

37th WEDC International Conference, Hanoi, Vietnam, 2014

**SUSTAINABLE WATER AND SANITATION SERVICES
FOR ALL IN A FAST CHANGING WORLD**

**Mobile phones for collecting WaSH data
in low-income countries**

E. Tilley & I. Günther, Switzerland

REFEREED PAPER 1871

Based on our experience using both paper-based and mobile phone data collection methods, we consider the advantages and disadvantages of each in relation to water and sanitation research in low-income countries. We compare the two methods in terms of Usability, Data Quality, Data Monitoring, Data Transfer, Cost, and Ethics and Corruption. We conclude that paper-based surveys are best suited for small sample sizes and/or in locations with unreliable mobile networks. Otherwise, mobile phones are an excellent, increasingly low-cost and easy to manage method for high-quality data collection.

Introduction

Essential to continuing and improving the provision of water and sanitation in low-income countries is monitoring and evaluation; the foundation of which is data... The Joint Monitoring Program (JMP) run by the WHO and UNICEF (World Health Organization and UNICEF 2010) is one of the most extensive examples of WaSH data collection in low-income countries, but research programs at all levels are interested in collecting data in order to understand and improve all aspects of water and sanitation needs, provision, use, and maintenance.

Typically, field data have been collected by enumerators wielding clipboards and pens; toilets were counted, residents were interviewed, and the numbers were carefully recorded on paper forms. The problem is that pages are easy to lose; messy hand-writing is sometimes impossible to read; rain and mud can destroy many days of carefully written notes; transferring the data to a digital format induces another source of error; and carrying huge stacks of multi-page questionnaires can be back-breaking work.

Collecting data digitally is not new, but it has recently become cheaper, easier and more flexible than in the early 2000s when Personal Digital Assistants (PDAs, the larger, pre-cursor to the smart phone) and laptops were gaining interest; programmed checks were found to improve the accuracy of financial data on the quality of data collected by PDAs (Fafchamps, McKenzie et al. 2012). Similarly, questionnaires delivered using Ultra Mobile Personal Computers (7 inch touch screen) were found to contain fewer errors than the same questionnaire administered with paper (Caeyers, Chalmers et al. 2012). However, PDAs were expensive, big and difficult to program. Moreover, unlike mobile phones, PDAs could not transmit the information over the cellular network (Shirima, Mukasa et al. 2007, Seebregts, Zwarenstein et al. 2009).

The rise of mobile phone ownership and the associated networks and services that grow along with it, have made it not only possible, but logistically and financially compelling to consider the use of mobile phones to collect field-data in low-income countries. Mobile phones have now been used in various locations to conduct verbal surveys with dispersed, difficult to reach populations (Croke, Dabalen et al. 2012).

Between 2011 and 2013 we conducted extensive data collection on water and sanitation in eThekweni Municipality, South Africa using both traditional paper methods and mobile phones. Based on these experiences, we are able to evaluate the relative merits of paper and mobile phone based data collection. Tomlinson et al. (2009) evaluate the advantages of mobile phones compared to PDAs. In terms of data-collection capabilities, PDAs were the predecessor to mobile phones, but they had few of the advantages and yet retained many of the disadvantages of paper: PDA-based data still had to be downloaded manually and

could not be reviewed in real-time away from the field. And though a survey could be revised more easily than on paper, the additional hardware costs were hard to justify in terms of reduced paper copies. Paper methods are not obsolete, though PDAs may be. Hence, we are interested to know the contexts in which mobile phones or paper-based methods would perform best.

In the following sections we compare the two methods in terms of Useability, Data Quality, Data Monitoring, Data Transfer, Cost, and Ethics and Corruption.

Methods

Using mobile phones, we conducted two household surveys (one in early 2011 and another in mid-2013) with about 1000 households as well as two household urine measurement campaigns (one in late 2011 and one in early 2013) with about 500 households. We also ran field experiments over the course of 9 months (in-between of the household surveys) and collected data using paper forms.

The goal of the household surveys was to obtain socio-economic data and information about sanitation investment and practices. The questions were in the form of multiple choice (e.g. what is your first language: isiZulu, Xhosa, other), numeric (how many people live in your household?), and some open questions (e.g. “why do you not like your toilet?”).

The goal of the urine production survey was to determine how much urine was produced at each household: urine tanks were weighed 3 times a week (Monday, Wednesday, Friday) for 4 weeks. The input data was numeric (e.g. 4.72 kg). For both types of survey, date, time and GPS coordinate data were taken automatically.

Paper forms were used to record information about people who participated in a field experiment that financially compensated people for the volume delivered to a central collection point. Enumerators recorded the names and ages of participants as well as the conductivity, pH and volume of the urine delivered. The participants were paid, the value was recorded and the participant signed the form to acknowledge that they had received the correct payment. The decision to use paper-forms for this work was based on the need to have a visual record of signatures so that customers could confirm their current and past payments.

There are a variety of platforms that can be used to program the user-interface on the phone and/or monitor the data as they are sent. A list of 12 alternative platforms are listed in the Notes section at the end of paper. We worked with the South Africa-based company Mobenzi (www.Mobenzi.co.za), partly because there was an existing relationship between them and the university with which we cooperated (University of KwaZulu-Natal), because they were local, and because they provided a range of services in addition to the software (e.g. phone rental). The Mobenzi console allows administrators to see and sort data on a computer as soon as the enumerator sends it. Although we are most familiar with the Mobenzi set-up, the general experiences should be comparable to other services/programs available on the market.

Discussion

Useability

We used a mid-range phone: a Sony Ericsson Elm with photo and GPS capabilities. The phones instilled a sense of pride among the enumerators; they saw their work as being “high tech” and the phones gave them a sense of prestige. Several-day workshops were held to ensure that the enumerators using the mobile phone were trained to enter the data properly and how to trouble-shoot in the event of problems; practice with the surveys and clear explanations about the type of data that should be collected was given to both paper and phone-based enumerators. One common misconception is that mobile phones can be given to respondents for them to answer questions themselves. While some researchers have equipped local experts (e.g. farmers) with phones and training (Dillon 2012), one cannot assume that general mobile phone literacy means that an average household can handle foreign phones, follow instructions, and enter data properly. We therefore consider here (and generally recommend) the situation where trained staff uses mobile phones to collect data without anyone else using the phone.

Unfortunately, the phones were too complicated for some of the enumerators, and after extensive training they had to be let-go as they still could not master the key-strokes needed to input the data. These enumerators may have been able to administer a paper survey, however their limited ability to learn may have meant that they were neither ideally suited to the more complicated task of following “skips and jumps” in paper based surveys without a pre-programmed logic. “Skips and jumps” refers to a way of

bypassing certain questions based on the response to a previous one (for example, no questions about schooling would be asked to a woman without children).

Because the phone was in use all day, the fieldworkers had to charge the phones every night, and not only was this difficult (during power-outages) it was a cost that was inflicted on the fieldworker (although they were later compensated). Fortunately, we had only one lost phone, and two broken ones (out of, approximately 35 phones used during the surveys). However, since the data were uploaded immediately, we had no data loss and replacing the broken screens was a small price to pay for not losing a full day of data.

Our first household survey consisted of 204 questions (which would have filled over 20 pages) and each enumerator was responsible for administering 100 surveys. Having the survey on the phone meant that the enumerator not only did not have to carry 2000 pages around, in hot, humid conditions, but they also did not stand out. Safety is a constant concern for women in South Africa, and being able to quickly and easily slip their mobile phone away, meant that they could become anonymous- not an obvious employee of a government or foreign project- and move easily without lugging a large bag of paper in the event that they would need to move quickly.

Data quality

One of the biggest challenges associated with paper-based data collection is the quality: both in terms of the visual quality (clarity of writing) but also completeness (not missing questions) and data type (not writing “yes” when the question requires a number). Moreover, intensive training and practice is required when issuing a paper-based survey, so that enumerators can skilfully skip over questions that don’t apply, and ensure that extra information is obtained when a key question is answered.

Complicated logic, question and section skips, as well as data quality checks can all be programmed into the mobile phone interface, but this does not negate the need for enumerator training, though it may be considerably reduced. Depending on what type of toilet the respondent had (pit latrine, urine-diverting, VIP), we had separate modules of questions that the surveyor would be presented with; the programmed logic ensured that the enumerator did not accidentally ask question about urine separation to someone with a pit latrine. Similarly, when measuring the weights of the urine tanks, the respondent was forced to enter a value with two decimal places; also, values for the 20L tanks were not accepted over 25.00 kg.

Collecting information in a matrix format, e.g. the names and ages of each family member is generally easier on paper; programmed loops that repeat over the same question set for each family member are possible, but cumbersome to enter.

Programming logic does not prevent mistakes and indeed the programmers (us) made several. The difference between programming mistakes and enumerator mistakes is that the programming mistakes are systematic and can be quickly noticed in the data (i.e. if a question is consistently skipped). However, the freedom to program logical jumps, might push the survey designer to try ever more complicated combinations of loops and skips, which, may be unnecessary and may cause more errors in the end. In both paper and phone-based surveys, open-text questions, though a source of rich data are prone to spelling mistakes. Text based answers are, however, easier to read and analyse using digital methods. The drawback is that the length of the answers is more limited than applying paper-based surveys.

One of the biggest advantages of mobile phone based interviews is the fact that if, during the course of the survey, the researcher notices that there is a problem with the survey, she can simply change the survey send the update to each phone instantly, for the price of an SMS. With paper-based surveys, the researcher would have to print out new copies and deliver the revised paper version to all of the enumerators, which could be expensive, difficult, or in the worst case, impossible.. Hence, mobile phone surveys mean that fewer of the old surveys are completed and less time and financial resources are required.

Data monitoring

Depending on the type of data collection, paper-based data are usually returned to the researcher at the end of the day, so that they can be checked for mistakes and missing data. This can involve extensive transport and logistics and, for the researcher reviewing the work, many sleepless nights. Data collected via mobile phones are transferred over the cellular network nearly instantly, so that a researcher, sitting at a computer, can monitor the data as soon as they are uploaded. The benefit here is that a researcher can quickly flag poorly programmed logic, bad or missing answers. Enumerators can be contacted instantly (they are carrying a mobile phone after all) and asked to re-do the question or survey, without having to back-track too much. Alternatively, if it is not feasible to re-enter the data digitally, the researcher can make a note of the real response (e.g. the age was 44 not 14) before the enumerator forgets by the end of the day. This type

of rapid review and correction is nearly impossible with paper-based data. However, the data can only be monitored if they are transferred, a problem we address in the following section.

Data transfer

Transferring paper-based data requires physical movement, either by person or vehicle or both. In the event that distances are very far, data transfer may be infrequent, which may therefore, reduce the frequency of data monitoring. The communities where we conducted the household surveys were between 30-50 kms from our research office. Although we made frequent visits to the field, it was difficult to visit more than one community a day, and it would have been impossible to visit each of them daily. Furthermore, data must usually be digitized from the paper version and transferring the data from paper to a new format induces another source of error (through typographical errors).

Transferring data from mobile phones to a central collection point (the researcher) is done across a cellular network. Within regions that have adequate networks, data transfer happens quickly, almost without the risk of data loss. Additionally, mobile phones can be used by enumerators in remote or distant areas without having to return to a city or central collection point to physically transfer the data. This allows a researcher to employ local staff who, may not only know the area better, but who don't need to spend extra time and money travelling back and forth to the research areas.

Unfortunately, all of the advantages of mobile phones only apply if and when there is a functioning, consistent mobile network. We (incorrectly) assumed that, because our enumerators testified to the strength of the network signal in their area, that it was independent of service provider, which was not the case. All of our research phones were connected to the same network (Vodacom), which, while very good in some areas, was completely absent in others. Enumerators with no signal, resorted to doing surveys in the day, and travelling to urban centres to upload their data in the afternoons.

Moreover, even though surveys can be easily saved on the mobile phones until they are uploaded, in our experience, we found that about nine to ten surveys could be stored before it became difficult to upload-trying to upload ten full length surveys at the end of the day was painfully slow, and took much longer than the sum of the time needed to upload the individual surveys. We therefore would not recommend that more than one day worth of surveys are stored on the phone before uploading. Furthermore, we had no way of analysing centrally how much mobile credit had already been used, or how much remained; the ability to centrally track credit usage was not available to us and could only be seen on the phones. On several occasions, we shipped enumerators back and forth to cities, trying to get their data to download, only to find out that they were out of credit. The lack of credit was partly because we had under-calculated the requirements, and partly because enumerators used the mobile phones for personal use (messaging, Facebook, etc.).

Ethics and corruption

Although some of the phones and the credit got used for personal use, the issue of "abuse" was not a concern; enumerators took exceptional care of their phones.

When a paper survey is administered, it is good practice to ensure that a field-manager is within easy reach to ensure that questions and problems can be resolved, but in reality, many enumerators do not see the field manager until the end of the day when they return their completed pages. During the enumerator training, the concept of data transfer and the monitoring console were explained; enumerators fully understood that we were watching their activity on the computer. Knowing the time and location (GPS marked) allowed us to ensure that fieldworkers were not only doing their work, but doing it in the right location (i.e. not sitting in one spot and filling in fictitious answers to meet the daily quota). While it may seem invasive, enumerators were happy to know that their work was acknowledged and not lost ("do you see my surveys?" was often asked, especially at the start of the project), as well as knowing that they could keep their own hours as long as the work was finished. (The temperature in KwaZulu Natal can reach 40°C by mid-day, so enumerators who wanted to start at 5:00, were free to do so and didn't have to wait for a field manager to show up).

Collecting panel data, i.e. taking two measurements from the same place or person at two different points in time, is notoriously difficult in poor settings. House numbers, phone numbers, family names, assigned ID numbers and increasingly, GPS have been used to match data points across time.. New technology also allows fingerprints to be scanned and stored at the time of an interview, so that the exact same respondent can be identified and matched precisely. Digital data collection allows multiple data types to be linked perfectly, which is ideal for the researcher, but could be devastating to the interviewee if it was lost - but the

same problem also applies to paper based surveys. The Mobenzi platform uses encryption. Still data that is transmitted digitally is at greater risk of corruption than paper based surveys.. The quest for panel data raises some serious questions about anonymity, data storage and the concept of “informed consent”, i.e. whether or not interviewees fully understand what they are committing to (Whittington 2004).

All types of data collection require careful storage and use, but the ability to collect greater quantities of more detailed data with mobile phones will require researchers to apply higher standards to their protocols.

Costs

We tried to choose the most cost efficient method for the location and amount of data that was required.

The 20 page household survey (with various skips) was conducted over a month, by 12 different fieldworkers, spread over 100 square kilometres. Although the survey set-up time, phone rentals, and network fees (air time) were expensive there were no paper-copy, transport, supervisor, and data entry costs. All data was collected independently in the field, and reviewed immediately by the researchers in the office. Phone rentals were about \$2 USD/day, but transport costs to visit 2-4 fieldworkers to collect paper surveys would have cost 10-20 times that amount per day; the presence of a supervisor in each area would have added an additional \$30 USD/day and data entry another \$30 per day (for about 10-20 questionnaires per day and person). For the geographically dispersed household survey with multiple, parallel fieldworkers and a complex questionnaire, the higher up-front costs were quickly recovered by removing the need for transport, on-site supervision, and lengthy data entry.

Conversely, the experimental data were collected by only 3 fieldworkers who were each employed for over 8 months. The data collected consisted of 8 short answers from each person who participated, so the information could clearly and easily be recorded on a simple table; there was no need for complicated programming, and the simple tables could be photocopied at any local shop for a few cents a copy. Furthermore, because the data required was quite simple and repetitive, and because enumerators were employed for so long, they quickly became experts in their work, and so made very few mistakes. Visits were made by the researchers sporadically to collect the data sheets and provide new ones. Transport was minimal and could not have been avoided, even if digital data collection was employed, since it was necessary to visit the field sites regardless (for deliveries, to pay salaries, meet community members, etc). In this context, airtime and phone rentals would not have helped produce cleaner data or reduced the need for transport: paper-based data collection in this long-term, study with highly-trained fieldworkers and infrequent travel was the most cost effective method.

Conclusions

The choice between paper and mobile phone-based data collection methods is not clear-cut. Our general conclusion, based on extensive experience using both, is that paper-based data collection is better for small sample size, short-term, geographically contained research programs, and where the results can be checked frequently. Paper-based data collection is necessary when there is a weak or unstable cellular network and are easier to apply when more complex, matrix based or open-text questions are the dominant form of information collection.

Mobile phone-based data collection is, in general, more flexible and scalable. It is therefore best suited to long-term research that covers a large region and where constant contact with the data-collectors is not possible. Complicated questionnaires with different logic-based skips and modules are completed more efficiently and reliably with mobile phones, though mistakes from enumerators and programmers are not eliminated.

At the end of the day, the quality of the data will depend on the amount of time that is spent on the design, the training, the piloting, and the clarity of the research question: if you don't know what you're looking for, not even the newest smart phones will help.

Acknowledgements

The author/s would like to extend thanks to Chris Buckley and Kerry Philip at the University of KwaZulu-Natal, Teddy Gounden and Scelo Xulu at the EWS unit of the eThekweni Municipality, the staff at Mobenzi, especially Marc Yunnie, Leeanne MacGregor, and all the enumerators, in particular Bongekile Mngwengwe, Nomvula Ncwane and Mondli Ngidi without whom this research would not have been possible.

References

- Caeyers, B., Chalmers, N. and De Weerd, J. (2012). *Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment*. Journal of Development Economics Vol. 98, No 1, pp 19-33.
- Croke, K., Dabalén, A., Demombynes, G., Giugale M. M., and Hoogeveen J. (2012). *Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews*. World Bank Policy Research Working Paper. The World Bank: Washington.
- Dillon, B. (2012). *Using mobile phones to collect panel data in developing countries*. Journal of International Development Vol. 24, No. 4, pp 518-527.
- Fafchamps, M., McKenzie, D., Quinn S. and Woodruff C. (2012). *Using PDA consistency checks to increase the precision of profits and sales measurement in panels*. Journal of Development Economics Vol. 98, No. 1, pp 51-57.
- Seebregts, C. J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A. J., Seebregts, C., Mukoma W. and Klepp, K.-I. (2009). *Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm Pilot interviewing system*. International Journal of Medical Informatics Vol. 78, No. 11, pp 721-731.
- Shirima, K., Mukasa, O., Schellenberg, J.A., Manzi, F., John, D., Mushi, A., Mrisho, M., Tanner, M., Mshinda H. and Schellenberg D. (2007). *The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania*. Emerging Themes in Epidemiology Vol. 4, No. 5, pp 1-8.
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai A. and Jackson D. (2009). *The use of mobile phones as a data collection tool: A report from a household survey in South Africa*. BMC Medical Informatics and Decision Making Vol. 9, No. 51, pp 1-8.
- Whittington, D. (2004). *Ethical Issues with Contingent Valuation Surveys in Developing Countries: A Note on Informed Consent and Other Concerns*. Environmental and Resource Economics Vol. 28, No. 4, pp 507-515.
- World Health Organization and Unicef (2010). *Progress on sanitation and drinking water*. WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP).

Note/s

Other available platforms, which may be free, low-cost, or full-service is provided below, though we have no experience with them, and cannot comment on the quality or availability of the services. Cortex Software (cortexsoftware.com); Formhub (<http://formhub.org>); Frontline SMS (frontlinesms.com); iFormBuilder (iformbuilder.com); Jana (jana.com); Text to Change (texttochange.org); Viewworld (viewworld.net); dooblo (dooblo.net); Magpi (datadyne.org).

Contact details

Elizabeth Tilley
 Clausiusstrasse 37, 8092 Zurich, Switzerland
 Tel: +41 44 632 94 58
 Fax: +41 44 632 12 07
 Email: tilley@nadel.ethz.ch
<http://www.nadel.ethz.ch/people/tilleye>

Isabel Günther
 Clausiusstrasse 37, 8092 Zurich, Switzerland
 Tel: +41 44 632 87 50
 Fax: +41 44 632 12 07
 Email: guenther@nadel.ethz.ch
<http://www.nadel.ethz.ch/people/isabelg>
