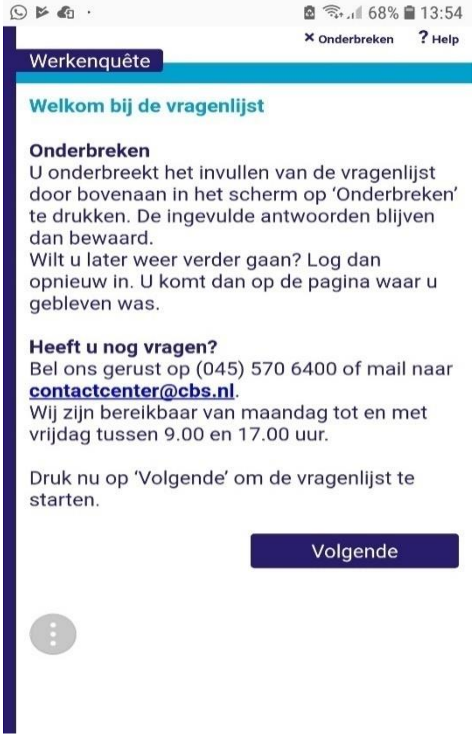
Welkom bij de vragenlijst

**Onderbreken**
U onderbreekt het invullen van de vragenlijst door bovenaan in het scherm op 'Onderbreken' te drukken. De ingevulde antwoorden blijven dan bewaard.
Wilt u later weer verder gaan? Log dan opnieuw in. U komt dan op de pagina waar u gebleven was.

**Heeft u nog vragen?**
Bel ons gerust op (045) 570 6400 of mail naar contactcenter@cbs.nl.
Wij zijn bereikbaar van maandag tot en met vrijdag tussen 9.00 en 17.00 uur.

Druk nu op 'Volgende' om de vragenlijst te starten.

*Figure 5.1b: Welcome page LFS survey on smartphone*



Werkenquête

Welkom bij de vragenlijst

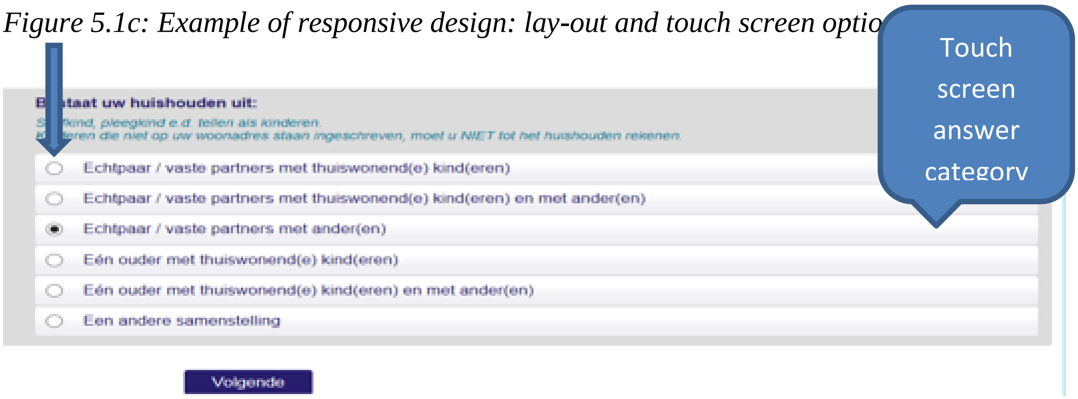
**Onderbreken**
U onderbreekt het invullen van de vragenlijst door bovenaan in het scherm op 'Onderbreken' te drukken. De ingevulde antwoorden blijven dan bewaard.
Wilt u later weer verder gaan? Log dan opnieuw in. U komt dan op de pagina waar u gebleven was.

**Heeft u nog vragen?**
Bel ons gerust op (045) 570 6400 of mail naar contactcenter@cbs.nl.
Wij zijn bereikbaar van maandag tot en met vrijdag tussen 9.00 en 17.00 uur.

Druk nu op 'Volgende' om de vragenlijst te starten.

The style sheets of CBS incudes for all devices a responsive design by Blaise 5 meaning touch navigation with touchscreen/big buttons for the entire answers categories (no need to click on the circle, possible to click on the complete space of the banner of answer category). Blue, italic text is used for instruction texts. Dark blue bold text for question text. Light grey shadows for the answer categories.
See figure 5.1c below.

*Figure 5.1c: Example of responsive design: lay-out and touch screen optio*



**B taat uw huishouden uit:**

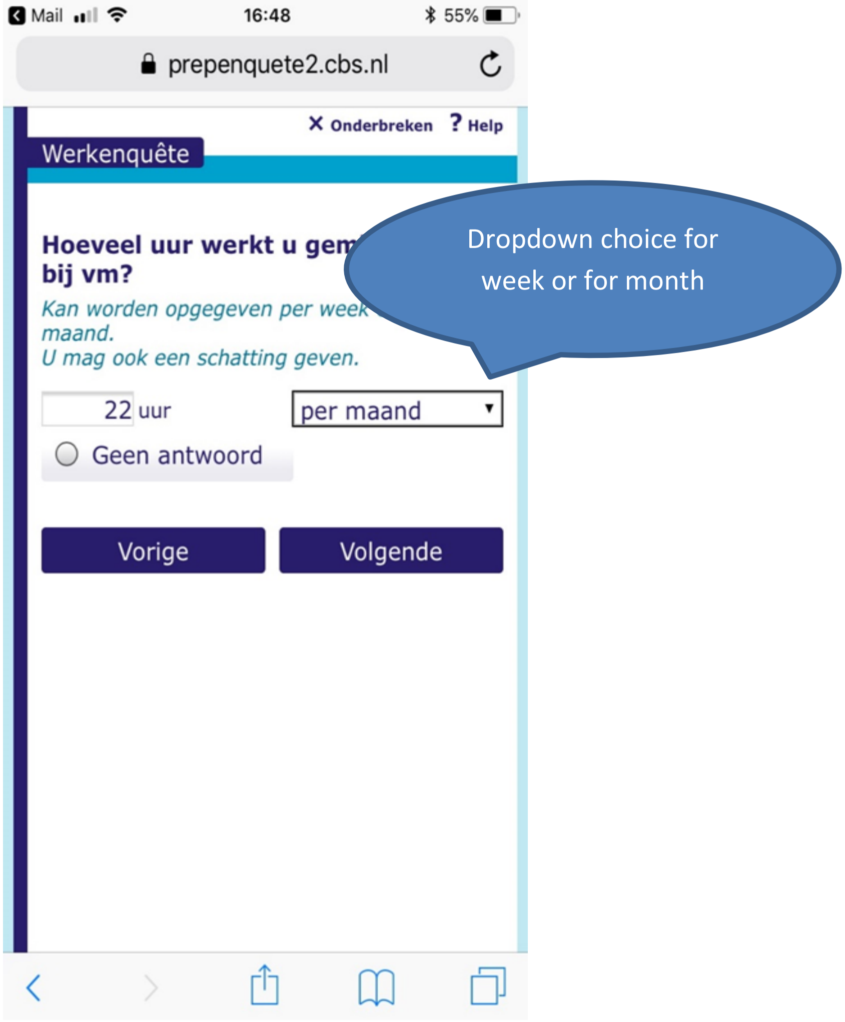
*S kind, pleegkind e.d. tellen als kinderen.*
*K eren die niet op uw woonadres staan ingeschreven, moet u NIET tot het huishouden rekenen.*

○ Echtpaar / vaste partners met thuiswonend(e) kind(eren)
○ Echtpaar / vaste partners met thuiswonend(e) kind(eren) en met ander(en)
● Echtpaar / vaste partners met ander(en)
○ Eén ouder met thuiswonend(e) kind(eren)
○ Eén ouder met thuiswonend(e) kind(eren) en met ander(en)
○ Een andere samenstelling

General usability findings from testing at CBS:

1. Several respondents confirmed by their comments and/or their behavior what is known from the literature: it takes more effort to read information from a small screen. The required effort to find relevant information is higher if this information is outside the viewable space.
2. The colouring and font for questions was generally appreciated. However, several respondents indicated that slightly larger letters and more contrast would be better. The legibility of the text in the answer buttons was considered somewhat less. The italic, blue letters for the introduction text were often not read and considered more difficult to read (see also the more detailed discussion of this in findings on introduction/instructions below).
3. For the touchscreen devices, and especially smartphones, we saw that selecting the correct answer options required more effort from respondents. This was partly caused by the fact that the touch targets in our questionnaires seemed too small and too close together. Respondents had to be very careful to select the right option.
4. Possibly, the buttons for the answer options on the touch screen devices are also too broad. In some cases that posed problems in case of scrolling, as respondents then accidentally selected an answer.
5. The ‘no answer’ button is too close to the substantial answer buttons. This increases the chance of erroneous answers. A visual distinction is needed here, more space.
6. Some respondents expressed appreciation for the fact that there is mostly only one question per screen.
7. Vertical scrolling is no problem for any respondent, even the ones with less experience using touchscreen devices, and all scroll spontaneously. Sometimes however, an answer is erroneously selected by scrolling.
8. Although most respondents did not have problems with the dropdown choice for week / month in the LFS question on number of contract hours, it did not work as intended for all respondents. The chance of faulty answers is real. It might be better to offer the two answers as multiple choice.  See figure 5.2.
9. It was not always clear for respondents that an open text field was active, they expected some visual; indication as for example a blinking  cursor to indicate that they could start typing.
10. On the current CBS smartphone stylesheets there is no visual difference between “select all that apply” questions and single choice questions. This sometimes confused respondents.
11. We noticed a few times how respondents on smartphones accidentally navigated in the questionnaire using the back button from the phone itself (the symbol ‘<’see red arrow in figure 5.3 below) instead of the back-button of the survey ‘vorige’ (blue arrow). This causes them to leave the questionnaire and having to login again.

*Figure 5.2: Question on working hours and dropdown menu, average hours for week or month*



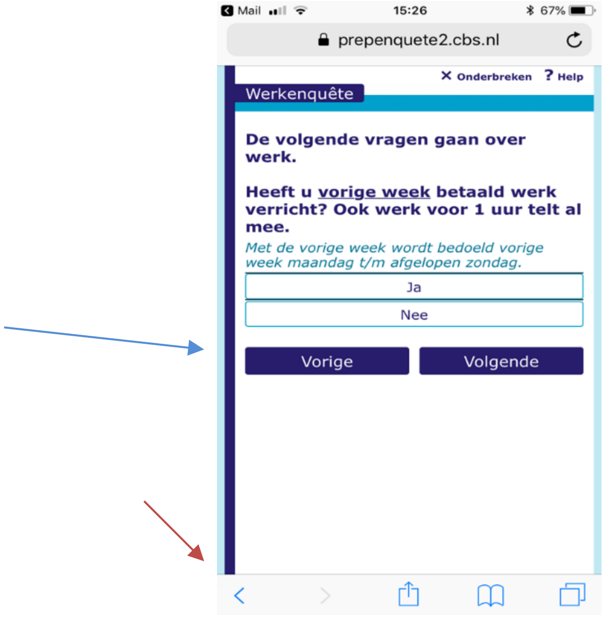
**Hoeveel uur werkt u gem bij vm?**

*Kan worden opgegeven per week maand.*
*U mag ook een schatting geven.*

22 uur per maand ▾

○ Geen antwoord

*Figure 5.3: Navigation issue: respondent use back arrow '<' instead of back button 'vorige',*

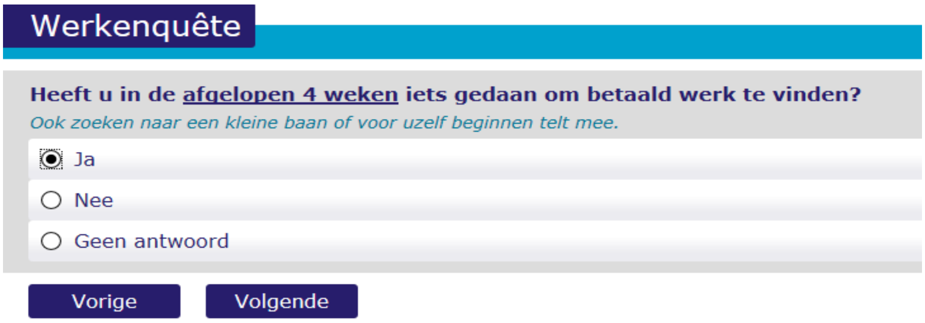


**De volgende vragen gaan over werk.**

**Heeft u vorige week betaald werk verricht? Ook werk voor 1 uur telt al mee.**

*Met de vorige week wordt bedoeld vorige week maandag t/m afgelopen zondag.*

*Figure 5.4a: Screenshot of  LFS question "In the 4 weeks ending last week, have you done anything to find work?"  Answer categories; Yes, No, Cannot say*



Werkenquête

**Heeft u in de <u>afgelopen 4 weken</u> iets gedaan om betaald werk te vinden?**

*Ook zoeken naar een kleine baan of voor uzelf beginnen telt mee.*

● Ja

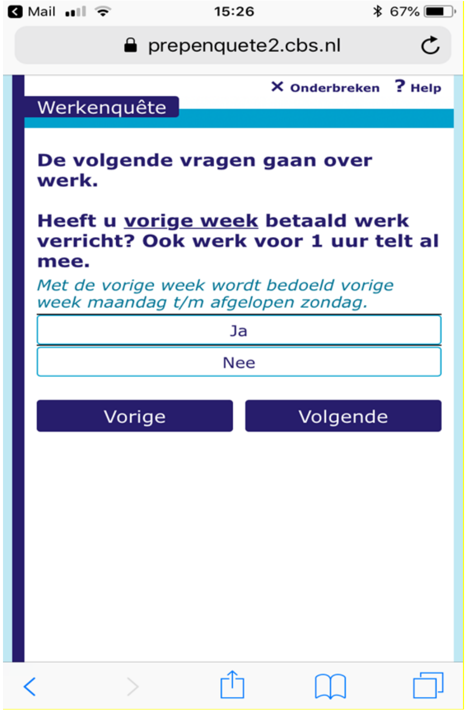
○ Nee

○ Geen antwoord

## Criterion Item introductions/instructions

Survey items may have an introduction text to explain terminology and conditions and to provide instructions to derive answers. Long introductions require more screen size. On smartphones introductions are often shortened, placed on a separate screen, hidden behind help buttons or avoided completely by changing wording of the questions. In the CBS LFS questionnaire the instruction is included under the question texts. By presenting this text between the question and the answer options we hope to improve the likelihood that respondents notice the text. See figures 5.4a and 5.4b for examples.  The instruction text is usually presented in italic, and in a different colour, light blue. This is for example in figure  5.4a, the instruction "This includes looking for a job of only a few hours or any activities to start a business". Additionally, the reference period of four weeks is emphasized with underlining. In this design of the LFS questionnaire no additional instructions are included using clickable info buttons or the question mark as research pretends those are not used or read or on a smartphone covers the original question and answer texts.  Also the instruction text is the same for all the devices. The introduction text, as for example the first two lines in figure 5.4b, is given in the same letter font and colour with one white space between the question text.

*Figure 5.4b: Screenshot of  LFS question "In the week from Monday the [date] to Sunday the [date], have you done any work for pay or profit?"*



**De volgende vragen gaan over werk.**

**Heeft u <u>vorige week</u> betaald werk verricht? Ook werk voor 1 uur telt al mee.**

*Met de vorige week wordt bedoeld vorige week maandag t/m afgelopen zondag.*

Ja

Nee

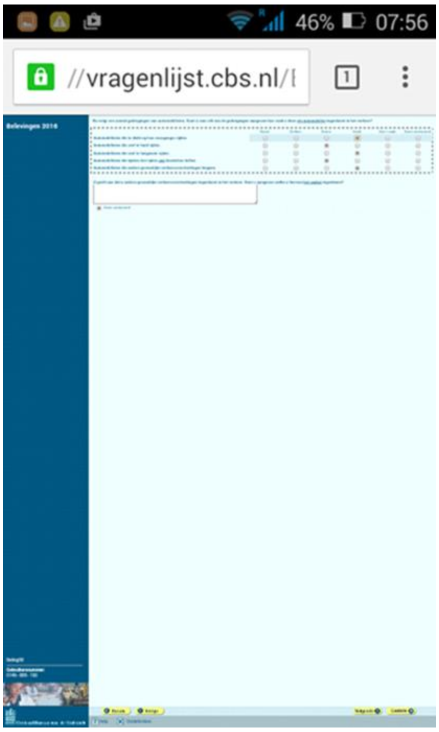
*Findings introductions/instructions*

1. In all devices, we found that respondents often did *not* read/notice the blue instruction texts. This seemed to be more often the case for respondents working on a smartphone. Partly not reading the instruction texts has to do with the operationalization, where the instruction has small light blue lettering. Many respondents indicated that the font just was not legible enough. This may be easily remedied. Partly, however, skipping the introduction also had to do with respondents' feeling that the question was posed, they understood the question, and found that the answers provided sufficient context for their answer. As the question text was comprehensible in their opinion they switched right away to the answer buttons. As one respondent indicated: *"I read the instructions if I have trouble finding the right answer"*. We observed that the instruction to the open question on occupation and work contents was read by respondents. Although it is in italic and blue like the instructions with other survey questions in the LFS with these specific questions, there was no context to be claimed from the answers. As this instruction with the occupation question is quite long, and fills half of the screen (see figure 5.9 and  part describing finding open answer fields). That may also have improved visibility.
2. In some cases, not reading the instruction lead to the wrong answers. If it is imperative that the instruction is read, it would be advisable to incorporate the instruction in the question text. As this LFS question together with the question on working hours is crucial to collect labour market data the risk of underestimation is quite high as the concept of last week can vary significant between respondents, as shown by research of Campanelli, Martin & Rothgeb (1991).
3. Stressing words or phrases with an underline was not sufficiently clear. As seen in figure 5.4b, the line is too subtle. If the underlined phrase is just above the answer boxes, the problem is exacerbated.
4. Even more care needs to be given to the need and placement of introductions. Some introductions can be suppressed, other introductions, especially if they introduce long questions or questions with may answers, could perhaps better be on a separate screen. We tested this in the grid question on cloud use and overall, respondents seem to (quickly) look at this introduction screen and did not comment negatively about a separate introduction screen.

### Criterion Grid questions

Grid questions are a series of survey items with the same answering categories on a relatively similar topic. On larger screens they are presented as a whole and the labels of the answering categories are shown only once. Grid questions form a block of items that demand more screen size or a different type of navigation. See figure 5.5 for an example for smartphone in a non-responsive design. The presentation is not readable, and has a risk that if a respondent tries to zoom in, answer categories and text are not visible anymore. As described in section 3.5, we tested 4 options for grid designs for smartphones; a vertical scrolling version without stemfix (for the LFS looking for work question), stemfix vertical scrolling-by-respondents, stemfix autoscroll and paging.

*Figure 5.5: Example of grid questions in non-responsive design*



*Findings comparing grid options by respondents*

In the test, we explicitly discussed the four options as developed for grid questions 2-7 (the ICT and environment questions, see table 4). When comparing these options, nine respondents preferred the classic grid presentation (option 1); four respondents preferred fixed-stem scroll by respondent (option 2); one respondent preferred the fixed-stem autoscroll (option 3) and two respondents preferred the paging design (option 4). One respondent could not choose between option 1 and option 2, for three respondents the overall preference was not assessed. During the test we noticed that the stem-fix autoscroll option clearly confused and disturbed many respondents. This was understandable, as the scrolling went so quickly they did not understand immediately what happened. Many found this option really annoying. The other options (the classic grid, stem not-fixed, the fixed-stem scrolling by respondent and the paging versions) all seemed to work rather intuitively.

Reasons mentioned for preference for classic grid [2]

- The classic grid is clearer and better structured / more overview (7)
- General preference for reading from / working on larger screen (3)
- Good that you do not have to scroll (2)
- Goes quicker that completing on phone (2)
- Easier for comparing answers between items (1)

[2] Note: one respondent may provide several reasons and may also provide input on several options. Reported are statements made by respondents when asked why they prefer a certain format over another. Later in the interview more respondents mentioned they prefer a larger screen over a small screen, but this point is only listed here for the respondents who explicitly mentioned this as a reason to prefer the classic grid.

- You have to read less than on the phone (1 –probably means that for the phone options the answer options are written out for each item).
- Distinction between individual questions clearer than the phone options, not the feeling you are looking at the same question you just answered (1)
- Classic grid is what you are used to, what you expect (1)
- Prefer horizontal presentation of answer options (1)
- In horizontal presentation easier to find midpoint than in vertical presentation (1)
- Easier to select answer options (1 – comparing to autoscroll version on the phone).

Reasons mentioned for preference for phone presentation over classic grid:

- On the phone the question in clearer and better structured / more overview (3)
- Easier to find midpoint on phone presentation (1).
- On the phone you have to judge each item, you take it more seriously (1)

Reasons mentioned for preference for stem-fix scrolling by respondent option over paging

- Less “clicking” (2)
- Clearer and better structured /more overview (1)
- Easier to change answer (1)
- Distinction between individual questions clearer, not the feeling you are looking at the same question you just answered (1)

Reasons mentioned for preference stem-fix scroll by respondent option over stem-fix autoscroll

- Prefer to control scrolling (3)

Reasons mentioned for preference over stem-fix autoscroll over paging

- Easier for comparing answers between items (1)

Reasons mentioned for preference for paging

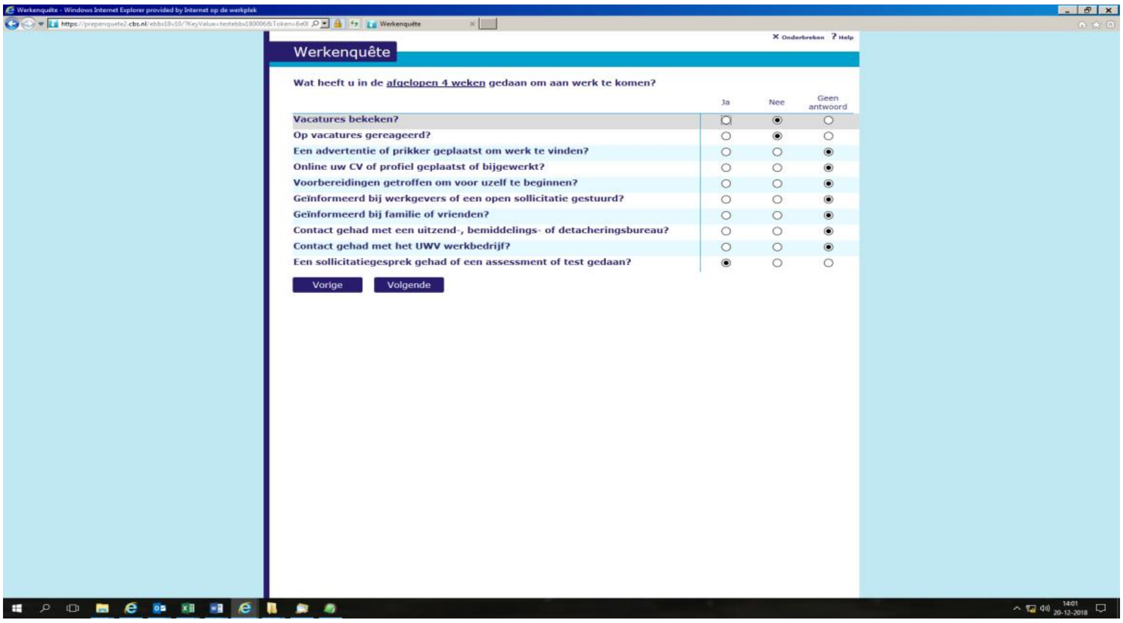
- Easier to focus on one question at a time (2)

*Findings using information in question stem when answering questions*

In the test we found that several times respondents did not use all relevant information from the question text when answering the questions. This was for example the case where in discussing the answers with respondents, it turned out they had not used the stated reference period or had not used the instruction to only report private ICT use. These errors occurred in all versions tested. However, as described in the literature and as also observed in this test, the likelihood of respondents reading and when necessary rereading the question text is affected by the amount of effort it takes for them to see the text.

Respondents typically focus their eyes and attention on the answer options. In the stemfix presentation and paging design for the phone, the question text is always close to these answer options. In the classic grid presentation as implemented for this test, the question text is vertically rather far from the answer options (see for example figure 5.6 below). In the smartphone grid question in the LFS (the looking for work question) respondents would even have to scroll back to always be able to read the question text (see figure 5.7 below).

*Figure 5.6: Classic grid presentation of LFS looking for work question for tablet and PC.*



✕ Onderbreken ? Help

Werkenquête

Wat heeft u in de afgelopen 4 weken gedaan om aan werk te komen?

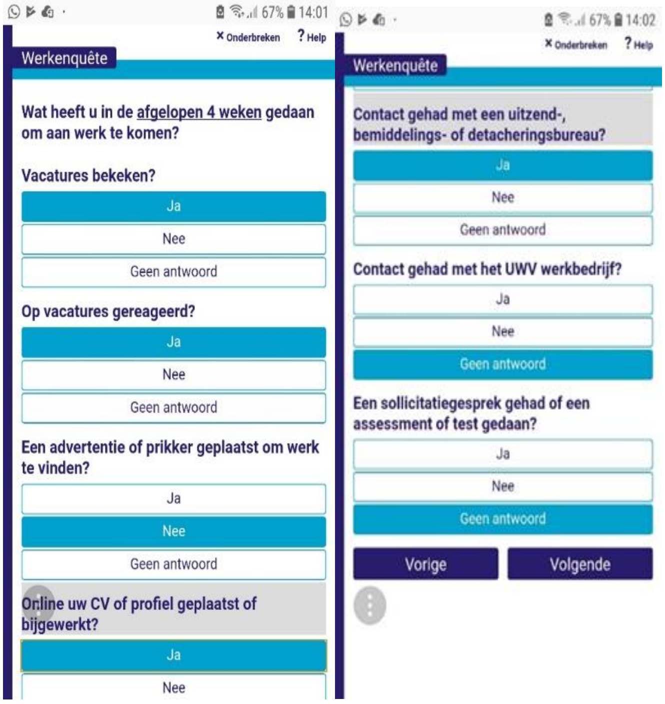
| | Ja | Nee | Geen antwoord |
|---|---|---|---|
| Vacatures bekeken? | ○ | ● | ○ |
| Op vacatures gereageerd? | ○ | ● | ○ |
| Een advertentie of prikker geplaatst om werk te vinden? | ○ | ○ | ● |
| Online uw CV of profiel geplaatst of bijgewerkt? | ○ | ○ | ● |
| Voorbereidingen getroffen om voor uzelf te beginnen? | ○ | ○ | ● |
| Geïnformeerd bij werkgevers of een open sollicitatie gestuurd? | ○ | ○ | ● |
| Geïnformeerd bij familie of vrienden? | ○ | ○ | ● |
| Contact gehad met een uitzend-, bemiddelings- of detacheringsbureau? | ○ | ○ | ● |
| Contact gehad met het UWV werkbedrijf? | ○ | ○ | ● |
| Een sollicitatiegesprek gehad of een assessment of test gedaan? | ● | ○ | ○ |

Vorige Volgende

*Findings ability to compare answers over items*

We observed for several respondents that they answered a specific item in relation to the other items. This was especially the case for the questions on how various actors would be able and willing to contribute to a better environment , for example:  "I think industry has more impact that my household". Also for the sharing of information via the internet we saw and heard respondents comparing answers to other items. For the general statements about the environment some respondents also seemed to adjust their answers to a specific item based on another related statement  (for example "I think it is good the government wants to improve our environment, but it should not cost me anything" and "I am willing to pay extra taxes to improve the environment").  As described above, some respondents also mentioned the ability to easily compare answers over items as a benefit of some of the tested options.

*Figure 5.7: Scrollable version of LFS looking for work question for smartphone*



67% 14:01

✕ Onderbreken ? Help

Werkenquête

Wat heeft u in de afgelopen 4 weken gedaan om aan werk te komen?

Vacatures bekeken?

[✓ Ja] Nee Geen antwoord

Op vacatures gereageerd?

[✓ Ja] Nee Geen antwoord

Een advertentie of prikker geplaatst om werk te vinden?

Ja [✓ Nee] Geen antwoord

Online uw CV of profiel geplaatst of bijgewerkt?

[✓ Ja] Nee

67% 14:02

✕ Onderbreken ? Help

Werkenquête

Contact gehad met een uitzend-, bemiddelings- of detacheringsbureau?

[✓ Ja] Nee Geen antwoord

Contact gehad met het UWV werkbedrijf?

Ja Nee [✓ Geen antwoord]

Een sollicitatiegesprek gehad of een assessment of test gedaan?

Ja Nee [✓ Geen antwoord]

Vorige Volgende

*Findings ability to recognize a new question*

In all tested options it happened that respondents did not immediately see that a new question had appeared. This resulted in extra burden (confusion, going back to see if they had already answered a question) and occasionally in an error, as they thought they had not provided their answer yet and unintentionally answered the next question. For the classic grid this only happened for the two questions on being able and on being willing to do something for the environment. These questions were almost identical. For the presentation on the phone this mainly happened in the paging design (and also in the auto-scroll option but probably caused by the fact they did not see the scrolling happen).

### Criterion Question length

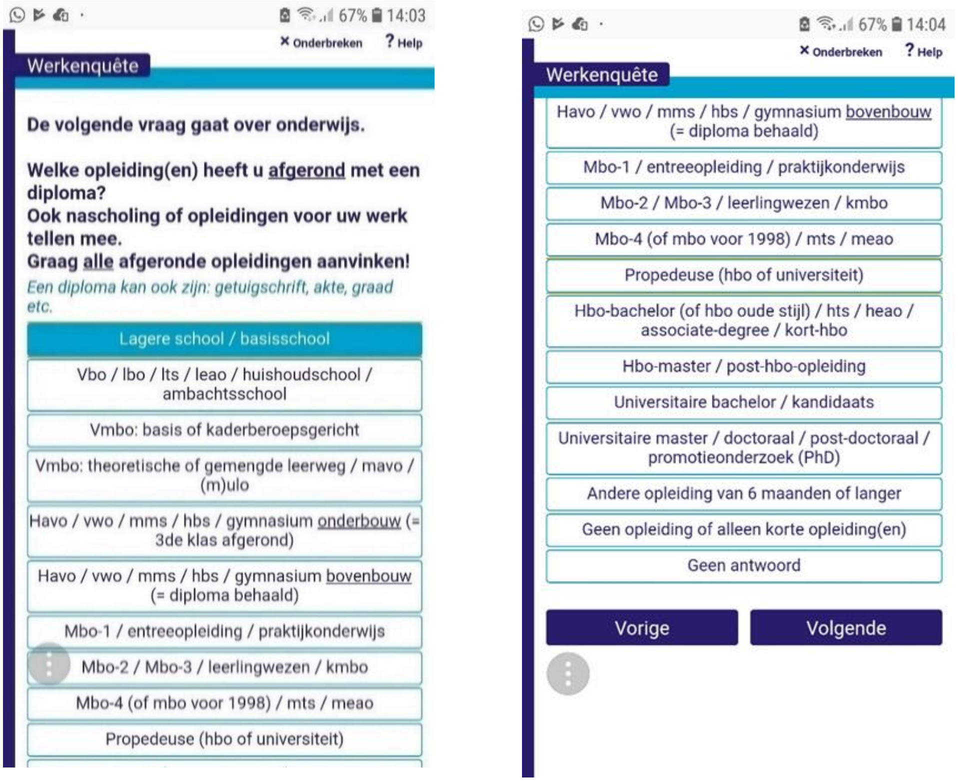
Survey items with longer question text demand more space. On smartphones question text are shortened or scrolling is needed. In general, the question text is shortened as much as possible. However, as also described on the criterion on the screen size, a question with a shortened text might in combination with a high number of answer categories look also quite busy on a smartphone screen. This is related to aim to include crucial content in the question text of the instructions to get the same stimulus to each respondent independent of the device used. Also this is related to needed content of the survey questions and variables that are based on the Model Questionnaire of Eurostat.

### Criterion Number of answer categories

The number of answer categories determines the required screen size of the item. On smartphones, items may be split into multiple items by introducing a hierarchy in the answer categories, thus avoiding scrolling, or require more scrolling.

The question with the largest number of answering categories was the LFS question on attained education (16 substantive categories, plus one 'no answer' category. Some of the answers were also long, with multiple names for certain levels of education in one button. See figure 5.8 for an example.

*Figure 5.8 Question on educational attainment on smartphone screen (scrolling design)*



**De volgende vraag gaat over onderwijs.**

**Welke opleiding(en) heeft u afgerond met een diploma?**
**Ook nascholing of opleidingen voor uw werk tellen mee.**
**Graag alle afgeronde opleidingen aanvinken!**
*Een diploma kan ook zijn: getuigschrift, akte, graad etc.*

- Lagere school / basisschool
- Vbo / lbo / lts / leao / huishoudschool / ambachtsschool
- Vmbo: basis of kaderberoepsgericht
- Vmbo: theoretische of gemengde leerweg / mavo / (m)ulo
- Havo / vwo / mms / hbs / gymnasium onderbouw (= 3de klas afgerond)
- Havo / vwo / mms / hbs / gymnasium bovenbouw (= diploma behaald)
- Mbo-1 / entreeopleiding / praktijkonderwijs
- Mbo-2 / Mbo-3 / leerlingwezen / kmbo
- Mbo-4 (of mbo voor 1998) / mts / meao
- Propedeuse (hbo of universiteit)

- Havo / vwo / mms / hbs / gymnasium bovenbouw (= diploma behaald)
- Mbo-1 / entreeopleiding / praktijkonderwijs
- Mbo-2 / Mbo-3 / leerlingwezen / kmbo
- Mbo-4 (of mbo voor 1998) / mts / meao
- Propedeuse (hbo of universiteit)
- Hbo-bachelor (of hbo oude stijl) / hts / heao / associate-degree / kort-hbo
- Hbo-master / post-hbo-opleiding
- Universitaire bachelor / kandidaats
- Universitaire master / doctoraal / post-doctoraal / promotieonderzoek (PhD)
- Andere opleiding van 6 maanden of langer
- Geen opleiding of alleen korte opleiding(en)
- Geen antwoord

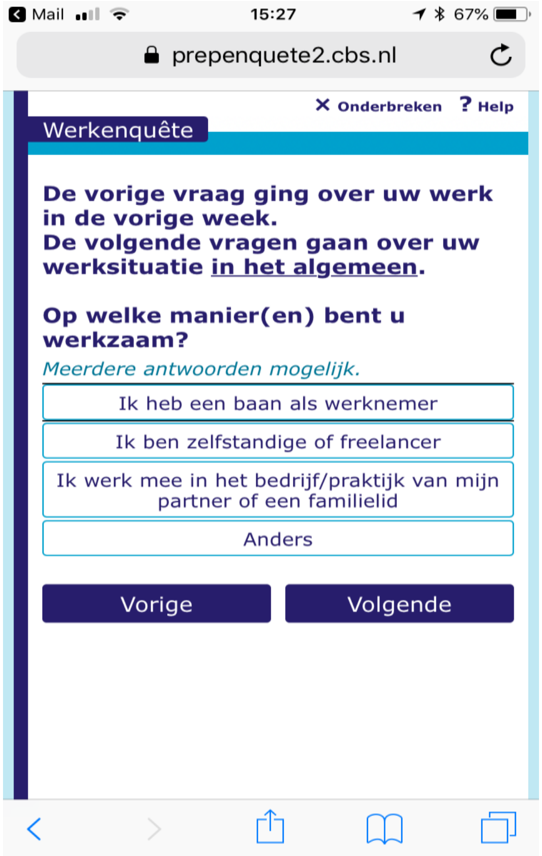
*Findings*

1. The vertical scrolling is no problem, the test respondents used this easily.
2. A number of respondents were confused and overwhelmed by the large number of answers, and indicated to have ‘just picked an answer’.
3. Respondents mention that compared with the other questions, this question text itself, plus the introduction is long. The fact the all finished education should be selected, was not read by *half* of the respondents.
4. Respondents overlooked their education. Also it was not clear and not noticed that they need to fill in all their educations they had finished with a degree
5. Some abbreviations appeared not to be clear or known.
6. It was not clear to several respondents that the last two options are single choice (whereas the other options are ‘select all that apply’). This combined with the fact that meaning of the last-but-one option (‘No education or only short education”) was not clear for everybody, caused several respondents to unintentionally deselect all previously selected options when choosing that option.

### Criterion Answer category length

Survey items with longer answer category text demand more space. On smartphones answer category text are shortened or made smaller in order to avoid overly large buttons. See the example in figure 5.9. The third answer option “Working in the business owned by a partner or family member” has a lager answer category box due to answer question length.

*Figure 5.9: Screenshot of different answer category text.*



**De vorige vraag ging over uw werk in de vorige week.**
**De volgende vragen gaan over uw werksituatie in het algemeen.**

**Op welke manier(en) bent u werkzaam?**

*Meerdere antwoorden mogelijk.*

Ik heb een baan als werknemer

Ik ben zelfstandige of freelancer

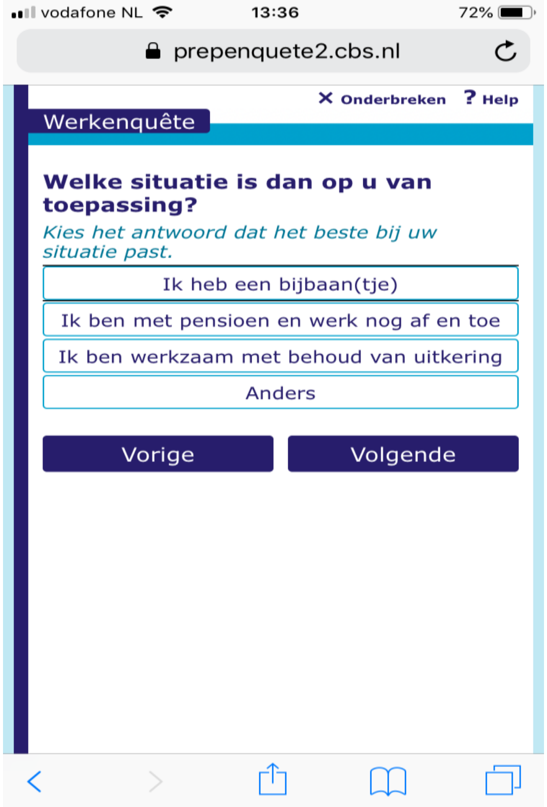
Ik werk mee in het bedrijf/praktijk van mijn partner of een familielid

Anders

*Findings*

1. The size of the button on the longer answer is preferable to the smaller buttons, as larger buttons are easier to select. But we should be careful to use a larger button to use more text; more text on a button, especially when more than one line is used, seems to reduce the readability of the text.
2. The context of the questions appeared to be very important. The question on the screen in Figure 5.9 shows the question asking for the kind of work situation, answer categories; 1) working as an employee 2) being self-employed, 3) working in business owned by family or partner and 4) other. One respondent needed the context of the follow up question to know if she had to answer that she was working as an employee or not. Eventually, she chose the last answer category 'other' (button 'anders') and subsequently, the next screen appeared (see Figure 5.10) with additional categories for the work situation like 1) working in a side job 2) being retired but work sometimes and 3) work with retention of benefit. Here, for this respondent the relevant category is included 'working in a side job (in Dutch: bijbaantje) that reflects correctly respondents working situation ( respondent is a student with a side job for a couple of hours a week in the supermarket). In that particular case, putting the question on the same page would be advisable and maybe necessary.

*Figure 5.10: Follow up question on work situation additional categories*



**Welke situatie is dan op u van toepassing?**

*Kies het antwoord dat het beste bij uw situatie past.*

Ik heb een bijbaan(tje)
Ik ben met pensioen en werk nog af en toe
Ik ben werkzaam met behoud van uitkering
Anders
Vorige
Volgende

### Dimension Touch navigation

In all cases, the touch navigation criteria evaluate the conflict between visibility on the screen and the simultaneous need to use the screen for navigation. Such navigation may lead to typing errors and respondent fatigue. In a smartphone optimized design, visual design features in the lay-out that aims to stimulate respondent in a user friendly way to fill in the questionnaire on a smartphone. However, due to a small screen and the touch screen options, there is a risk of unintended selections, resulting in switching to a next screen or clicking on non-applicable answers.

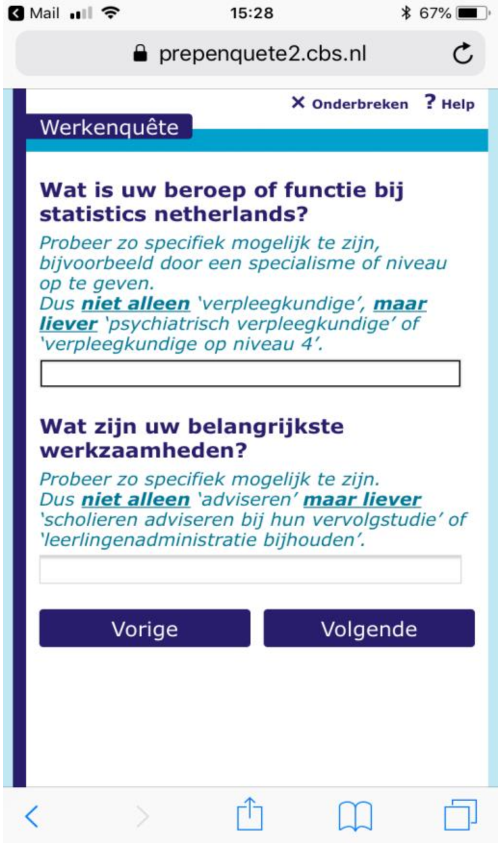
### Criterion Open questions

Open questions require typing in the answer. For smartphones, a keyboard will appear which may overlap with the survey item. Furthermore, the open question text box needs to be touched first.

An example is presented in figure 5.10. Two questions on one screen with open text field to fill in name of occupation and profession. In the instruction (in blue) information of being specific and elaborating on the profession, so for example not methodologist, but survey methodologist. In this LFS pilot a research goal is to see whether this yields the same quality of data used to code occupations in the Netherlands and international CASCOT.

Two versions of this question were tested. All respondents with this question on the route first answered the question with a shorter instruction, see figure 5.11, upper part. They were subsequently shown the alternative with a longer instruction, see figure 5.12. Because of this set up and because of the small number of respondents, it could not be determined if a longer instruction would have led to more elaborate answers. Out of 16 respondents, 9 indicated that they preferred the *longer* instruction, because of the larger number of examples, but also because the table format is more easily legible than just text. 5 respondents indicated that they preferred the shorter version, as it was more compact, and already made sufficiently clear what the intention of the question was. The implication of using the longer version, is that the two questions on occupation and  work duties need to be on separate screens. A quantitative test should determine if the longer text also leads to more elaborate answers. Further results from the pilot are expected in the next months. Another important finding with these questions is that the open answering field should be larger to invite elaborate answers. For some respondents it was not clear they could type more text than the size of the answer box. Other respondents took the size of the box as an indication of the number of words that was desired for answering this question.

*Figure 5.11: Version 1 Screenshots of open questions asking for occupation name and working activities on one screen*



**Wat is uw beroep of functie bij statistics netherlands?**

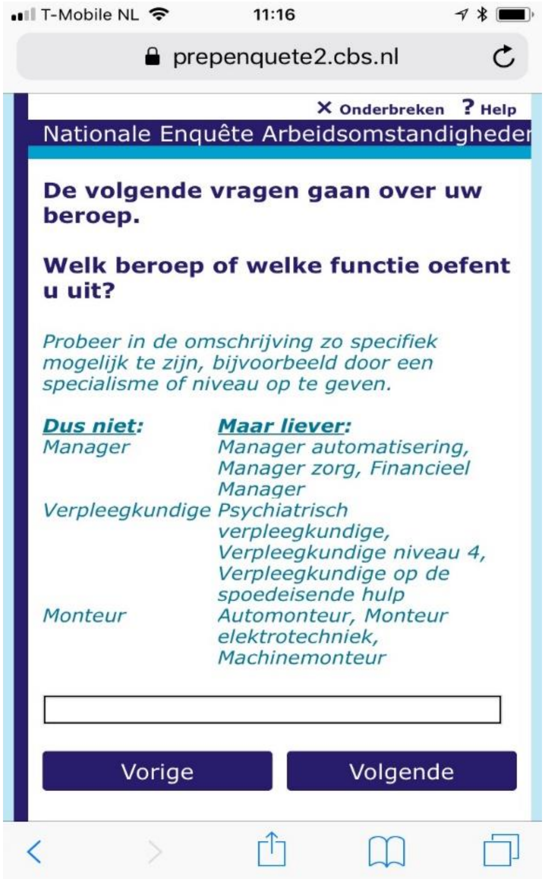
*Probeer zo specifiek mogelijk te zijn, bijvoorbeeld door een specialisme of niveau op te geven.*
*Dus* ***niet alleen*** *'verpleegkundige',* ***maar liever*** *'psychiatrisch verpleegkundige' of 'verpleegkundige op niveau 4'.*

**Wat zijn uw belangrijkste werkzaamheden?**

*Probeer zo specifiek mogelijk te zijn.*
*Dus* ***niet alleen*** *'adviseren'* ***maar liever*** *'scholieren adviseren bij hun vervolgstudie' of 'leerlingenadministratie bijhouden'.*

The lay-out of figure 5.12 is an alternative visual design that has a more elaborated instruction on giving a description of the occupation and is visually more structured. This version was included in the usability testing. Also, the questions on occupation and working activities are divided over two screens instead of a combination of two

*Figure 5.12: Version 2; Question on occupation with elaborated version of instruct text and visual more structured.*



**De volgende vragen gaan over uw beroep.**

**Welk beroep of welke functie oefent u uit?**

*Probeer in de omschrijving zo specifiek mogelijk te zijn, bijvoorbeeld door een specialisme of niveau op te geven.*

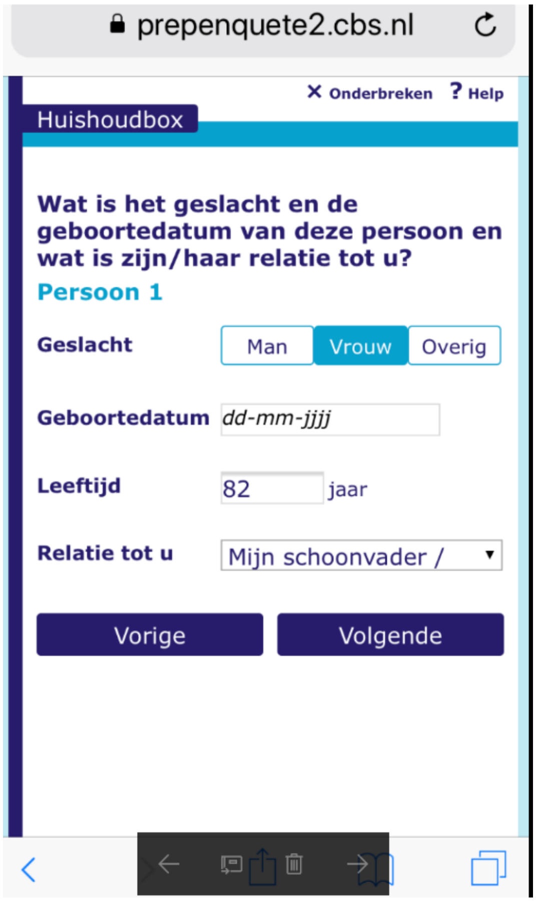
| ***Dus niet:*** | ***Maar liever:*** |
|---|---|
| *Manager* | *Manager automatisering, Manager zorg, Financieel Manager* |
| *Verpleegkundige* | *Psychiatrisch verpleegkundige, Verpleegkundige niveau 4, Verpleegkundige op de spoedeisende hulp* |
| *Monteur* | *Automonteur, Monteur elektrotechniek, Machinemonteur* |

*Findings*

1. An important results of the test was that almost all respondents mention that the box for texting an answer was too small. Respondents are not invited to use a lot of text. In the pilot LFS this is evaluated in the field whether there is a risk of device effects using smartphone in measurements of occupational status.
2. Also we observed that the keyboard covers the answering field and the help instruction text.
3. For some respondents, typing the open answers on the smartphone was a struggle. These respondents would most probably not have chosen to do a survey on smartphone, however.
4. The version in figure 5.12 got more preference form the test respondents; *"it is more elaborate, more clear as more examples are given".* But it requires more scrolling, and do you choose for clear communication that supports the definition of the response task or choose for compacter formulating reading in a sentence at once with clear usability that the survey is easier to fill in?".
5. For the household roster questions tested, we found that typing in a birthdate was very challenging for respondents, especially on the touchscreen devices. This was caused by several implementation decisions, such as forcing respondents to provide dates in one field not helping them to provide the correct format, asking them to use the "-" sign and an error message that was not very informative. However, a general

finding seems to be that it is demanding for some respondents if they have to use a keyboard on touchscreen devices to enter numbers and especially specific  symbols such as "-" that respondents do not regularly use and may not even be accessible easily on their keyboards (i.e. they should switch to another keyboard to be able to use the required sign).

*Figure 5.13: Household roster screen question on birth date format date-month-year*



**Wat is het geslacht en de geboortedatum van deze persoon en wat is zijn/haar relatie tot u?**

**Persoon 1**

**Geslacht** ○ Man ● Vrouw ○ Overig

**Geboortedatum** dd-mm-jjjj

**Leeftijd** 82 jaar

**Relatie tot u** Mijn schoonvader / ▾

## Criterion Items with many answer categories

An often applied solution to survey items with many answer categories is the drop-down box which requires scrolling to search the right answer. Such scrolling can be (partly) avoided by typing in the first letters of the answer (auto-complete). For smartphones, such solutions  demand for navigation on the touch screen which can be cumbersome.

In the version of the LFS tested there were no drop down boxes with many answer categories included.

## *Duration, Relevance and Burden and device*

Enjoyment-relevance-burden: Surveys differ in their general enjoyment-relevance-burden scores to respondents. For online surveys, response rates may vary from 15% to 45% with the exact same data collection strategy in terms of invitation and reminder letters, text messages or emails. Such large differences express the perceived enjoyment-relevance-burden ratio to the general population. For smartphones, surveys that score weaker are at larger risk.

### *Effort needed to read and answer survey questions*

Several respondents confirmed what is known from the literature: it takes more effort to read information from a small screen. Crucial instructions such as a reference period or ‘only include private internet use’ seem to be read and applied less by respondents using smaller screens. The required effort to find relevant information is higher if this information is outside the viewable space.

In the test we saw serval examples of how respondents used answers provided on previous items to decide on an answer (especially in a set of questions on how well various actors were willing and able to contribute to solving environmental issues). How accessible previously answered questions are depends on the size of the screen and the implemented layout.

We saw for most respondents, even for younger ones, that touchscreen devices and especially smartphones seem to require more effort from respondents. This was partly caused by the fact that the touch targets in our questionnaires were too small. Respondents had to be very careful to select the right option.

Contrary to the above, the classic grid question as presented on the large screen of tablets and PCs/Laptops seemed more demanding for some respondents than the presentation we tested for the smartphone. Understanding a grid and selecting an answer on the correct place is quite challenging for some respondents. Also, depending on the actual implementation of a questionnaire, the usability of certain aspects on the large screen may be worse than on the small screen. In our tested questionnaires the text lines seemed to too wide for comfortable reading (respondents had to move their head horizontally to read), the vertical distance between stem and item was very large in some grid questions Also, on the large screen, the navigation buttons were too much to the left – too far away from where the mouse and the eyes are when moving to the next button.

### *Clarity of response task*

On the smartphone it was not always clear if it was check all that apply or a single answer format, this creates additional burden and more risk of reporting error.

*Table 5.3: Response, device us and break off LFS pilot 2018*

| | | *Break off* | | | *Response* | *overall 36,9%* |
|---|---|---|---|---|---|---|
| | | *Log in* | *Break off* | | | |
| Device | | | | % | | % |
| | *Smartphone* | 716 | 49 | 6,8 | 667 | 20 |
| | *Tablet* | 558 | 36 | 6,5 | 522 | 16 |
| | *Other device* | 16 | 1 | 1 | 15 | 1 |
| | *PC* | 2215 | 96 | 4,3 | 2119 | 63 |
| | | 3050 | 182 | | 3323 | |

With respect to burden, and relevance, in our test respondents in general seemed to fill in the LFS rather easily and said questionnaire was OK for them, especially for those with a general employed situation. For people without work or others situations like combining one or two jobs some questions where cognitively more burdensome. Also, for non-working respondents the line of questioning may convey that they should be working and should explain why they are not – in at least two instances.

In 2018, CBS conducted a pilot with the LFS persons household and the questionnaire as tested for WP5. As in table 5.3, the overall response on the LFS was 36,9% from which mobile devices responded for 36%

(smartphone 20 % and tablet 16%).The average duration times was 6,5 minutes. The average break off was 5,9% (see Bakker & Robberts 2018).

# 6 Conclusions and recommendations

We performed tests on two surveys, the ICT and LFS. We briefly summarize findings, provide a list of recommendations per fitness dimension, list possible design actions, and end with a look ahead.

## 6.1 Summary of tests

The ICT test questionnaire, optimised for smartphones, performed satisfactorily on mobile for the test group at SSB. This group was young adults, in general higher educated and they expressed that mobile was their preferred device to use. SSB did not directly test preferred device, but several test persons stated that they in a regular situation would try to open survey links on their mobile first. (The reason for this could be that standard data collection procedures for SSB and other data collectors in Norway is to send text messages with direct survey links. Password or 2 step authentication solution is most commonly used with code sent by SMS to mobile as well. This set up makes mobile the most convenient device to use, and SSB has seen a mobile device completion as high as 80 % on some surveys.) With the high mobile penetration in Norway, it is fair to assume that this will be the general inclination in large parts of the population. Further, the tests showed that speed, navigation and usability is expected to function without problems and delay. The expectation of flawless technology increases with development. What was acceptable in terms of loading time, usability etc. five years ago no longer is, and we need to meet these expectations. We have learned from the tests that there are several measures that can be taken to reduce response burden and increase relevance for the respondents. For future revisions of the ICT survey we urge that terminology to be updated, to think “mobile first” and to design a one screen question solution. For the future we suggest running similar tests for a broader spectre of respondent groups, particularly also less proficient smartphones users.

The smartphone-optimised LFS questionnaire was tested at CBS with test respondents with different characteristics according to age, gender, education and employment status. For most respondents the questionnaire performed good and test respondents in general could easily fill in the LFS questionnaire, especially those with a general employed situation. Observation shows that even for younger ones, touchscreen devices and especially smartphones seem to require more effort from respondents. This was observed in the usability test of the LFS questionnaire, but also in the related tests of the household roster and grid questions. The observations also confirmed what is known from the literature: it takes more effort to read information from a small screen. Crucial instructions such as a reference period seem to be read and applied less by respondents using smaller screens. The required effort to find relevant information is higher if this information is outside the viewable space.

## 6.2 Recommendations mobile device questionnaire design

Based on existing literature (e.g. Antoun et al. 2017a, 2017b; Bakker, 2018; Nielsen & Budiu 2013) and based on evaluations of the fitness criteria using results from the usability tests, we make a number of recommendations for each dimension: screen size, touch navigation and duration. In addition, we make a number of more general recommendation not necessarily related to mobile devices. Some recommendations are repeated as they apply to more than one design aspects.

## Recommendations dimension Screen size

*Introductions and instructions:*

1. Carefully consider if and which types of different font should be used. The font differentiations used in our test did not perform well. Possibly instructions texts should be included in the same font, not bold and not italic and same colours as the question text.
2. Reduce long texts in introductions and instructions.
3. Consider to rephrase very important instruction in the form of a question

*Grids & #items per grid:*

4. Main recommendations for grid questions
   a) For the smartphone: the best of the tested options seems the fixed-stem scroll by respondent option as it:
      - improves likelihood respondents read relevant question text
      - facilitates comparing answers over items
      - for comparing answers over items / consistency between items: all items scrollable on one page probably decreases device effects when combining classic grids for larger devices and phones
      - fixed-stem and autoscroll may work if implemented in a way respondents immediately see what happens

   b) To improve the likelihood that respondents use relevant information in the question and to prevent device effects, it is recommended to also redesign the classic grid question for larger screens. This should probably be done in a way that a) the question stem is very close to the answer options and b) the answer options are directly below the item (and not presented as headers of columns as in the classic matrix presentation) c) it is easy to see answers on previous items.

   c) For all presentations of questions that are very similar (overlap in wording and same answer options) make it easier for respondents to see that a new question has appeared. This can be done by numbering (using for example item 3/6 when presenting items of one grid) and visual presentation (using colors/font to highlight differences, animation of changing of screen).

   d) Limit the number of sub questions in questions batteries and make sub questions visually distinct using numbers, letters.

*Question text*

5. Reduce long texts in questions.
6. Avoiding repeating text from screen to screen especially when it is a lot of text.
7. Limit the number of sub questions in questions batteries and make sub questions visually distinct using numbers, letters.

*#answer categories*

8. Reduce number of answer categories, for example by combining options or by hierarchical displaying.

*Answer text*

9. Reduce long texts in answer categories and avoid repetition of question text.

## Recommendations dimension Touch navigation

10. Develop responsive questionnaires that also work well on smaller / older phones and test it.
11. Develop responsive questionnaires that also work in portrait for tablets and test it.

12. Avoid horizontal scrolling.
13. Avoid vertical scrolling in combination with auto forward, as for questions with many answer options this may led to a situation in which the respondent does not see all answer options.
14. Develop questionnaires that also work well on smaller / older phones.
15. Develop questionnaires that also work in portrait for tablets.
16. Make sure font size is large enough to read easily. . Antoun et al. (2017b) refer to industry guidelines recommending font size (17-18 pixels (4,8 mm high).
17. Make sure touch targets can be easily selected. Antoun et al. (2017b) refer to industry guidelins recommending a size of touch targets of about 8 mm in length and width.
18. Make sure there is enough space between touch targets as used in general guidelines.
19. Utilize design and navigation to avoid confusion about which answer is selected.
20. Check for consistency in visual language (e.g. alignment) and effects in the design for various devices.

*Open question*

21. Visible lay-out of open text fields on a smartphone should support and motivate the respondent to filling in more text (display the maximum number of words in the text window).
22. Open text fields should be big and marked with a cursor and keyboard function.

*Many answer*

23. Reduce number of answer categories, for example by combining options or by hierarchical displaying.
24. Avoid repeating text from screen to screen especially when it is a lot of text.
25. Limit the number of sub questions in questions batteries and make sub questions visually distinct using numbers, letters.
26. Use design that distinguishes between single and multi-answer questions. Further research is needed for this, possibly the use of bullets and checkboxes may help to visualize the difference
27. Avoid using “other, please specify” after long questions batteries with numerous categories

### Recommendations dimension Duration

28. Prevent lengthy questionnaires especially for the smartphone.
29. Avoid repeating text from screen to screen especially when it is a lot of text.
30. Potential options for the use of visual design features like icons and pictograms should be researched and tested how to increase motivation, ease the response task and decrease burden on a smart phone.

Some findings in the usability testing of the LFS and ICT and evaluation of other ESS surveys gave input to recommendations about the content of the measurement of variables. As in some cases these findings were device specific. Device specific measurement problems may arise when there is interference between the response task, the burden and visual lay-out and touch navigation aspects that may introduce a potential risk for measurement error. For example, the vertical distance between an instruction text and the answer options may be different for large screens and small screens, leading to different costs for the respondent to process that instruction. Testing and monitoring can help to attend device and mode specific measurement risks.

### Additional general recommendations

31. Avoid long recall periods and attempt to make such period appropriate for what you need information about .
32. Avoid using “other, please specify” after long questions batteries with numerous categories as the overview is missed.
33. Collect paradata to monitor device use to allow for analyses of possible device effects.

34. Conduct usability pre-testing on all devices/browsers to monitor risk of possible device effects.
35. Use relevant, intuitive and understandable terminology to prevent device but also mode effects as explanation from interviewers is lacking.

### 6.3 Possible design actions to improve mobile device fitness

In line with the recommendations of section 6.2, we list possible design actions per dimension and fitness criterion.

*Table 6.1: Possible design actions for the fitness criteria*

| *Dimension* | *Criterion* | *Design options* |
|---|---|---|
| Screen size | Introductions | • Limit amount of and text length of instruction and introduction text<br>• Position introductions on separate pages with separate back-forward buttons<br>• Include shortened instruction that is crucial in the question text<br>• Make introductory information that is not essential for all respondents available behind expandable links that indicate which information may be found behind it. |
| | Grid questions | • Fix the stem of the questions and either allow for swiping between items on multiple pages horizontally or collapsing items vertically<br>• Allow for scrolling vertically (but avoiding horizontal scrolling)<br>• Use numbers, letters or visual features to distinguish between the different items<br>• More research testing and developments are needed for choices of most device friendly design |
| | Question text | • Shorten question texts<br>• Make clarifying information that is not essential for all respondents available behind expandable links that indicate which information may be found behind it (further testing needed) |
| | # answer cat’s | • Reduce number of categories<br>• Split into multiple questions, possibly hierarchically |
| | Answer text | • Shorten answer texts and avoid repetition from question text<br>• Make clarifying information that is not essential for all respondents available behind expandable links that indicate which information may be found behind it (further testing needed) |
| Touch navigation | Open question | • Attempt to replace open questions by closed questions<br>• Position open questions on separate pages<br>• Choose a visual design with an open text field with enough space where respondents see all the words or the last complete lines what they are typing in<br>• Introduce visualization to indicate completeness/richness of answer data with for instance colours, red, orange, green or indicate a maximum number of words |
| | Many answers | • Use auto-search or auto-fill<br>• Split into multiple questions, possibly hierarchically<br>• Remove response categories that are very rarely used, and add an “Other” option.<br>• Use numbers or letters to distinguish between the different items |
| Duration | # of items | • Reduce length of the survey<br>• Split into multiple waves in a panel design<br>• Consider to use a split questionnaire design with randomization |
| | Household | • Move from household to person survey |

| *Dimension* | *Criterion* | *Design options* |
|---|---|---|
| | Database | • Rely more heavily on processing and analysis afterwards<br>• Perform checks at data collection web server<br>• Further developments and testing is needed to evaluate the use and fill in of retro-data on usability and data quality. |
| | Complexity | • Reduce number of complex questions<br>• Inform respondents that using larger screens and quiet surroundings when filling in the questionnaire is advisable in the survey invitation/information materials Use contact modes that are compliant with access to appropriate response mode<br>• Research possibilities to make the survey more enjoyable (to some extent) |
| | Enjoyment<br>Relevance<br>Burden | • Make the survey attractive and appealing on a smartphone<br>• Utilize proven design of existing apps and websites<br>• Add value to the survey by afterwards feeding back information to the respondent |

## 6.4 Look ahead

We like to propose two general topics for future discussion within the ESS: mobile device first questionnaire design and questionnaire length of ESS surveys. We view these topics as beyond the scope of the current WP, but they naturally arise from an assessment of fitness for mobile devices, in particular smartphones.

We are advocates of a mobile device first questionnaire design, or, at the least, of a rigorous account of the mobile device option in questionnaire design. We have two main reasons for this. The first reason is that smartphones have become a dominant communication channel and cannot be ignored in design. The second reason is that issues with usability and comprehension on smartphones reveal the measurement error prone questions and question blocks. Such a viewpoint, however, has implications for ESS model questionnaires and ESS survey guidelines. Multi-device surveys introduce additional challenges for the questionnaire design.

An obstacle that is often put forward to introduction of new devices is questionnaire length. As most ESS surveys are long and, consequently, demanding when filling in on a smartphone, it is imperative to prevent 'speed' and 'stimulate' a relaxed manner leading to better quality and less measurement errors. A responsive design should facilitate the respondent filling in on a smartphone screen, however, most questions and answer texts are cognitively demanding due to their specific content or response task, as for example long reference periods. When reflecting on the fitness criteria and the experiences from the test interviews at CBS and SSB it is a harsh job to find a good modus in redesigning, i.e., responsive design and collecting valid and reliable data comparable over devices and modes (also interviewer based). This leads to a question for future discussion: Is it feasible to find possibilities to shorten/redesigning a ESS model questionnaire making it user friendly to fill in on a smartphone?

We recommend that these topics are addressed in both general discussion on ESS procedures and in specific working groups for ESS surveys.

# References

Android Developer’s Guide. (2016). Material design components. Retrieved from https://material.google.com/components/bottom-navigation.html

Antoun, C., Couper, M. P., & Conrad, F. G. (2017a). Effects of mobile versus PC Web on survey response quality: A crossover experiment in a probability web panel. *Public Opinion Quarterly, 81, 280–306.*

Antoun, C., Katz, J., Argueta, J. and Lin (2017b). Design Heuristics for Effective Smartphone Questionnaires *Social Science Computer Review, 1-18.*

Bakker, J. & Robberts, A. (2018). Mobile device login and break-off in individual surveys of Statistics Netherlands. Discussion paper. Statistics Netherlands.

Bakker, J., Wijnhoven, H. & van der Steen, M. (2018) Designing the questionnaire of tomorrow: Current best-practices and future goals. Presentation on IBUC conference Baltimore, October 2018. On request: j.bakker@cbs.nl

de Bruijne, M. (2015). Designing web surveys for the multi-device internet. Doctoral dissertation, Tilburg University, the Netherlands.

de Bruijne, M., & Wijnant, A. (2013a). Can mobile web surveys be taken on computers? A discussion on a multi-device survey design. *Survey Practice, 6, 1–8.*

Campanelli,P., Nicolaas, G., Jäckele, A., Lynn, P. Hope, S., Blake, M. and Gray, M. (2011). A Classification of Question Characteristics Relevant to Measurement (Error) and Consequently Important for Mixed Mode Questionnaire Design. Paper presented at the Royal Statistical Society, October 11, London, UK.

Couper, Mick P., Christopher Antoun, and Aigul Mavletova. 2017 ‘Mobile Web Surveys: a Total Survey Error Perspective”. In *Total Survey Error in Practice.* Edited by Paul P. Biemer, Staphanie Eckman, Brad Edwards, Edith de Leeuw, Frauke Kreuter, Lars E. Lyberg, Clyde Tcker, and Brady T. West, pp. 133-54, New York: John Wiley

Gravem, D. & Berg, N. (2019). Recommendations for key questionnaire elements, questions and question types in mixed mode settings. Deliverable 4.3 of the MIMOD project. January 31 2019.

Gravem, D.F., Holseter, C. Falnes-Dalheim, E., Signore, M., Luiten, A., and Meertens, V. (2018) Mixed-mode experiences of European NSIs. Deliverable 4.1 of the MIMOD project, June 22 2018.

Nielsen, J. & Budiu, R (2013). Mobile usability. Nielsen Norman group Berkeley.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pre-testing survey questions. *Journal of Official Statistics, 7, 349–365.*

Schouten, B., Blanke, K., Gravem, D., Luiten, A., Meertens, V., & Paulus, O. (2018). Assessment of fitness of ESS surveys for smartphones. Deliverable 5.1 of the MIMOD project. July 20 2018.

Statistics on the Usage of Information and Communication Technologies 2016, questionnaire improvements (2017). Final report WP5: Improving, designing and testing questions on e-commerce, e-mediaries and sharing economy for the ICT Household survey

Tourangeau, R., Maitland, A., Rivero, G., Sun, H., Williams, D., and Yan, T.; Web Surveys by Smartphone and Tablets: Effects on Survey Responses, *Public Opinion Quarterly*, Volume 81, Issue 4, 12 December 2017, Pages 896–929, https://doi.org/10.1093/poq/nfx035