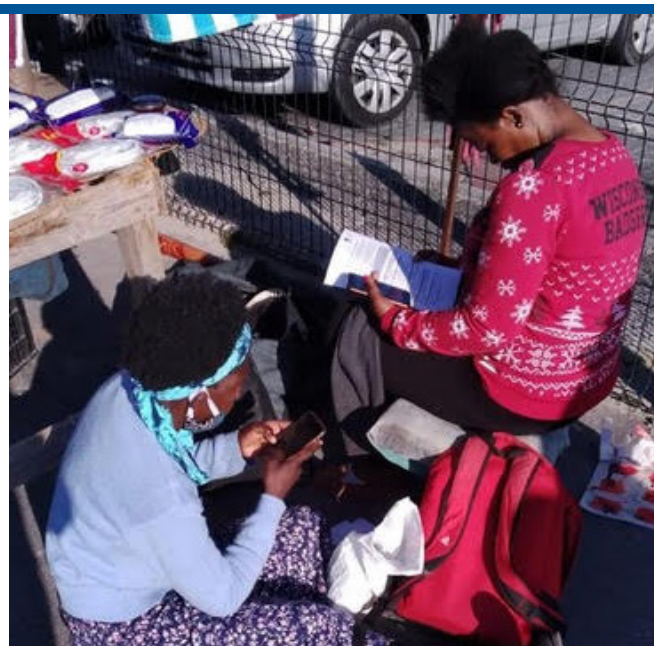


Implementing a mobile survey using KoboToolbox

Johanna Hansmann

Abstract

Remote and digital research formats were increasingly used during the COVID-19 pandemic due to national lockdowns, worldwide travel restrictions, and protective measures. The use of mobile phones for remote data collection provides an inexpensive tool to quickly collect real-time and large-scale data sets. This paper focusses on implementing mobile surveys using the open-access tool of KoboToolbox. After briefly looking at the advantages and challenges of mobile surveys, an experience-based guideline is presented based on practical insights gathered in Benin, Cambodia, Kenya, Nigeria, and South Africa. Lastly, good practices and lessons learned are shared.



Introduction

The worldwide increased internet connectivity and mobile device ownership, together with the large availability of software and web hosts for data collection unlocks alternative remote and digital research formats (Sue & Ritter, 2012; Henze et al., 2020). Already common in challenging environments, the use of mobile phones provides an inexpensive tool for quickly collecting large-scale data sets (Andrew et al., 2003; Henze et al., 2020). However, switching to remote data collection during COVID-19 outbreak went beyond the factors of cost and scalability. For many, it was the only response feasible for complying with national lockdowns, travel restrictions, and protective measures.

This paper focusses on conducting mobile surveys using the open-access tool of KoboToolbox. The purpose is to share practical insights on the experiences of four different SLE research projects collecting data remotely in Benin, Cambodia, the Cape Flats/

SLE method briefs are created from the practical experiences of our alumni in their interdisciplinary research projects. Lessons learned and good practices are compiled. In each brief, we present the method that is explained clearly, step by step, and with the help of practical examples. With its method briefs, SLE aims to support researchers and practitioners who are active in solution-oriented and transformative international development work by providing insights into hands-on methods in a structured manner, so that the wheel does not always have to be reinvented.

The Centre for Rural Development (SLE) is affiliated with the Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences in the Faculty of Life Sciences at the Humboldt-Universität zu Berlin. Its work concentrates on four branches: international cooperation for sustainable development as a post-master degree course, training courses for international leaders and experts in the field of international cooperation, research on sustainability issues, and advisory services for universities and organisations.

The views and opinions expressed in this brief are those of the authors and do not necessarily reflect the official position of SLE.

South Africa, and selected countries of the African Union (i.e. Kenya, Nigeria, and South Africa). After briefly describing the method and looking into the general advantages and challenges, this brief guides through the different steps of implementing a mobile survey. The last section will highlight good practices and lessons learned.

Description of the method

Collecting data through a survey belongs to one of the basic research methods of social science. Surveys form one end of empirical social research, and qualitative interviews form the other, which use individual qualitative interviews or focus group discussions to obtain more detailed or in-depth information to studying perceptions, processes, or attitudes and behaviours. In this brief, however, only the survey method is described. It is a systematic and standardized approach to collect information on individuals, households, or entities by mainly asking questions about preferences and lived experiences. Using different sampling and statistical techniques allows, for example, one to draw conclusions representative of a group or explore causations and complex relationships (Rossi et al., 2013).

Digitalisation and technological advances have made it possible to simplify data collection and data management activities of traditional paper-based surveys (Sherin et al., 2018). A well-known tool for this is KoboToolbox, which is extensively used among humanitarian agencies and international organisation, such as UNHCR, OCHA, WFP, UNDP, and the World Bank (KoBoToolbox, 2021).

This intuitive tool is an open source service platform, with creative space for researchers to customize their own survey questionnaire from scratch or using existing survey templates. It allows to collect data through more than 20 different types of questions, such as the traditional paper-based binary, multiple-choice, likert-scale, or open-ended questions. Other formats include the collection of data on the location through GPS data or on the setting/context through images, videos or audios. Through these features, remote and real-time field data can be collected by sharing survey's URL link or engaging local enumerators to collect data with their mobile device. This allows fast dissemination of information, which considerably reduces costs linked to time, personnel, and transportation.

Given the COVID-19 social distancing measures,

completing a survey individually via mobile device was also significantly convenient and safer for both survey participants and researchers. KoboToolbox allows for the survey to be filled out online as well as offline. When offline, the entered data is firstly stored on the device and uploaded when the network connection has been established again. Automatically generated reports on incoming data enables researchers to closely monitor and steer the data collection process. Also, data can be downloaded in a number of formats such as Excel or SPSS. Thus, the speed and practicality at which data is available allows for an overall timelier analysis, assessment and evaluation of information (Greenleaf et al., 2017; Henze et al., 2020; JDC, 2021; Kobotoolbox, 2021).

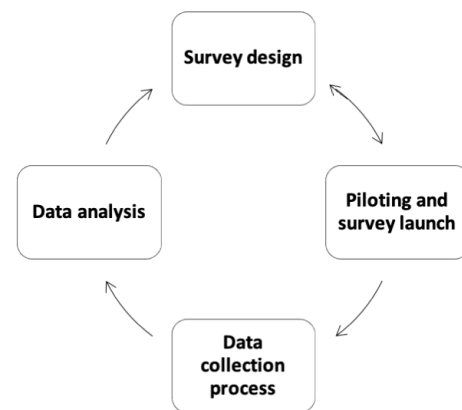


Figure 1: Phases of conducting a mobile survey (Source: Own illustration)

However, conducting mobile surveys come with challenges not found in conventional formats. The development of technology has outpaced corresponding advances in ethics in online and remote research, i.e. regulation on anonymity and confidentiality of data. Moreover, mobile surveys require participants and enumerators to have access to the internet, own or have access to a mobile device, and be literate. Therefore, this method works well for some projects but by no means is appropriate for all research objectives. Online formats tend to have a significantly lower response rate than face-to-face interview and observatory methods. This is due to the lack of interaction and personal connection, which are particularly important in establishing trust, especially when researching sensitive topics. Other challenges are related to the recruiting of study participants via the internet, which can prevent random sampling. This, together with general challenges on sampling can potentially affect data quality, accuracy and reliability (Fielding et al., 2008; Andrews et al., 2010; Sue & Ritter, 2012; Greenleaf et al., 2017; Hensen et al., 2021).

Approach

The topic, nature and objectives of conducting a mobile survey was very different for each of the four SLE overseas projects (See Annex). The experiences, important aspects to keep in mind, best practices and lessons learned were reviewed upon the four phases illustrated in Figure 1. This paper will refer to the different projects based on the country where the survey was implemented, i.e., team Benin, team Cambodia and team South Africa. As for the mobile survey conducted in Kenya, Nigeria and also South Africa, the reference of team 'African Union' will be used, since these surveys were conducted as part of a single overseas project in cooperation with the African Union. The next section will describe each phase in Figure 1 in more detail.

Implementing a mobile survey step by step

A timeline, organizing and detailing by week, and which steps to conduct is a good way to keep an overview of the survey objectives and processes. Figure 2 shows the average time all projects needed to accomplish each phase. The data collection phase depends on various factors such as the sample size or the numbers of enumerators engaged.



Figure 2: Implementation timeline (Source: Own illustration)

I. Design

The first steps in the design process of a survey are normally defining the study's aim and objectives, a literature review and consultations with specialists on the topic (Sue & Ritter, 2012). Interviewing project partners or members of the target audience can also be helpful. In the case of team South Africa, a hired academic advisor and co-researchers living on site were continuously consulted along the entire survey design process.

The questionnaire

A questionnaire is usually divided into modules, which contains questions on a similar topic or theme. For example, team Cambodia had a module on food security, a module on livelihood factors and a module on local knowledge and social inclusion. Creating an ad hoc new survey can take much time,

therefore, including already tested modules or questions is preferable (FAO, 2018). In the case of the research in South Africa, for instance, the Food Insecurity Experience Scale Survey Module (FIES-SM) developed by the FAO, which includes eight world-wide-tested questions to measure food security, was used.

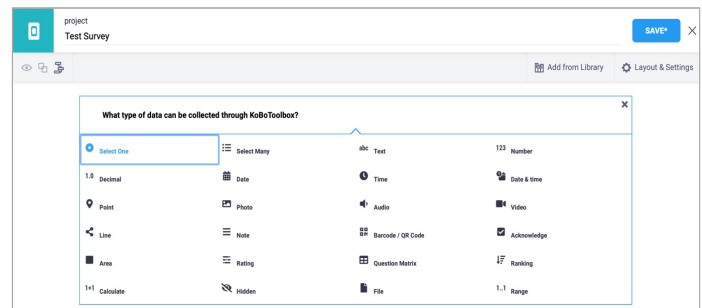


Figure 3: Snapshot of the KoBoToolbox form builder and question types (Source: KoBoToolbox, 2021)

Placement of modules

Once the survey modules are defined, it is essential to carefully consider appropriate placement of each module to ensure good quality data. The first module normally contains general demographic questions to build a respondent profile. It is advised that the module containing the most socially and emotionally sensitive questions in relation to the other modules should be placed after the demographic section (FAO, 2018). In the case of team Cambodia and team South Africa, the FIES-SM was placed as second module, as it contains questions on distressing experiences.

Formulating questions

Questions must be linguistically and culturally adapted to the context, especially when working with translations. The data quality is highly dependent on the wording, terminology and proper understanding of the questions by the respondents (FAO, 2018). Therefore, it is helpful to include co-researchers, tandem or project partners during the formulation process. For example, team Benin closely consulted back and forth with their local/on-site tandem partners to formulate questions in French.

Selecting question types

The KoboToolbox 'form builder' has 24 different types of question formats (Figure 3). These range from traditional single-choice questions, to ranking answers, uploading a photo or sharing GPS coordinates.

Although there are many interesting formats, it is important to know your target group and research

location. For instance, if mobile data networks are weak or too expensive and smartphone ownership is low, then selecting question formats that require participants, for example, to send pictures or GPS coordinates is not optimal (Henze et al., 2020).

Both teams, Benin and Cambodia, identified that multiple-choice questions save considerable time during data evaluation compared to open-ended questions. Although the analysis of open-ended questions is time consuming, as they require systematic coding and are not always optimal for applying statistical methods, they offer other benefits to the research, such as the opportunity to discover detailed, in-depth or unexpected insights. Hence, open-ended questions reduce the risk of bias. Moreover, when asking questions that require a numerical answer, for example 'How many years of work experience do you have', team Cambodia suggests to allow participants to enter a number themselves instead of giving multiple answer options, i.e., '0-5 years', '5-10 years', '10+ years'. By using this binary question format instead of an open-ended question format, team Cambodia was unable to calculate a more specific Propensity Score Matching (PSM) due to too inexact answers. This is not just a problem with PSM, but with all kinds of statistical conclusions. The more imprecise the answers, the more imprecise their interpretability. Therefore, whether binary, multiple-choice, open questions, etc., it is important to know what kind of data analysis will be conducted on a later stage when selecting the type questions.

The 'form builder' in KoboToolbox also allows to set up a question under the following three formats: (1) mandatory; (2) 'skip logic' or through the application of a (3) validation criteria. If a question is mandatory it means that one of the answer options must be selected, while the format 'skip logic' and validation criteria mean that a question/module is enabled only depending on the response of the previous question. This can help reduce respondent burden. For example, through 'skip logic' a question asking about the type of job a participant has is enabled only if he/she selected the answered option 'yes, I have a job'. Using a validation criterion is similar to 'skip logic' and a helpful feature if specific selection conditions are set, i.e., filter participants older than 21 years old. The difference of setting a criterion to the 'skip logic' format, however, is that validation criteria is used on open-end questions -- hence, no answer option is given. Possible responses are coded before-hand and this enables a reduction in respondent bias. For instance, the survey implemented in South Africa

asked about something similar to a ZIP-code. Participants were asked to enter a numerical value in the first question. The full questionnaire was then only displayed to participants or enumerators entering the Zip-code relevant for the study area.

The African union had a questionnaire of 60 questions. However, only two questions were mandatory. This means that the other 58 questions could be simply skipped by participants without giving an answer. The team was unsure whether questions were skipped because participants didn't want to answer or because answer options were not suitable, or other reasons. When questions are skipped systematically, the conclusions drawn from analyzing the data are biased. In contrast, team South African had 55 out of 57 questions marked mandatory, giving the answer option 'I don't want to answer' or 'not applicable' within all question formats. Although there was a high consistency in replies throughout the survey, the problem of systematically unanswered questions remains.

Final steps

Before deploying the questionnaire and starting the piloting phase, it is important to include short welcome words and contact information, as well as information on the incentives for participating, if applicable.

It is also important to have one questionnaire (in KoboToolbox they are named projects) per research location. A questionnaire can easily be duplicated. For example, the African Union had a project separately for Kenya, Nigeria, and South Africa. In this way, a close monitoring of incoming data for every research area is possible.

II. Implementation

Piloting the survey

Pre-testing and piloting the questionnaire is very important to ensure a smooth survey implementation process. This allows for the questionnaire to be revised and adapted if necessary (Sue & Ritter, 2012). All teams piloted their survey at least twice with their research teams on site.

Why hire enumerators?

Although mobile surveys can easily be shared and disseminated remotely, it is helpful to have local enumerators who know the context and speak the language. For example, they have access to local networks and digital communication channels where they can easily share the URL's link, e.g., Facebook and WhatsApp groups. Interviewing via telephone or

face-to-face is also possible and preferable from a data quality perspective. Engaging local people can reduce the challenge of not being able to build trust. For example, local enumerators working in team South Africa were provided with tablets and shared the survey with their networks and also went to public spaces, such as the shopping mall, to interview people and fill out the survey together. Team Benin also had tandem partners who visited different local markets and recruited participants there. Team Cambodia outsourced the data collection, meanwhile, but the enterprise also closely worked with local enumerators to increase participation.

Enumerator training

Enumerators are a central quality control factor of the data collection process. Investing time in training will pay off in an increased quality of data. Therefore, enumerators should be provided with clearly-written instructions and information (i.e. through an enumerator manual) on the target group, sample size, on how to recruit participants and how to ask survey questions. Beyond this, an 'in-person' (digital) training is highly recommended (FAO, 2018). Both team Benin and team South Africa conducted a half day enumerator training via Zoom. This enabled enumerators and field coordinators to familiarise themselves with the survey questions and with the technical aspects of KoboToolbox. For example, the South African team role played data collection, where enumerators practiced conducting the interview via WhatsApp call, while at the same time typing the answers in their mobile device. Ground rules for asking questions were explained and different scenarios on how to recruit participants were acted out.

It is important that enumerators know their responsibilities and are aware of the data collection time frame and who to contact in case of technical issue or general challenges. For both teams, Benin and South Africa, it was helpful to set up a tandem partnership between a team member in Berlin and a local enumerator. By creating WhatsApp groups, frequent communication was possible.

III. Data collection

The period between the launch of the survey and the downloading of data should be normally spent supervising and monitoring the data and preparing for analysis (Sue & Ritter, 2012). The 'summary' function of KoboToolbox allows for a good data overview. Good survey practice includes checking the incoming questionnaires at the end of every day. Downloading data directly to Excel or SPSS is helpful. This

way, erratic response patterns or delayed objectives of sample size can be quickly noticed (FAO, 2018). It is equally important to transparently communicate with the team on site (i.e. tandems, co-researchers, enumerators, partners) on a daily basis on the data collection process.

When working with large data sets many team members should have access to the survey project in KoboToolbox. This can facilitate the monitoring process. For secure data handling, it is recommendable nevertheless, that only few have full access to alter questionnaires and data (i.e. add, edit and delete submissions), while the others only own viewing and download permission.

IV. Data analysis

KoboToolbox has different data analysis and visualization tools integrated. All the teams agreed that downloading the data through SPSS or Excel for cleansing and analysing purposes is better suited for high quality data management. Team Benin conducted test analysis during the data monitoring phase, which was an advantage for posterior analysis.

Conclusions: Good practices and lessons learned

The survey design and training of enumerators are two important steps in assuring a smooth implementation and high data quality (Sue & Ritter, 2012). However, it is also important to take a holistic approach and consider all aspects of the survey process equally from the start. In doing so, the likelihood of fulfilling the survey and study objectives is increased. For example, having in mind how the answers will be analysed and what methods will be used (final phase) are key when formulating a question (first phase). In other words, it is better to orient your questions based on indicators you want to know and not the other way around. However, if the empirical subject cannot provide indicators upfront, indicators can be explored and developed through a survey by asking open-ended questions. This is, however, not part of this method description. All groups to some extent underestimated the statistical analysis portion of the project. To be prepared for data analysis, an indicator should be first determined, after which comes question formulation and the parallel development of the statistical formula.

All teams also emphasised the importance of taking time in piloting and conducting an enumerator

training. Team Benin found it challenging for tandem partners to recruit and approach participants in the market, while team Cambodia reduced this challenge by outsourcing data collection to an experienced local firm. For team South Africa, working with enumerators living in the communities relevant for the research was crucial not only for data quality factors, but because they were involved and identified with the project. Concerns related to lack of trust and context knowledge were thus reduced. However, this may have potentially led to a sampling bias. Therefore, explanation of the sampling strategy during the enumerator training is very important.

In the context of COVID-19 pandemic, the advantages of implementing a mobile survey using KoboToolbox were greater than the challenges. Nonetheless, remote research requires as much foresights and rigorous design considerations than more traditional research modalities (Sue & Ritter, 2012). Lastly, it also requires constant communication and consultation with partners on site.

Literature

Andrews, D., Nonnecke, B., Preece, J. (2003) Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*. 16, 2, 185-210.

Andrews, D., Nonnecke, B., Preece, J. (2003) Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*. 16, 2, 185-210.

Andrews, D., Nonnecke, B., Preece, J. (2003) Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*. 16, 2, 185-210.

Andrews, D., Nonnecke, B., & Preece, J. (2003). Conducting research on the internet: Online survey design, development and implementation guidelines. *International Journal of Human-Computer Interaction*, 16 (2), 185-210. https://doi.org/10.1207/S15327590IJHC1602_04.

FAO (2018). Implementing the Food Insecurity Experience Scale (FIES) in surveys. Available at: http://www.fao.org/fileadmin/user_upload/voices_of_the_hungry/docs/Implementing_the_FIES_in_surveys_28218_04.pdf (accessed 10.03.2021).

Fielding, N. G., Lee, R. M., & Blank, G. (Eds.) (2008). *The SAGE handbook of online research methods*. London: Sage Publications Ltd.

Greenleaf, A. R., Gibson, D. G., Khattar, C., Labrique, A. B., & Pariyo, G. W. (2017). Building the evidence base for remote data collection in low-and middle-income countries: Comparing reliability and accuracy across survey modalities. *Journal of medical Internet research*, 19(5), e140. <https://doi.org/10.2196/jmir.7331>.

Hensen, B., Mackworth-Young, C. R. S., Simwinda, M., Abdelmagid, N., Banda, J., Mavodza, C., Doyle A.M., Bonell, C. & Weiss, H. A. (2021). Remote data collection for public health research in a COVID-19 era: ethical implications, challenges and opportunities. *Health Policy and Planning*, czaa158. <https://doi.org/10.1093/heapol/czaa158>.

Henze, J., Paganini, N. & Stöber, S. (2020). Switching over instead of switching off: A digital field research conducted by small-scale farmers in Southern Africa and Indonesia. In: H. Kara & S. Khoo (ed) *Researching in the age of COVID-19. Volume I: Response and reassessment*. Bristol: Bristol University Press, Policy Press. <https://doi.org/10.2307/j.ctv18gffz2s.13>. JDC (Joint Data Center on Forced Displacement) (2021). Enhancing Kobotoolbox for data collection and analysis. Available at: <https://www.jointdata-center.org/enhancing-kobotoolbox-for-data-collection-and-analysis/> (accessed 10.03.2021).

KoBoToolbox (2021). About Kobo: The nonprofit behind KoBoToolbox. Available at: <https://www.kobotoolbox.org/kobo/> (accessed 21.06.2021).

Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.) (2013). *Handbook of survey research*. New York: Academic Press.

Sherin, S., Mathew, P., Johns, F., & Abraham, J. (2018). The feasibility of using remote data collection tools in field surveys. *International Journal Community Medicine Public Health*, 5, 81-85. <http://dx.doi.org/10.18203/2394-6040.ijcmph20175514>.

Sue, V. M., & Ritter, L. A. (2012). *Conducting online surveys*. Los Angeles: Sage Publications Inc.

Centre for Rural Development (SLE)

Hessische St. 1-2 & Robert-Koch-Platz 4
10115 Berlin
Telephone: +49 (0)30 2093-46890
Fax: +49 (0)30 2093-46891
Email: sle.agrar@hu-berlin.de

Projects	GIZ-African Union	GIZ-Cambodia	GIZ-Benin	Co-research South Africa
Aim of mobile survey	Explore participant's access to the health system and the effects it has on their health status	Measure food security and livelihood factors to compare agricultural cooperative members and non-members	Explore consumer's preferences in the poultry market	Explore the impact of COVID-19 on food security, agency and local food systems
Country	Kenya, Nigeria and South Africa	Cambodia	Benin	South Africa
Sample unit and target group	Individual people, especially migrants	Households of cooperative members and non-members	Individual consumers	Households of marginalised communities
Recruitment method(s)	Mostly face-to-face interviews, i.e. conducted in health facilities for migrants using mobile phones	Outsourced this to a company on site	Tandem partners visited local markets to interview participants face-to-face	Enumerators send out the link through WhatsApp, Facebook and interviewed face-to-face or via telephone
Length of survey	60 Questions	93 Questions	202 Questions	57 Questions
Language	English	Khmer	French	Afrikaans, English and Xhosa
Sample size	965	294	304	1827
Data analysis software	SPSS	SPSS, R	SPSS	Excel, SPSS, R

Table 1: Overview of the mobile surveys conducted during four SLE research projects in 2020