# Comparing smartphones to tablets for face-to-face interviewing in Kenya

**Survey Methods: Insights from the Field**

**Sarah M. Hughes | Samuel Haddaway | Hanzhi Zhou**

**How to cite this article :** Hughes S., Haddaway, S. & Zhou, H. (2016). Comparing smartphones to tablets for face-to-face interviewing in Kenya, Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=7031

**DOI :** 10.13094/SMIF-2016-00001

**Copyright :** © the authors 2016. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

**Abstract :** Research conducted over the past 30 years has demonstrated a reduction in errors and improvement in data quality when face–to-face social surveys are carried out using computers instead of paper and pencil. However, research examining the quality of data collected by interviewers using mobile devices is in its infancy and is based in developed countries. In a small pilot study conducted during the World Bank's Kenya State of the Cities Baseline Survey, a face-to-face survey on living conditions, infrastructure and service delivery, the authors compared the quality of data collected using smartphones to data collected using tablets. The study of mobile touchscreen devices showed that tablets outperformed phones in some cases, but that the results were highly dependent on the interviewer.

## Background

Research conducted over the past 30 years has demonstrated a reduction in errors and improvement in data quality when face-to-face social surveys are carried out using computers instead of paper and pencil (Banks & Laurie, 2000; de Leeuw, 2008; Schrapler, Schupp, & Wagner, 2010). Most studies of data quality by survey mode and by device type have been conducted in developed countries, where the vast number of surveys conducted for policy, marketing, and other purposes provide opportunities for methodological research. Caeyers, Chalmers, and De Weedt's (2010) study comparing paper and pencil interviews (PAPI) to computer-assisted personal interviews (CAPI) in Tanzania represents a rare example of an experimental study on survey methodology in a developing country, and the results confirmed that the internal validation checks that are programmed into CAPI questionnaires to detect skip errors, implausible answers, and impossible answers led to a substantial reduction in errors compared to surveys conducted using PAPI. An experimental study in Fiji by Yu et al. (2009) found that none of the errors observed in 20.8% of the paper questionnaires were found in the CAPI versions and that the PDA-programmed version led to cost and time savings compared to paper forms. Aside from these rare examples, most research on mode effects in developing countries consist of non-experimental studies carried out as pilots within ongoing surveys or conducted retrospectively when a new mode is adopted on a longitudinal survey. Most results have paralleled those found in earlier developed-country studies comparing PAPI to CAPI; researchers have found interviewer error on CAPI surveys is lower than they might have expected if they had used paper surveys, and this perceived reduction in error is likely attributable to enforced skips (Trott & Simpson. 2005; Siekmans, Ngnié-Teta, Ndiaye, & Berti, 2012; for an

alternative view, see Escobal & Benites, 2013).

International funders and organizations carrying out data collection projects are eager to adopt computerized methods for data collection. Indeed, the United Nations Economic and Social Affairs Statistics Division explicitly recommends that, "in all cases, data should be collected in electronic format wherever possible, as this facilitates data capture and editing" (United Nations Statistics Division, 2014). Yet, adoption of CAPI surveys for data collection in developing countries has been slow through the mid-2010s. Based on the authors' collective experience and anecdotal information from survey managers conducting surveys in developing countries, this is likely due to obstacles such as the cost and availability of hardware and software for surveys, a relatively short battery life for laptops, the need for frequent access to an electrical current, the relative fragility of the hardware, the lack of reliable mobile networks for data transmission, and limited experience with questionnaire programming and CAPI management on the part of in-country survey organizations.

Recent advances in lower-cost, lighter weight mobile devices such as smartphones and tablets with longer-lived batteries, user-friendly interfaces, and easy programming coincide with the rapid expansion of mobile networks to produce an opportune time for adopting CAPI instruments for surveys in developing countries. Tomlinsen et al. (2009), among others, suggest that the ease of use and familiarity of mobile phones could make them more useful for data collection than other CAPI hardware. The World Bank has taken a leading role in expanding mobile-platform surveys by developing a mobile questionnaire and survey management tool for use on the global Living Standards Measurement Survey (Carletto, 2015) and other World Bank-sponsored surveys. The United States Census Bureau also has developed a mobile version of its free CSPro survey questionnaire software. However, little is known about the impact smaller devices have on the quality of data when used for face-to-face interviews. Instead, research on device-mode effects on data quality have been carried out primarily on self-administered questionnaires (SAQ).

On SAQs, whether computerized or paper and pencil, respondents must process information that generally appears before them in a static form: textual, numeric, symbolic, and graphic (Redline & Dillman, 1999). In contrast to SAQs, CAPIs include the intervening presence of an interviewer, who delivers the question orally and provides, in the gold standard method, only pre-defined interpretations of the question. But while the contextual differences between SAQs and CAPIs are substantial, the existing research on SAQs is nonetheless instructive for implementers of any type of CAPI data collection, since human-computer interaction (HCI) is necessary for an interviewer to complete the survey. Studying the smaller size of mobile devices, Bruijne and Wijnant (2013) found that self-administered web surveys carried out on mobile devices took longer to complete, perhaps due to formatting differences, than the same survey on desktop computers. Mavletova (2013) also found that durations were longer when respondents used mobile devices to complete a survey compared to PC or laptop, although only a portion of the longer duration was due to respondents finding it more difficult to complete questions. Rather, slow question loading explained most of the difference. Lugtig and Toepoel (2016) found larger measurement error when smaller devices were used for an SAQ, although they surmised that this error might be due to respondent characteristics rather than device, per se, wherein respondents who choose to use smaller devices might differ in substantive ways from those who choose to use larger devices such as desktops or tablets. Other studies of HCI suggest that mobile phones may not be an optimal replacement for the larger screens of laptops and larger mobile devices such as tablets for completing questionnaires. Peytchev and Hill (2010) found that small keyboard size led to avoidance of open-ended questions in an experimental mobile self-administered survey. Peytchev and Hill point to a broader literature on HCI, which shows that task success rates, such as correct selections, are lower on smaller screens. Applying this growing body of research on the effect of screen size on the quality of survey data, we suspect that device size could affect the quality of survey data entered by interviewers. Even if a

respondent provides a lengthy response to an open-ended question or a well-considered response to a closed question, the interviewer may short-cut or mis-select responses at a higher rate on a smaller device, thus altering responses and curtailing quality.

Experimental research on CAPIs under field conditions in developing countries is rare and to date we can find no experimental comparisons of device size impact on interviewer data quality in such settings. As a first effort, using data from a small pilot study conducted during a large-scale CAPI survey in Kenya, we compare the influence of device size on the quality of survey data collected by interviewers using tablets or smartphones. By assessing interviewer data quality in terms of thoroughness (low number of missing responses and high rate of GPS coordinate capture), accuracy (correct data entry), and consistency (mean duration), we explore the influence of device size on interviewers' administration behaviour. In our analysis we assume equality in experience across the interviewers, "John" and "Jane," but we also collected information on the interviewers' perceptions of the two devices to better understand the individual user experience. We hypothesize that data collected on smartphones will be of lower quality than data collected using tablets. We expect that lower quality will be seen through a higher number of missing responses and GPS capture, more errors in numeric or text entry, and shorter or implausible durations, and that these indicators of lower quality are linked to the use of the smaller screens and keyboards on smartphones.

*Thoroughness (low missing data)*: Item nonresponse is one of two main types of nonresponse error (the other being sample unit nonresponse). Rates of item missingness, including "Don't Know" (DK), "Refuse" (REF), and "Not applicable" (NA), are routinely used as markers of interviewer data quality in surveys under the expectation that "good" interviewer behaviour will lead to high cooperation and willingness from respondents to provide responses other than DK/REF/NA (Groves, 1989; de Leeuw, 2001; de Leeuw & Huisman, 2003; Jans, Sirkis, & Morgan, 2013). Recent research on questionnaire design suggests that item nonresponse differs by device type (Mavletova & Couper, 2014).

*Accuracy (correct data entry):* Training interviewers to correctly enter numeric and text strings is a strategy for reducing other interviewer-related measurement error, such as out of range responses or mis-recorded responses (Biemer & Lyberg, 2003; Fowler, 2004). Whether entering case ID codes, monetary values, or telephone numbers, correct and complete numerical data entry is a key interviewer skill for ensuring data quality.

*Consistency (mean duration)*: Survey managers track the average duration of survey interviews as part of process management and as a useful indicator of interview quality (Olson & Peytchev, 2007). For process management and budget control, the expected duration of the interview is determined during pretesting of the instrument and re-estimated in the early field period. These benchmarks are used during the field period to identify outlier cases for further scrutiny or to identify interviewers whose average duration is outside the expected range. Duration in a personal interview can correlate to cooperation and rapport (Holbrook, Green, & Krosnick, 2003), is simple to measure, and acceptable ranges are relatively easy to set and monitor. Differences in duration can be understood in a variety of ways. Shorter times may indicate a high degree of rapport and cooperation between respondent and interviewer or suggest efficiency on the part of the interviewer. On the other hand, shorter duration might suggest shortcutting or speeding on the part of the interviewer. In their 2013 study of response times, Couper and Kreuter found that questionnaire items with interviewer instructions took less time to administer than items without instructions, leading the authors to surmise that interviewers might not be reading the instructions. Unobtrusive computer audio recorded interviewing (CARI) studies support this finding; in a study of interviewer effect on data quality, Kosyakova, Skopek, and Eckman (2014) found that CAPI interviewers manipulate the triggering rate of filter questions and that this undesirable behaviour

increased over the field period. When interviewer pay structure is per-completed-case, speeding might be a logical approach to maximizing wages.

**Methodology**

Data for the World Bank's Kenya State of the Cities Baseline Survey was collected from July 2012 to March 2013. The survey supports the Kenya Municipal Program (KMP), a long-term effort to improve living conditions through infrastructure investment and service delivery in 15 municipalities in Kenya. The survey portion of the State of the Cities project included two main tasks: 1) creating a sample frame based on listing a projected 194,000 households in 2,087 enumeration areas (EAs) in 15 of Kenya's largest cities and, 2) carrying out interviews of 30-45 minutes' duration with approximately 14,600 households randomly selected from the sample frame. Listing and interviewing were carried out concurrently using tablet computers. Teams of data collectors used tablet-programmed listing forms to enumerate all households contained within each EA. Next, interviewers uploaded the listing data to a server via the mobile network using their SIM card-enabled tablets. The data were captured in a server accessible via a web interface. The data collection team sampled households from each fully listed EA using the web interface, and then transmitted the selected case data, including household identifier, location, and descriptive data, to interviewer tablets. Finally, interviewers contacted the selected households for interviewing. At the end of each day, all completed survey response data were transmitted to the server via the mobile network, and all data were accessible for review and processing through a web interface.

As part of a grant from the Center for Excellence in Survey Research at NORC at the University of Chicago, the research team selected two KMP survey interviewers, "John" and "Jane," to carry out 200 of their assigned household interviews using smartphones instead of tablets. The selected interviewers had several years' experience working with the data collection company, demonstrated high production on social surveys, and were considered to collect high quality data, according to the data collection manager (*n.b.* the criteria for this determination was not clear, and no specific data supporting the rating were provided to the authors). Midway through the tablet data collection period, the two interviewers conducted approximately 50 interviews each using smartphones in two cities, Nairobi and Thika, to reach a total of 200 interviews. To complete these interviews, interviewers simply switched devices until they had completed 50 interviews in each city. The application and interface was identical on both the phone and the tablet, with no differences in functionality; the sole difference between the devices were the screen and keyboard size. The cases completed on phones were all "fresh"; in other words, respondents had not been previously contacted by the interviewers and were not pre-screened in any way. By performing the pilot study in the middle of the field period, we were able to ensure that the interviewers were already familiar with the software and that any differences in quality would be attributable to device effects.

The purpose of the research was to permit comparison of the quality of data collected using smartphones to the quality of data collected using tablets. We compared the data in terms of missing responses (thoroughness), mistyped phone numbers (accuracy), and mean duration of the interview (consistency). We also carried out qualitative interviews with the interviewers and their supervisor to gain a more textured understanding of their experiences with the smartphone and tablet, and their preferences in using the two different devices.

We performed two-sample t-tests to determine whether there was a significant difference in the several indicators of survey quality between interviews conducted with phones compared to interviews conducted with tablets using several different comparison methods. Below, we discuss the results for each indicator.

## Results

In our initial research design, we planned to compare the data collected using phones to the data collected using tablets by combining our two phone interviewers' results and comparing to tablet data collected by all interviewers in all 15 cities. However, we found significant differences in outcomes between the two interviewers participating in the experiment. This made the analysis more challenging as the two distinct interviewer profiles reduced our ability to make generalizations to other interviewers. This also reduced our sample size since combining their results distorted the output; therefore we analysed their results separately. However, the interviewer-specific differences provided an opportunity to explore individual interviewing styles and experiences, and how these interacted with the two mobile devices. Below, we present the results of our quantitative analysis of the data and include illustrative or explanatory qualitative data where appropriate. For each dimension, we present each interviewer's tablet results compared to his or her own phone results. The two interviewers' results are presented side-by-side to create an easy visualization of the differences between their results.

### *Thoroughness (missing data: DK/REF/NA)*

For this analysis, we compared our two interviewers' rates of missing items and found that only one of the two interviewers demonstrated a significant difference in item missingness by device mode. As shown in Table 1, interviewer Jane showed a statistically significant lower proportion of missing data for tablet interviews compared to phone interviews, using all variations of comparison groups, at a level of 0.84% to 1.5% lower, while John's rate of item missingness was similar on both devices under all comparison scenarios.

**Table 1: Difference in mean percent of missing values (Refused, Don't Know, Not Applicable)**

| Comparison Groups, by device used | Difference in mean % by observations with missing values, by interviewer | |
|---|---|---|
| | Tablets compared to phones (John) | Tablets compared to phones (Jane) |
| Thika tablets compared to Thika phones | -0.15 | -1.32*** |
| Nairobi tablets compared to Nairobi phones | -0.01 | -0.99* |
| Thika & Nairobi tablets compared to Thika & Nairobi phones | -0.11 | -1.14*** |
| All cities compared to Thika phones | 0.01 | -0.84** |
| All cities compared to Nairobi phones | -0.26 | -1.53*** |
| All cities compared to Thika & Nairobi phones | -0.12 | -1.17*** |

*Statistical significance indicated as follows: *=p<0.1, **=p<0.05, ***=p<0.01*

When asked to compare her use of the phone to the tablet, Jane indicated that she was able to type faster on tablets because of the size of the keys. Jane and John both indicated that they were more likely to accidentally mis-select responses on the phones than on tablets. While both of these statements suggest potential drawbacks of phones, we cannot draw a clear link to the higher rate of missing items for Jane's phone.

***Accuracy (typing errors)***

In this survey, both interviewers collected significantly more valid phone numbers on tablets than on phones by nearly every measure of comparison, as shown in Table 2. (Phone numbers were deemed "valid" if they contained the correct number of digits and started with Kenyan prefixes.)

**Table 2: Difference in the mean number of valid phone numbers listed**

| Comparison Groups, by device used | Difference in mean number of valid phone numbers listed, by interviewer | |
|---|---|---|
| | Tablets compared to phones (John) | Tablets compared to phones (Jane) |
| Thika tablets compared to Thika phones | 0.10 | 0.63* |
| Nairobi tablets compared to Nairobi phones | 0.21** | 0.18** |
| Thika & Nairobi tablets compared to Thika & Nairobi phones | 0.15** | 0.17*** |
| All cities compared to Thika phones | 0.10 | 0.20** |
| All cities compared to Nairobi phones | 0.19** | 0.16** |
| All cities compared to Thika & Nairobi phones | 0.14** | 0.18*** |

*Statistical significance indicated as follows: *=p<0.1, **=p<0.05, ***=p<0.01*

The difference between tablets and phones for interviewers accurately collecting phone numbers may have been due to differences in the interaction between the interviewer and the device. First, the keyboard size is smaller on the phone, which might lead to accidental "typos," or errors in numbers when interviewers' fingers touch more than one key. Second, thumb typing with either or both thumbs is typical for keyboard data entry on the smartphone, while interviewers could more easily use all fingers on one hand or, possibly, both hands, to enter data on the tablet keyboard. This study, which was carried out under normal field conditions, did not include capture of typing method by device, although the participating interviewers indicated thumb-typing was most typical on the phones and both thumb-typing and one-handed typing were used on tablets.

Although the analysis showed significantly poorer results for accurately capturing numbers on phones, the two interviewers in our experiment described different experiences typing with the phones. Jane said that she tended to type less (fewer words in text strings) with the phone than when using the tablet and she

found typing easier on the tablet because of the larger size of the keys. This could mean that she skipped some typing tasks on the phone, including entering phone numbers. John did not find that one device was easier for typing than the other.

Alternatively, respondents may have felt uncomfortable giving out their phone number when it was being entered into what may have looked like the interviewers' personal cell phone; when the interviewer used a tablet, confidence may have been higher that the phone numbers were being collected for legitimate purposes. Therefore, we cannot rule out respondent reluctance as a source of error in phone number collection.

***Consistency (mean duration)***

By using the start time and the end time captured in the programmed questionnaire, we calculated the length of each interview on the KMP survey. The overall mean duration (all interviewers) was 24.3 minutes. We performed two-sample t-tests to determine whether there was a significant difference in the mean durations of interviews conducted using phones as compared to interviews conducted using tablets.

As shown in Table 3, John had significantly longer survey durations on tablets than phones in Thika and Nairobi. In Thika, his tablet interviews were, on average, five minutes longer than his phone interviews, and in Nairobi, they were over 11 minutes longer. His mean duration on tablets in all 15 cities was also significantly longer than his mean duration on phones in Nairobi by an average of over eight minutes.

**Table 3: Difference in mean survey durations (in minutes)**

| Comparison Groups, by device used | Difference in mean survey durations (minutes), by interviewer | |
|---|---|---|
| | Tablets compared to phones (John) | Tablets compared to phones (Jane) |
| Thika tablets compared to Thika phones | 5.03* | 0.75 |
| Nairobi tablets compared to Nairobi phones | 11.66*** | 1.22 |
| Thika & Nairobi tablets compared to Thika & Nairobi phones | 0.59 | 1.06 |
| All cities compared to Thika phones | -6.10 | 3.28* |
| All cities compared to Nairobi phones | 8.30** | -2.39 |
| All cities compared to Thika & Nairobi phones | 0.51 | 0.51 |

*Statistical significance indicated as follows: *=p<0.1, **=p<0.05, ***=p<0.01*

In contrast, Jane showed no significant difference in the mean duration of interviews on phones compared to tablets in Thika or Nairobi (separately or combined). However, when comparing all tablet interviews in all 15 cities to Jane's Thika phone interviews, tablet interview durations were significantly longer than Jane's phone interviews by an average of three minutes. In discussions about the phone pilot, Jane indicated that she owned a smartphone and used it for texting. John indicated that he did not have a smartphone. It is possible that Jane's shorter durations on phones were a result of more familiarity with a

similar device and that John's longer durations on phones represent his longer learning curve, but Jane's higher rate of missing and mistyped values on the phone muddy this supposition.

### *Interviewer perceptions and preferences*

As described above, our research plan included gathering impressions from our interviewers on differences using tablets and phones for the data collection. The primary differences, as the interviewers experienced them, can be summarized as the following:

- The tablets attracted more attention than the phones in most interviewing areas or neighbourhoods. Both respondents and other observers wanted to know more about the tablets, such as how much they cost and how they work. As a result of peoples' curiosity, some of the interviewers' activities were stalled, as they felt obliged to "open ourselves up and answer the questions...it can be a problem, and we might not get to the respondents (on time)." Phones did not attract this kind of attention. "When you have the phone, people assume you are a visitor coming to see the neighbours. When you have the tablet, they ask questions about you, assuming you are coming for other reasons." The interviewers also had different perceptions of the smartphones and the tablets depending on their location and the sample to be interviewed:

1. The interviewers preferred tablets in higher income neighbourhoods in Nairobi, as the interviewers felt the tablets helped them appear more professional. They saw this as an advantage for gaining cooperation among white collar and other employed respondents.
2. The interviewers preferred phones in "slums," as they do not stand out like tablets, and are easier to hide in insecure locations.

- There was an adjustment period for the interviewers as they learned to use the smartphones, which could account for some differences in data quality. Smartphones were introduced three months into the field period, and the interviewers indicated that it took a little time to become familiar with the phones. "Our thumbs are used to the tablets and have been using them a longer time. As we continue using the phones, we'll get more used to the phones so it will be more or less the same."
- Jane typed faster and more on tablets than on phones, according to her own review of the experience. The reason, she stated, had to do with the size of the keys. John did not indicate any difference in typing on the two devices.

While discussing the smartphones and tablets, the interviewers described two differences in their interactions with the devices that were particularly revelatory for data collection planning:

- Phones require more scrolling to read questions and select response options, which the interviewers admitted led them to avoid fully scrolling to read questions as they were written. Instead, the interviewers stated that after having spent months doing many interviews, they no longer needed to scroll to read the questions and/or response options. These comments suggest a significant departure from the standard survey methodology and data quality step of reading each question exactly as it is written.
- Interviewers also indicated that the act of scrolling to read response options can lead to accidentally selecting a response with the touch-screen interface. The interviewers indicated that mis-selecting responses occurred more frequently on the phone due to more scrolling needed on the phone than the tablet to view screens. Our analysis is unable to detect mis-selected responses.

## Discussion

Despite our hypothesis that smaller screen size would lead to poorer quality data collected on smartphones than on tablets, our quantitative analysis was not overwhelmingly conclusive regarding differences in data quality collected on tablets versus phones. A lower proportion of valid phone numbers on phones compared to tablets (Table 2) was the only measure on which both interviewers showed significant differences between devices. This result should be taken into consideration when researchers adopt smartphones for this type of household survey data collection. Most social scientific surveys require gathering numeric data, not just limited to phone numbers but also including income, expenditure, quantities, and other numeric values. Accurately recording numbers is challenging for interviewers even on laptops with a full keyboard and, consequently, repeated practice forms an important module in interviewer training for many social scientific surveys. Even simple differences such as the layout of the numeric keypad can affect accuracy and speed of data entry for numbers (Armand, Redick, & Poulsen, 2013), as can the size of the numeric keys (Park & Han, 2010). The smaller keypads on phones may prove to be a source of error for this device type, but both tablets and phones require practice for interviewers to acquire accuracy.

Returning to the surprising result of very different outcomes for the two interviewers, who were selected using the same criteria, we believe that an "interviewer effect" has muddied some of our results. Data collected showed significant differences between our two interviewers in nearly all dimensions of data quality (not shown here); John had longer duration and lower GPS capture on tablets than Jane, and Jane had fewer missing values on tablets and more valid phone numbers than John. From the literature we know that differences in missing values may arise from respondent characteristics, such as an unwillingness to provide information for one device due to privacy concerns or systematic difference in the sample assigned to one interviewer, or from differences in interviewer behaviour, such as lower rates of probing or other causes (see de Leeuw, 2001). Our research is unable to uncover respondent reluctance associated with device, but a thorough examination of the pilot interviewers' cases revealed that differences in sample characteristics did not explain between-interviewer differences on quality measures (not shown). Instead, it is possible that we are detecting differences in quality that are associated with the capabilities or experience of the two interviewers in the pilot rather than differences attributable to their interactions with the two different devices. Of particular note in this regard is the much shorter duration of interviews by Jane on both devices, despite no differences in her sample compared to John's that would lead to shorter interviews.

The unexpected admission of poor adherence to survey administration protocols (using memory instead of scrolling for long questions or response options) and the problem of mis-selecting responses while scrolling suggest several recommendations for programming and interviewer monitoring when using tablets or smartphones for data collection. First, surveys must be optimized for the screen size of the data collection device, including cutting lists into segments that fit onscreen. Second, programmers must take care to place selection buttons in the centre of the screen, away from the edges of the form, where users place fingers for scrolling or paging. Third, programmers should weigh the benefits of programming confirmation screens against the cost of lengthier surveys. Inconsistencies between initial response and confirmation screen should produce a flag immediately visible to the interviewer to allow for correction during the interview. Fourth, interviewer training should include demonstration and practice on correct use of the touchscreen to avoid mis-selections. Finally, interviewer monitoring should include, if feasible, recording portions of the interviewers' survey administration.

While implementing these recommendations could reduce interviewer-produced errors, ultimately the

quality of survey data largely depends on the technical skills of the interviewers and the investment in training, data review, and continuous interviewer feedback made by the research team.

### Limitations of the analysis

This research has a number of limitations, briefly listed below. Budget constraints were the major driver of our choice of a non-experimental method, while client reluctance to extend the pilot to a larger portion of the total survey sample was another design consideration. Thus, readers should keep in mind that the research is limited by:

- Non-experimental method
- Small sample (2interviewers, 2 cities, 100 respondents per city, 50 in each arm)
- No independent verification of the response data (call-back data verification with respondents was carried out by the interviewer team supervisor, not by independent researchers and did not include full re-interviews)
- Unknown influence of interviewer effects and interviewer interaction with the devices
- May not be generalizable to other contexts or survey content
- Survey was used “out of the box” and not optimized for use on phone

### Conclusion

Adoption of new, faster, and cheaper devices for data collection is tempting on any survey project, perhaps particularly so in developing countries where alternatives are few and data collection budgets are low. However, researchers should incorporate methods for identifying in advance the ideal screen size and functionality of data collection device depending on the content and length of questionnaires, as well as other relevant requirements of the survey project. In software and systems engineering, analysts define “use-cases” appropriate for different purposes. Our research suggests that there could be some use-cases for which tablets are most appropriate, others in which phones are best, and still others in which phones and tablets are interchangeable. In addition, further study is needed to better understand how human-computer interaction affects data quality on CAPI studies that adopt mobile devices. Researchers must focus efforts on reducing errors that could be tied to device size and screen layout when selecting a device, and to modify hiring, training, and monitoring of interviewers to take into account different interviewer experience and interviewing styles.

# References


1. Armand, J., Redick, T., & Poulsen, J. (2013). Task-Specific Performance Effects with Different Numeric Keypad Layouts. *Applied Ergonomics, 45*(4), 917-922.Banks, R., & Laurie, H. (2000). From PAPI to CAPI: The Case of the British Household Panel Survey. *Social Science Computer Review*, 18*(4)*, 397-406.
2. Banks, R., & Laurie, H. (2000). From PAPI to CAPI—the case of the British household panel survey. *Social Science Computer Review, 18*(4), 397-406.
3. Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley & Sons.

4. Bruijne, M., & Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment with a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review, 3*(4), 482-504.
5. Caeyers, B., Chalmers, N., & De Weerdt, J. (2010). *A Comparison of CAPI and PAPI Through a Randomized Field Experiment*. Social Science Research Network. Retrieved from http://ssrn.com/abstract=1756224.
6. Carletto, C., Jolliffe, D., & Banerjee, R. (2015). From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *Journal of Development Studies*, *51*(2), 133-148.
7. Couper, M. P., & Kreuter, F. (2013). Using Paradata to Explore Item-Level Response Time in Surveys. *Journal of the Royal Statistical Society, 176*(1), 271-286.
8. de Leeuw, E. D. (2001). Reducing Missing Data in Surveys: An Overview of Methods. *Quality & Quantity*, *35*(2), 147.
9. de Leeuw, E. D. (2008). Choosing the Method of Data Collection. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 113-135). New York: Taylor & Francis Group/Lawrence Erlbaum Associates.
10. de Leeuw, E. D., & Huisman, M. (2003). Prevention and Treatment of Item Nonresponse. Journal of Official Statistics, 19(2), 153.
11. Escobal, J., & Benites, S. (2013). PDAs in Socio-Economic Surveys: Instrumental Bias, Surveyor Bias, or Both? *International Journal of Social Research Methodology, 16*(1), 47-63.
12. Fowler, F. J. (2004). Reducing Interviewer-Related Error Through Interviewer Training, Supervision, and Other Means. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 259-278). Hoboken, New Jersey: John Wiley & Sons, Inc.
13. Groves, R. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.
14. Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires. *Public Opinion Quarterly*, *67*(1), 79-125.
15. Jans, M., Sirkis, R., & Morgan, D. (2013). Managing Data Quality Indicators with Paradata-Based Statistical Quality Control Tools: The Keys to Survey Performance. In F. Kreuter (Ed.), *Improving Surveys with Paradata* (pp. 191-229). Hoboken, New Jersey: John Wiley & Sons, Inc.
16. Kosyakova, Y., Skopek, J., & Eckman, S. (2014). Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach. *International Journal of Public Opinion Research, 27*(3), 417-431.
17. Lugtig, P., & Toepoel, V. (2016). The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, *34*(1), 78-94.
18. Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review, 31*(6), 725-743.
19. Mavletova, A., & Couper, M. P. (2014). Mobile web survey design: Scrolling versus paging, SMS versus e-mail invitations. *Journal of Survey Statistics and Methodology*, 2, 498–518.
20. Olson, K., & Peytchev, A. (2007). Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes. *Public Opinion Quarterly, 71*(2), 273-286.
21. Park, Y. S., & Han, S. (2010). One-Handed Thumb Interaction of Mobile Devices from the Input Accuracy Perspective. *International Journal of Industrial Ergonomics, 40*(6), 746-756.
22. Peytchev, A., & Hill, C. A. (2010). Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations. *Social Science Computer Review, 28*(3), 319-335.
23. Redline, C., & Dillman, D. (1999). *The Influence of Auxiliary, Symbolic, Numeric, and Verbal Languages on Navigational Compliance in Self-Administered Questionnaires*. U.S. Census Bureau. Retrieved from https://www.sesrc.wsu.edu/dillman/papers/1999/theinfluenceofauxiliary.pdf.
24. Schrapler, J., Schupp, J., & Wagner, G. (2010). Changing from PAPI to CAPI: Introducing CAPI in a Longitudinal Study. *Journal of Official Statistics, 26*(2), 233-269.
25. Siekmans, K., Ngnié-Teta, I., Ndiaye, B., & Berti, P. (2012). Experience with Digital Entry of National Iodine Survey Data in Senegal. *African Journal of Food, Agriculture, Nutrition & Development, 12*(7),

6987-7000.
26. Tomlinsen, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., ... & 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, 9*(51), 1-8.
27. Trott, D. L., & Simpson, A. M. (2005). Computer-Assisted Personal Interviewing—The Bermuda Experience. *Statistical Journal of the UN Economic Commission for Europe, 22*(2), 133-145.
28. United Nations Statistics Division. (2014). *Guidelines for Producing Statistics on Violence Against Women—Statistical Surveys*. Retrieved from http://unstats.un.org/unsd/gender/docs/Guidelines_Statistics_VAW.pdf.
29. Yu, P., de Courten, M., Pan, E., Galea, G., & Pryor, J. (2009). The Development and Evaluation of a PDA-Based Method for Public Health Surveillance Data Collection in Developing Countries. *International Journal of Medical Informatics, 78*(8), 532-542.