**ESS surveys and mobile device data collection**

Barry Schouten (CBS), bstn@cbs.nl

Karen Blanke (DESTATIS), karen.blanke@destatis.de

Dag Gravem (SSB), dag.f.gravem@ssb.no

Vivian Meertens (CBS), vmts@cbs.nl

**Abstract**

*The coverage and use of mobile devices have increased very strongly over the last five years and are anticipated to grow for the coming years. This change in communication has already led to an increasing percentage of respondents that start online surveys on mobile devices, despite the fact that the surveys are often not fit for small screens and mobile device navigation. ESS national statistical institutes have begun to react to this trend and started to explore smartphone options. In 2017, ESSnet MIMOD has been launched with a separate work package on mobile devices. We describe the goals and some first results of this work package.*

**Keywords:** Multi-mode surveys, Multi-device surveys, ESS surveys;

**1. Introduction**

The MIMOD – Mixed Mode Designs for Social Surveys was awarded a Eurostat Grant, in December 2017. The project is led by Istat (Italy) in partnership with CBS (Netherlands), SSB (Norway), STAT (Austria) and Destatis (Germany). MIMOD aims at supporting NSIs in facing a range of challenges when implementing multi-mode and multi-devices data collection. It has five substantive work packages. One of them (WP5) deals with the use of mobile device data collection for ESS surveys, which is the topic of this paper. There are two purposes of mobile devices: 1) conducting surveys on mobile devices as the data collection mode with new features and limitations and 2) employing mobile device sensors (such as GPS, camera, microphone, accelerometers) to enrich surveys. The first purpose is important as more and more respondents attempt online surveys on a variety of devices. The second purpose is more advanced and requires programming and a separate backend infrastructure.

The objective of MIMOD WP5 is threefold:

1. Determine which of the ESS surveys is fit for mobile devices with relatively little modifications to content and length;
2. If an ESS survey is not fit, decide what (potentially infeasible) changes to the survey content and questionnaires would be needed;
3. Explore the potential use of mobile device sensors in ESS surveys;

To give some background to the implementation of smartphones, table 1 provides an overview of countries and smartphone options for seven ESS surveys. From the table, it becomes immediately clear that many countries do not yet employ the web mode, and, if they have an online option, then few have made any adaptation to the surveys for smartphones. Countries that do not have an online option yet, thus, do have the advantage that they can redesign a survey and account for mobile device use. Most countries that included the web mode did so before the strong rise of smartphone coverage started.

*Table 1: Number of ESS countries that implemented different online and smartphone options*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Survey* | *No web* | *Smartphone 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 discuss the three objectives in this paper. Work for objective 1 started early 2018 and we present fitness criteria. In section 2, we discuss the fitness of surveys for smartphones. In section 3, we give a general discussion on mobile device sensors. We end with a few points for discussion in section 4.

**2. Fitness of ESS surveys for smartphone questionnaires**

*2.1. Disparity of modes and devices*

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 focus the assessment of fitness of surveys for smartphones relative to traditional online devices, desktops and laptops, i.e. we state that ESS surveys can be online and self-completed. 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 where 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, e.g. a mouse or touch pad.

*2.2. From disparity to fitness 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, 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.

* Item introductions: Survey items may have opening 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.
* Grid questions: Grid questions are a series of survey items with the same answering categories on a relatively similar topic. They are presented as a whole and answering categories are shown only once. Grid questions form a block of items that demand more screen size or a different type of navigation. On smartphones grid questions that require scrolling or navigation are changed through accordion, carrousel or other modifications of the navigation.
* Length of item: Survey items with longer question text (excluding introduction text) and/or longer answer category text demand more space. On smartphones text are shortened or scrolling is needed.
* Number of answer categories: The number of answer categories determines 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.
* Filter questions: When survey items are closely related, e.g. one survey item is a follow-up question to another, then they are often displayed on the same screen. A filter question then implies that one or more new items appear on the same screen and requires larger screen size. On smartphones, scrolling is needed or questions are separated to different screens.

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.

* 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.
* Items with many answer categories: A 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.

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.

* Number of survey items: The number of survey items directly affects the interview duration. Since most surveys contain filter questions, the number of survey items may vary per respondent. To date, long surveys are often blocked or discouraged for smartphones. There is a revived interest in split questionnaire design and panel questionnaire design.
* Person or household survey: In a household survey, multiple persons need to answer questions, making the survey longer and/or requiring the need to switch between persons or allow for proxy reporting. To date, household surveys are rarely if not encouraged for smartphones.
* 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.
* Task complexity: Survey items that require extra effort from the respondent, e.g. calculations or searching for information in personal administration, take more time and may not be compatible with the time and location in which the respondent performs the survey.

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.

*2.3. Operationalization of fitness criteria*

The fitness criteria should be as clear as possible, allowing for as little room as possible for subjective assessment. We have operationalized the various criteria as given in table 2.

*Table 2: Fitness criteria*

|  |  |  |
| --- | --- | --- |
| *Dimension* | *Criterion* | *Operationalization* |
| Screen size | Introductions | Number of items with introductions |
| Grid questions | Number of grid questionsAverage number of items per grid |
| Item length | Number of items with > 20 words (excluding introduction text) |
| # answer cat’s | Number of items with > 5 answer categories |
| Filter questions | Number of (anticipated) filter questions with follow-up questions on the same screen |
| Navigation | Open questions | Number of open questions |
| Many answers | Number of items with > 25 answer categories |
| Duration | # of items | Total number of itemsAverage number of items asked 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 a average respondentNumber of (anticipated) items that require consultation of personal documentation by a average respondent |
| Enj-Rel-Bur | Response rate to traditional online devices |

Each ESS survey is scored on these 15 criteria. The assessment is executed on the ESS-model questionnaires. Where a model questionnaire is not available, guidelines in combination with country specific questionnaires are applied. The second step will be to translate this to an overall score or to a much smaller set of scores that represent fitness. Ideally, this then leads to a simplified recommendation such as

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

Yellow: Survey requires some adaptation

Red: Survey requires (near) total redesign

However, there is still discussion on the rating or weighting of each criteria, as some may have more, some less relevance. At the time of writing, the coding of the ESS surveys was on-going. Results will be reported in MIMOD deliverables and conference papers.

**3. Introduction**

The second main activity in WP5 of MIMOD is the exploration of mobile device sensor for ESS surveys. Mobile devices have become standard tools for communication, but they also support a range of automated measurements through their sensors. The range and type of sensors depend on the platform (Android and iOS) and model, but there is a fair list of sensors that is shared by all. The appendix contains an overview of available sensors. Many mobile device applications use at least one of these sensors, such as GPS or accelerometers. Employing multiple sensors and combining with advanced software, mobile devices can be powerful tools to collect data. In brainstorm sessions at Statistics Netherlands, six topics have been identified as promising for a combination of survey and sensor data: 1) Travel and time use, 2) Internet behaviour and use, 3) Expenses and buying behaviour, 4) Health and fitness, 5) Living conditions, and 6) Working conditions. These topics are hard or cognitively burdensome to fully measure by means of a questionnaire and may arouse relatively strong measurement error. Furthermore, these topics already attract a lot of research into the potential of sensor data, partially as in-depth, qualitative measurements and partially as big data. Some of these topics map directly to ESS surveys such as EHIS, ICT, HBS and HETUS. For other ESS surveys, such as LFS and EU-SILC, it is less apparent how sensors may be employed.

Mobile device sensor data require a different data collection infrastructure and different data processing, i.e. an investment in both systems and researchers/analysts. The strongest business case for sensor data, therefore, comes from questionnaire topics that are demanding in terms of respondent effort, time and/or cognition but that seem within reach for sensors. Out of the six themes, the most promising are sensor data around mobility/travel using GPS, Wi-Fi and accelerometer, for use in travel surveys and time use surveys. Travel data have already been explored extensively and commercial applications are available. Furthermore, a number of NSI’s started to explore the development of apps themselves.

Two other themes where sensor data are actively investigated are budget expenditure and ICT/internet behaviour. Within Eurostat there is an interest in innovating budget expenditure surveys, given the often low response rates, high response burden and data quality concerns of these surveys. The business case for budget expenditure, therefore, is very strong. Sensor data measurements can be collected using tailor-made applications or using plug-ins in browser-oriented questionnaires. The latter offer the advantage that respondents can visit the regular website and do not have to download and install an application. For ICT/internet behaviour, the browser-oriented approach may be promising.

A theme that is considered highly relevant and promising but complex and multi-dimensional is health and life style. In this theme, a lot of (inter)national parties are already active. It is promising, as respondents may see the benefits of collecting such data and learn about themselves.

Obviously, respondents need to consent to provide (and link) sensor data. Some sensors, such as the accelerometer, do not need a consent, technically, and some sensors, such as GPS or Wi-Fi, may already be enabled by the respondent him/herself. Other sensors, such as audio or video, always need explicit consent. However, regardless of the type of sensor, it is deemed ethical to inform respondents and to explicitly ask for consent.

Apart from ethical and methodological issues, there is also the confidentiality of sensor data. In theory, when secure data transmission and storage can be assured, it may seem that sensor data are just like traditional survey data. However, the main differences lie in the automation of the data collection, i.e. respondents do not fully know what data is collected and how it is used, and in the availability of the sensors to other applications and data collectors when consent is given. This requires developments and research of new communication strategies of NSI’s using sensor data in their data collection infrastructure.

In the second half of 2018, sensor measurements are explored for the ESS surveys and recommendations will be given about their utility and consent strategy.

**4. Discussion**

Given that MIMOD, and in particular WP5 about smartphone data collection, is still in its early stages, there are various open questions. We mention a few here.

Questions for discussion:

1. Should we consider a mobile device to be a separate mode in terms of selection and measurement properties when applied to ESS surveys?
2. Does accepting or stimulating mobile device response imply shortening of survey questionnaires or the use of advanced techniques such as split questionnaire/modular survey designs?
3. Shall/must we consider the smartphone in the close future as the “first mode”?
4. Can we still rely on the same level of data quality and comparability when collecting data on smartphones within the ESS?
5. Is coverage of mobile devices sufficient across ESS countries to stimulate sensor measurements?
6. How to find the balance between high quality, but expensive, and relatively inexpensive sensor data that are available in larger quantity?
7. Do sensor data measurements have ethical and privacy consequences beyond those of online data collection, and how does this vary across ESS countries?
8. What kind of additional or other communication strategies are needed compared to traditional surveys to get reasonable consent in sensor data collection?

**Appendix - Overview of sensors in mobile devices**

3D touch: This sensor measures the pressure exerted on the screen. Small objects up to 385 grams can be weighted. Only in iPhone 6S and 6S plus.

Accelerometer, gravity, gyroscope: A set of sensors measuring motion, acceleration and position of the device. Used for position tracking or step counters.

Ambient Light: Measures the intensity of the ambient light. More advanced versions also determine the light colour or –temperature. Commonly used to adapt the screen brightness and colour to the ambient conditions.

Bluetooth: Wireless communication protocol. Can connect to small low energy devices, such as wireless headphones, key fobs, smartwatches or smart scales. Also detects the presence of Bluetooth beacons.

Camera: Takes pictures or videos, or measures light intensity. Usable for image- or pattern recognition, scanning of QR- or barcodes and colour analysis. Can also coarsely measure gamma rays, a form of radioactivity.

Camera Flash: Usually used as a flashlight or as illumination for pictures or videos. Can create stimuli that can be picked up by other sensors.

Cellular: The core of all cell phones. Used to make and receive calls and text messages. Strength and ID of cell tower broadcasts can be measured. With the knowledge of tower positions the user location can be determined with a precision of ~500 meters. More advanced cell phones – almost all cell phones today – can also connect to the internet. The presence of the internet connection as well as its speed (upload, download, responsiveness/ping) can be determined. Not often found in wearables yet.

Fingerprint: Some devices are capable of detecting fingerprints. The raw data is not accessible, but it can be used as identification or simply as a button.

GPS: The Global Positioning System. Dozens of GPS satellites circle the earth and broadcast beacon signals. By measuring the time-of-flight of the satellite signals, the distance to that satellite can be calculated. With four or more satellites visible, the position on earth can be triangulated. The precision is ~5 meters outdoors. Indoor performance is poor. The satellite signals also includes time information.

Heart Rate: Measures the heart rate, usually optically on the finger (cell phone), wrist (smartwatch) or in-ear (headphones).

Humidity: Measures ambient humidity. Not very widely used yet

Magnetic Field: Usually used as a compass, but can also measure the strength of magnetic fields or can be used, within limits, as a metal detector.

Microphone: Detects speech and sounds that can be saved, streamed or analysed. Can also determine loudness and detect ambient noise. Multiple microphones in one device allow for determining the directionality or distance of sound sources. This is used to filter out ambient noise in phone calls. Microphones can also be used to record the heartbeat or estimate lung function/spirometry.

NFC (near field communication): The same technology as contactless payments. Can be used to pay with the cell phone or smart watch, as “contactless QR code” / “NFC tags” to change phone settings (muting) or trigger the start of certain apps. A phone can be a tag as well, so two devices can identify each other and initiate a data channel for communication.

Pressure: This sensor measures the ambient air pressure and functions as a barometer. The precision is so high, that height differences of a few meters (ground floor vs first floor) can be detected. Is also used in combination with GPS for more precise height determination.

Proximity: Measures the presence of objects close to the screen, usually binary (object present or not). Switches off screen and touchscreen during phone calls.

Screen: Display of static or dynamic images. Can also illuminate the surroundings. Commonly used for user interaction.

Speaker: Plays sounds or speech. Can be used for feedback to the user or in combinations with other sensors such as microphones.

Thermometer: Usually placed in or near the battery to prevent overheating. Measures the battery/cell phone temperature which might be higher than the ambient temperature.

Vibration: Feedback mechanism, induces vibration in the device. Can be used as feedback or combined with other sensors such as the accelerometer.

Wi-Fi: Usually used for internet access. Can detect presence and strength of different wireless networks in different frequency bands. Measuring of connection speed (upload, download, responsiveness/ping) is possible.

Wireless Charging: Neither sensor nor feedback, strictly speaking. Some devices can be charged wirelessly over very short distances (1-2 millimeters). There are competing non-compatible standards, such as Qi and PMA for cell phones. The device can detect the presence of a certain charging station.