**Assessment of fitness of ESS surveys for smartphones**

Barry Schouten (CBS), Karen Blanke (DESTATIS), Dag Gravem (SSB), Annemieke Luiten (CBS), Vivian Meertens (CBS) and Oliver Paulus (DESTATIS)

*Summary: WP5 of the ESSnet MIMOD (Mixed-Mode Designs for Social Surveys) 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 this first deliverable, we focus completely on the fitness for smartphones. We present a set of criteria related to smartphone screen size, smartphone navigation and interview duration, we apply the criteria to Eurostat model questionnaires and country-specific implementations of the EHIS, EU-SILC, ICT and LFS, and we end with recommendations for these four surveys.*

1. **Introduction**

ESSnet MIMOD started in December 2017 and runs until April 2019. It has five topical work packages: 1) mixed-mode data collection strategies, 2) mode-effect adjustment, 3) mixed-mode case management, 4) multi-mode questionnaire design, and 5) mobile device questionnaires and sensors. We focus on WP5.

Modes affect answering behaviour and may lead to method-specific measurement bias, item-nonresponse bias and break-off. In general, modes are disparate on four features:

1. Presence of an interviewer: Interviewers may motivate respondents and keep them concentrated, may assist in explaining terminology but may also cause social desirability bias;
2. Speed and pace of the interview: Surveys over the phone tend to go faster than in person which in turn tend to go faster than online and on paper;
3. Presentation of the survey items: The distinction between oral and visual presentation and the type of visual presentation may affect answers;
4. Timing of the interview: The flexibility to choose a time and location to fill in the questionnaire may affects answers;

We assess fitness of surveys for smartphones relative to traditional online devices, i.e. desktops and laptops. We assume that the online mode already is an optional mode in the survey and the questionnaire is fit for these traditional devices. For this reason, when considering the effect of devices we consider mostly the disparity due to presentation and due to timing of the interview. Obviously, the presence of an interviewer does not play a role when comparing devices. Apart from the difference in loading time of web pages on different devices, also the speed and pace are similar between devices. However, as mobile devices offer more flexibility in the location at which the questionnaire is completed, devices may affect the timing of the interview. The difference in presentation between devices is again obvious, mobile devices have smaller screen sizes and navigation is performed through the screen rather than through separate tools.

1. **From disparity to criteria**

We do not intend to come up with absolute criteria from which it can be concluded that a survey is fit for smartphones or not. Rather, we construct a number of dimensions, and within these dimensions a number of scales, that allow for a gradual assessment of fitness. If a survey questionnaire scores high on multiple scales, then there is a higher risk of device-specific measurement error and/or item-nonresponse. The criteria themselves should be as clear as possible. Nonetheless, we suspect that some country/culture differences will always exist when it comes to the impact or risk.

The difference in presentation and timing of the interview between devices amounts to a focus on screen size, touch navigation and duration when developing criteria.

* 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.
* Touch navigation: In all cases, the 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.
* Duration: Interview duration in combination with timing and location affect respondent motivation and concentration and break-off. Interview duration is deemed even more an issue for mobile devices due to the longer loading time of web pages and due to potential pauses in the internet connection.

Criteria should be as clear as possible, allowing for as little room as possible for subjective assessment. Fully objective assessment will not be possible, however, and in the assessment it is recommended to score a survey by at least two experts. We operationalize the various criteria as given in table 1.

*Table 1: Fitness criteria*

|  |  |  |
| --- | --- | --- |
| *Dimension* | *Criterion* | *Operationalization* |
| Screen size | Introductions | Number of items with introductionsNumber of items with instructions included |
| Grid questions | Number of gridsAverage number of items per grid |
| Question text | Number of items with > 20 words (excluding introduction text) |
| # answer cat’s | Number of items with > 5 answer categories |
| Answer text | Number of items with > 10 words in at least one category |
| Touch navigation | Open question | Number of open questions |
| Many answers | Number of items with > 25 answer categories |
| Duration | # of items | Total number of itemsAverage duration of survey per respondent |
| Household | Is survey a household survey? Yes/no |
| Database | Does survey require interaction with a database? Yes/no |
| Complexity | Number of (anticipated) items that require calculations by an average respondent, i.e. are cognitively burdensomeNumber of (anticipated) items that require consultation of personal documentation by an average respondent |
| Enj-Rel-Bur | Response rate to traditional online devices |

Initially, we also envisaged including the number of filter questions, where the follow-up questions appear on the same screen. There are however no criteria given by ESTAT on how to handle this, and we, therefore dropped this criterion

1. **The use of smartphones in ESS surveys**

Within WP1 of MIMOD, a survey was fielded to all 31 ESS NSI’s. One section of the survey concerned the possibility of completing ESS surveys on a smartphone. As a follow-up to this question, several other questions were asked. Seven ESS surveys were selected for this purpose within MIMOD. These surveys are:

* European Health Survey (EHIS)
* EU Survey on Income and Living Conditions (EU-SILC)
* ICT Survey (ICT)
* Labour Force Survey (LFS)
* Adult Education Survey (AES)
* Harmonized European Time Use Survey (HETUS)
* Household Budget Survey (HBS)

Table 2 shows the number of countries that have an online option and, if they have, whether smartphones can be used and to what extent the questionnaire has been adapted. A first observation is that the majority of questions do not have an online option. A second observation is that if a survey can be done online, then only a few countries made slight adaptations. Hence, smartphones do not yet play a dominant role in questionnaire design.

*Table 2: Possibility of using smartphones (Number of NSI’s)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Survey* | *No web* | *Blocked* | *Not adapted but usable* | *Slightly adapted* | *Profoundly adapted* |
| LFS | 25 | 1 | 5 |  |  |
| EU-SILC | 24 | 2 | 4 | 1 |  |
| EHIS | 20 | 1 | 10 |  |  |
| AES | 21 | 1 | 8 | 1 |  |
| ICT | 16 | 3 | 10 | 2 |  |
| HBS | 26 | 2 | 3 |  |  |
| HETUS | 31 |  |  |  |  |

We conclude that smartphones are yet a future option for the selected ESS surveys and not yet daily practice in the ESS countries.

1. **Selection of ESS surveys and model questionnaires**

For time reasons, but also because HBS and HETUS innovation and use of smartphone applications are considered in separate task forces, we restrict ourselves to four surveys: EHIS, EU-SILC, ICT and LFS.

For each survey, we decided to consider both the Eurostat model questionnaire and two country-specific implementations. We do this, because model questionnaires leave room for specific choices by NSI’s such as introduction texts to items or batteries of items, explanation of terminology, and, most importantly, extra country-specific questionnaire modules. The latter depends on the availability of administrative data and other, usually national, stakeholders that make use of the same survey data.

For the EU-SILC and LFS no model questionnaires exist, but, instead, Eurostat has disseminated a set of guidelines and recommended implementations. This implies that for these two surveys, there likely is more variation between NSI’s in their country-specific implementations. As a consequence, recommendations about smartphone fitness become more uncertain as country-specific versions of the survey may have properties, e.g. length or extra blocks of grid questions, which complicate smartphone interview. This has to be kept in mind throughout the following.

Each country-specific survey design of EU-SILC offers a different mix of survey (main) modes (interviewer- & self-administered, computer- & non-computer-assisted), which has also impact on some question designs too. Mostly interviewer-administrated modes and rarely self-administrated modes like PAP or CAWI. The total numbers of items differs from about 300 items to over 600 items. The variance of modes and number of items leads to a country-specific average duration of interviews ranging from 12 minutes to more than 50 minutes. Moreover there are some but not all countries using register-data to reduce the extent of primary data collection. In case of EU-SILC, several countries make us of national registers, when complex income- and tax-related variables are collected, so that response burden can be reduced by using other data sources as complement. All those differences arising from the flexible survey design of EU-SILC led to the need to check the country-specific EU-SILC survey with its individual survey design when assessing its fitness for smartphones. Keeping that in mind, the following assessment on the fitness criteria is based on the recommended EU-SILC guidelines and a detailed assessment of the German EU-SILC PAP-questionnaire, complemented by a brief inventory on EU-SILC country-specific designs within the MIIMOD consortium members. .

Table 5 shows the difference in design of the country-specific EU-SILC surveys within the ESSnet MIMOD consortium.

*Table 5:Selected, country-specific EU-SILC key survey indicators of surveys (MIMOD consortium-members)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *ESS-Countries (consortium)* | *Survey modes\** | *Total number of items* | *Avg. duration of interview(min.)\** | *Use of registerdata* | *Household- or individual-survey?* |
| DE | PAP | ~330 | 55 | No | HH |
| NL | CATI/CAWI | - | 18.4 | Yes | HH |
| IT | CATI/CAPI | ~530 | 12.1 | Yes | HH |
| NO | CATI | ~350 | 55.1 (2015) | - | HH |
| AT | CAPI/CATI | ~660 | 23.9 | Yes | HH |
| FR | CAPI | ~690 | 28.2 | Yes | HH |
| FI | CATI | ~320 | - | Yes | HH |
| PL | PAPI/CAPI | ~540 | 45.1 | No | HH |

\*Source: EU-SILC ESQRS 2016 Annex 6 (Reference Metadata in ESS Standard for Quality Reports Structure)

1. **Results for the selected surveys**

We discuss the results for the four selected surveys separately.

*6.1 European Health Interview Survey (EHIS)*

For the EHIS only two assessments are available, for the model questionnaire and for the specific implementation of the EHIS in Norway. The specific implementation at CBS is in preparation for 2019 and could not be scored yet. At CBS, the EHIS is implemented as a mix of web and CAPI. At SSB, it is implemented as CATI. The EHIS is usually integrated with an existing Health Survey that is on-going. The EHIS is fielded once every four years. Table 6 shows the scores on the fitness criteria. CBS assessed the 2019 model questionnaire and SSB the 2015 Norwegian EHIS.

Conclusions based on model questionnaire

* The review is done on the model questionnaire of the EHIS as it will be fielded in the Netherlands in 2019. The number of questions with instructions is a prominent feature. Also the items or answer options with additional instructions or examples is noticeable.
* For the introductions and instructions, is included both information to be read out loud for all respondents by the interviewers and including fully on the web designed questionnaire and clarification for the respondents.
* The inclusion of national questions complicates the assessment, both in terms of calculating average time, and regarding the questionnaire structure. Without the national variables, the number of questions and duration time would have been lower.
* There are a relative large number of questions that require calculation for determining the frequency of events or for consumption.
* Some questions have more than five answer categories.

Conclusions Norway:

* The number of filtering questions is the most predominant trait[[1]](#footnote-1). This appears to have been done to adapt the health variable descriptions from the EHIS regulation for an aural mode by portioning out the stimulus. In a visual mode, it might be feasible to merge filter and follow-up without ending up with too long questions or too many response categories.
* The number of questions with instructions is the second most striking feature. Here, we have included both information to be read out loud for all respondents, and clarification for either respondent- or interviewer-initiated dialogues.
* The inclusion of national variables complicates the assessment, both in terms of calculating average time, and regarding the questionnaire structure.
* 14 questions have more than 20 words. Questions where the response options were an integral part of the question and intended to be read out loud are included here.
* There are quite a few questions that require calculation, for determining the frequency of events or for the consumption of alcoholic beverages.
* Several of the questions *could* have been implemented as grids in desktop web.
* Purely transitional introduction texts are included.
* The full Norwegian EHIS including national variable appears too long for mobile web. The subject matter, with many sensitive questions, is an argument in favour of self-administration.

*Table 6: Score on fitness criteria for EHIS*

|  |  |  |
| --- | --- | --- |
| *Dimension* | *Criterion* | *Scores* |
| *Model* | *SSB* |
| Screen size | Introductions | 19 | 15 |
| Instructions | 23 | 40 |
| Grids | 0 | 0 |
| #item per grid | NA | NA |
| Question text | 1 | 14 |
| # answer cat’s | 5 | 4 |
| Answer text | 0 | 0 |
| Touch navigation | Open question | 0 | 0 |
| Many answers | 0 | 0 |
| Duration | # of items | 118 | 153 + 150 |
| Av duration | NA[[2]](#footnote-2) | 34 min |
| Household | No | No |
| Database | Yes | No |
| Cogntv burden | 29 | 21 |
| Consultation | 2 | 6 |
| Enj-Rel-Bur | 35,9% (web)[[3]](#footnote-3)54,5% (total) | 59% (CATI) |

*6.2 EU Survey on Income and Living Conditions (EU-SILC)*

Due to a missing EU-SILC model questionnaire, DESTATIS evaluated the EU-SILC Guidelines and the country-specific questionnaire, complemented by CBS, assessing the EU-SILC Dutch implementation. Consequently, the assessment was limited to target variables rather than to questions. At DESTATIS, EU-SILC is a paper questionnaire. At CBS, EU-SILC is a mix of web and CATI.

To manage the issue of a missing model questionnaire with specific questions, we checked the methodological guidelines with its description of EU-SILC target variables provided by Eurostat. We identified 198 target variables that are recommended to survey via primary data collection. Then we checked our specific German EU-SILC survey and identified 148 of those 198 target variables. We took the 148 target variables transformed into 112 German EU-SILC questions as a base for the assessment of the fitness criteria below.

*Table 7: Score on fitness criteria for EU-SILC*

|  |  |  |
| --- | --- | --- |
| *Dimension* | *Criterion* | *Scores* |
| *DESTATIS* | *CBS* |
| Screen size | Introductions | 0 | 34 |
| Instructions | 43 | 42 |
| Grids | 17 | 7 |
| #item per grid | 7 | 5.7 |
| Question text | 5 | 20 |
| # answer cat’s | 33 | 37 |
| Answer text | 37 | 2 |
| Touch navigation | Open question | 5 | 16 |
| Many answers | 5 | 0 |
| Duration | # of items | 329 | 262 (140 proxy) |
| Av duration | 55 min | 32 min |
| Household | Mix[[4]](#footnote-4) | No[[5]](#footnote-5) |
| Database | No | No |
| Cogntv burden | 1 | 9 |
| Consultation | 87 | 16 |
| Enj-Rel-Bur | NA | 33% |

To get a better understanding of the quantified indicators below the condition of EU-SILC in Germany is as follows: EU-SILC in Germany is a self-administered Paper and Pencil (PAP) survey. As it is containing a household-concept, it consists of a household questionnaire as well as an individual questionnaire for respondents over 16 years of age. In total the survey encompasses 157 survey questions. During the assessment we identified 35 questions that are not directly related to any target variable in the guidelines. Therefore we excluded those questions and focused on the remaining 112 survey questions.

Conclusion EU-SILC Germany:

* In total the assessed German EU-SILC survey covers 112 survey questions with 329 items in total.
* Considering the integrated routing, an average respondent (employed, quite healthy adult with children in a rental apartment) has to deal with about 200 items, leading to an average interview-duration of more than 50 minutes.
* The survey is a mix of a personal and a household questionnaire and the latter contains questions on both household level as well as on individual level. That concept needs to be restructured for smartphone use, as the changes between household and individual questions will become very confusing.
* The household questionnaire covers 17 multi-person grids, where the respondent has to reiterate those questions depending on the number of persons in the household.
* Moreover - besides the 17 multi-person grids - a respondent has to deal with 17 regular grid-questions. With an average of 7 items per grid, the amount and design of those grids/ variables have to be revised.
* A central characteristic of the EU-SILC survey is the detailed collection of numeric variables in the context of ‘income’, ‘taxes’, ‘social benefits’, ‘credits’ and ‘housing costs’ leading to 87 items that require the consultation of personal information by a respondent.
* 33 questions contain more than 5 answer categories. In context of small screen sizes, those questions need a careful re-design in order to ensure all necessary information are displayed at once. Furthermore the amount of those questions needs to be reduced.
* More than one third of the questions (43) provide instructions. On the technical perspective the implementation of instructions on smartphones with small screens sizes might not be a problem, but in combination with the large amount of questions, this might be burdensome for smartphone users, as each instruction is to be clicked on, as space is limited for direct presentation.
* As a consequence, the German EU-SILC PAP on smartphones is problematic on all three dimensions.

Conclusion based on EU-SILC guidelines[[6]](#footnote-6):

* There are 198 defined target variables to be provided by each country. If this data is delivered by primary data collection, the respondent will be burdened. Especially when responding via mobile devices the amount of target variables is extensive.
* Due to the possibility of using register-data, the need for primary data collection can be reduced. In case of Italy and the Netherlands it seems that the use of register-data reduces the average interview-duration to less than 20 minutes, which would be feasible for a self-administrated survey on mobile devices, while the average interview-duration in case of Germany and Poland (more than 45 minutes) is burdensome.
* About 30 target variables contain detailed items that require consultation of personal documentation. If those need to be answered by the respondent – as in case of e.g. Germany and Poland – this cannot be executed “mobile” while travelling and need later completion. So, easy open and close functions continue later is recommended and make the questionnaire functionalities demanding but feasible. If – as in case of e.g. Austria and Italy – the use of register-data provides all those complex variables without bothering the respondent, there won’t be any problem for smartphone respondents at all.

Consequently, the assessment of the EU-SILC target variables is ambivalent and cannot be observed without looking at the design of the country-specific EU-SILC surveys. Basically, the target variables surveyed only by mobile devices are not recommended due to amount and structure. Thus, register-data is extremely helpful, but not feasible in several countries. These aspects do advocate for a fundamental redesign of the entire survey, rather than adapting on and on

Conclusions Netherlands:

* Many survey items demand for screen size in terms of question length, introductions and instructions.
* Remarkable is the large number of open questions related to survey items with many answer categories. No use is made of long classification lists or databases, but this feature will be burdensome on a smartphone.
* A large number of items require consultation of personal files or are cognitively demanding. Here and there, we anticipate that a respondent will have to check with household members. The survey is not fit for completion anywhere and anytime.
* The online response rate has an average value, indicating that the survey is neither considered very interesting or very burdensome.
* A disclaimer is needed that at Statistics Netherlands quite a few survey items are completed using administrative data from tax income data.

*6.3 ICT Survey*

The ICT is assessed by CBS and DESTATIS. CBS evaluated both the model questionnaire and the Dutch ICT. DESTATIS looked at the German ICT. At CBS, the ICT is a mix of web and CATI. At DESTATIS, the ICT is a paper questionnaire with a small part of the questions referring to the household as a whole.

Conclusions CBS for model questionnaire:

* The ICT is a relatively short survey.
* Question texts are long for many questions, due to the complexity of the topic. There also some questions with long answer texts.

Conclusions DESTATIS:

* ICT is a household survey, but using a household and an individual questionnaire.
* Generally, we suffer from long, detailed answering categories, besides of long lists of answering categories.
* Instructions are somewhat long, but could be managed with appropriate layers in smartphone design
* At least three times, data base connection is needed: economic sector, profession and education. All are demanding on smartphones.
* The questionnaire starts with more demographic basics than expected: 15 question.
* The list of country of birth is critical: > 50 countries; would be nice by a pull-down menu, but less advisable on Smartphones.
* Answering categories integrate explanations, as the topic is rather difficult; therefore, redesign on wording and how to provide additional help is needed. This will be a big challenge.
* The observation reference period is problematic (at least for paper questionnaire): There are two observation reference periods, 3 and 12 month. This will end up critical on the smartphone, to be overlooked, more than on the PC/Laptop.

*Table 8: Score on fitness criteria for ICT*

|  |  |  |
| --- | --- | --- |
| *Dimension* | *Criterion* | *Scores* |
| *Model* | *CBS* | *DESTATIS* |
| Screen size | Introductions | 5 | 26 | 17 |
| Instructions | 4 | 4 | 16 |
| Grids | 1 | 23 | 0 |
| #Items per grid | 6 | 2 | NA |
| Question text | 13 | 39 | 26 |
| # answer cat’s | 7 | 13 | 10 |
| Answer text | 3 | 16 | 5 |
| Touch navigation | Open question | 0 | 5 | 9 |
| Many answers | 0 | 0 | 0 |
| Duration | # of items | 39 | 140 | 61[[7]](#footnote-7) |
| Av duration | NA | 23 min | NA |
| Household | No | No | No |
| Database | No | No | 2 |
| Cogntv burden | 7 | 7 | 1 |
| Consultation | 3 | 4 | 2 |
| Enj-Rel-Bur | NA | 33.1% (web)36.9% (CATI) | NA |

Conclusions CBS:

* The CBS ICT is much longer than the model questionnaire, roughly 100 questions are added, and will make it harder to redesign to smartphones.
* The ICT has many items that are problematic in screen size; question texts are often long, answer texts are sometimes long and quite a few items have introductions and/or instructions.
* ICT has a web response rate of around 33%, which is a moderate to good response rates compared to other surveys.

*6.4 Labour Force Survey (LFS)*

The LFS first wave is assessed by CBS and SSB. SSB considered both the model questionnaire/guidelines and the Norwegian LFS. At SSB, the LFS is CATI, but will be redesigned to include web. At CBS, the LFS is a mix of web, CATI and CAPI. There are very many paths through the questionnaire, dependent on employment status, job seeking activities and participation in education. This means that the number and complexity of questions will vary considerably per respondent. For this reason, it was decided to consider the Norwegian LFS for three distinct groups of respondents:

* A regular employee with a fixed working contract;
* An unemployed person seeking work;
* A student with a part-time job without a working contract.

*Table 9: Score on fitness criteria for LFS*

|  |  |  |
| --- | --- | --- |
| *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 | 4 |
| #Items per grid | NA | NA | NA | NA | NA |
| 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)54% (overall) |

Conclusions SSB for the model questionnaire:

* The main problems with the LFS “model questionnaire”/un-operationalized variable list is the frequency of questions with five or more answer categories, and answer categories with 10 or more words.
* These features are of course at least partly connected to the fact that it is not an actual model questionnaire. It requires major adaptation into an actual questionnaire.

Conclusions SSB for Norwegian LFS:

* The most prominent characteristic is the number of questions with instructions, either displayed on the screen or available by clicking on icons.
* In the pilot LFS, instructions that are short and/or considered relevant for many or most respondents are always visible. Instructions that are relevant for smaller groups of respondents are available through clicking on icons.
* The part-time working student persona received the highest number of questions, the highest number of questions with additional instructions, and would most likely have had the most use of these instructions.
* The unemployed persona received the lowest number of questions and the lowest number of questions with additional instruction, but more questions with six or more response options.
* Using the mobile web fitness criteria, the first wave LFS pilot questionnaire is best suited for respondents that are unemployed, followed by employees with fixed working contracts. For students working part-time and other groups in a liminal position in the labour market, desktop web or interviewer-administered modes may be better suited due to many questions with additional clarification and instructions.
* As a side remark: The second to eight wave questionnaires are shorter than the firstt wave, and use previous wave data as well as register information, for dependent interviewing. For employed respondents who have not changed jobs, these follow-up interviews will be shorter and even better suited for mobile web than the questionnaire assessed.

Conclusions CBS:

* Out of the total number of 346 questions more than half is problematic on at least one criterion. 51 questions have two issues and 26 have three or more issues.
* The LFS is a household survey with average duration of 27 minutes in web. On the person level the average duration is, obviously, shorter and may reduce to around half, i.e. between 10 and 15 minutes.
* The LFS has relatively low web response rate of little over 20% (without incentives).
1. **Translating scores to recommendations**

The next step is to translate the scores on the 16 criteria to an overall score or much smaller set of scores that represent fitness. Ideally, this leads to a simplified recommendation:

Green: Survey requires no specific adaptation (other than general mobile layout);

Yellow: Survey requires some adaptation;

Red: Survey requires (near) total redesign.

We have made such an overall assessment in two steps: First, we scored the three dimensions by averaging the number of criteria that scored good or bad. Second, we combined the three dimensions.

Table 10 shows an assessment of the three dimensions for the four surveys:

* EHIS: The EHIS screen size dimension is somewhat problematic because of the long introduction and many instructions. The navigation shows no problems. The duration is problematic.
* EU-SILC: The EU-SILC questionnaire is problematic on all dimensions. It has many questions that require sufficient screen size. There are open questions and questions with many answer categories. It is also a long survey that requires consultation of personal archives. The duration dimension may be less problematic when administrative data are used, as is done in some ESS countries, but even then will remain lengthy.
* ICT: The ICT survey scores good on both the navigation and duration dimensions for the model questionnaire. Country-specific implementations may be problematic on duration. The screen size dimension is problematic due to the large number of instructions, introductions and long questions/answers.
* LFS: The LFS is 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.

*Table 10: Scores on the three dimensions screen size, navigation and duration for each survey.*

|  |  |  |  |
| --- | --- | --- | --- |
| *Survey* | *Screen size* | *Touch navigation* | *Duration* |
| EHIS |  |  |  |
| EU-SILC |  |  |  |
| ICT |  |  |  |
| LFS household |  |  |  |
| LFS person |  |  |  |

What general conclusions can we draw? From table 10, we deduce that EU-SILC and LFS on the household level require a (near) total redesign for smartphones, i.e. score red. The picture for EHIS, ICT and LFS on the person level is more mixed and subtle. Here, we choose a yellow score, i.e. the surveys require some adaptation. Since the EHIS’ duration is problematic, we do not recommend to encourage smartphones without a larger redesign in which split questionnaire or panel questionnaire designs are considered. For the ICT and LFS on the person level, smartphones may be encouraged, once adaptations have been implemented.

1. **Discussion**

Out of the four surveys of this deliverable, two have scores that indicate a redesign for smartphones may be feasible, but with modifications. The ICT survey is relatively short and has a moderate to good online response rate. It will require a considerable redesign of the wording of questions and answers and solutions to handle the introductions and instructions. The LFS on a person level may also be doable. Online response rates are lower, but are likely to increase for a person level LFS. Navigation needs to be looked at as it has many filter questions and a number of open questions. Like the ICT, wording of questions and answer categories require a thorough redesign. The EHIS and, especially the EU-SILC, do not seem suitable for smartphones without major changes to the length. EU-SILC is problematic on all fronts and seems the hardest to transform to smartphone application.

The handling of filter questions, and specifically the advantages and disadvantages of follow-up questions on the same screen, needs attention and further consideration.

We have a few topics for discussion:

* To what extent should we focus on smartphones in surveys?
* What implications does a focus on smartphones have for data collection strategies?
* Under what conditions should we discourage, accept or stimulate smartphone participation?
1. Norway included the scoring of filter questions in their summary; other countries did not, see page 3. [↑](#footnote-ref-1)
2. The duration of the model questionnaire is unknown. The CBS EHIS takes 34,5 minutes on average, but includes extra items, as it is integrated in the Dutch Health Survey [↑](#footnote-ref-2)
3. The reported response rates are for the CBS specific implementation of the EHIS that is integrated with the Dutch Health Survey. [↑](#footnote-ref-3)
4. 142 of the 329 items refer to the household, although they are answered by one person (proxy reporting is allowed) [↑](#footnote-ref-4)
5. For part of the survey items proxy reporting is not allowed and a member of the household core needs to answer the questions. [↑](#footnote-ref-5)
6. Based on EU-SILC target variables, German PAP-assessment and key indicators of EU-SILC surveys [↑](#footnote-ref-6)
7. DESTATIS also has 8 household items. These are currently not included in the assessment. [↑](#footnote-ref-7)