Implementing a CAPI Survey with *Survey Solutions*

The Decision to Use CAPI

The decision to use computer-assisted personal interviewing (CAPI) instead of paper and pencil (PAPI) is an important one and should not be taken lightly. It is a step into modern technology that may be more expensive for the first one or two surveys, but subsequent surveys will be far less expensive, will provide much higher data quality, and the data will be available for analysis much faster than when using PAPI. The staff members who implement a CAPI survey will gain experience in a technology that is rapidly becoming the standard throughout the world of data collection, so your entire organization will be more competitive.

The decision to use CAPI should go hand-in-hand with the choice of data entry and management software that is used. Many factors should be considered when making these decisions:

- Should CAPI be considered at all? For example, will a paper trail be required, or will a statistical database be sufficient? Do local laws and regulations permit the electronic storage and transmission of survey data?
- Will the fieldwork require constant or frequent monitoring? For some software this would suggest the need for widely available cell phone or Internet service.
- Will the interviews be conducted in multiple languages?\(^1\)
- Does the questionnaire include more than one respondent?
- Are you willing and able to pay the initial costs of tablets and training for the long-term benefits of CAPI?
- Will the listing be conducted prior to or during the interviewing process?

The answers to these questions will help to guide the decision whether to use CAPI, and if so, which software to use. Every CAPI software package has strengths and weaknesses, and they should be taken into account during this decision-making process. For a comparison of the different CAPI packages (except *Survey Solutions*) see [this World Bank-funded report from 2011](#).

Direct Costs

One component of CAPI costs is for the tablets on which the data are collected. The most popular of these devices, and at present the least expensive, run on the Android operating system. Each interviewer must have her or his own tablet computer for data collection, so this can represent a substantial initial investment. Fortunately these devices can be reused on future surveys, so after this initial investment the only costs will be for replacements of lost or damaged units.

---

\(^1\) All languages supported by Microsoft Windows are available in *Survey Solutions*.
Cost should not be the only consideration when selecting the appropriate tablet for data collection. The operating system should be the first consideration, since most data entry packages run on either Windows or Android tablets, although some run on both. These considerations include battery life, charging time, screen size and resolution, connectivity options (such as cell phone, Bluetooth, USB, wifi, and connectivity between tablets), and durability under harsh environmental conditions (such as hot and cold temperatures, rain, and sand storms). Note that prices of both Windows and Android tablets have been dropping dramatically over the past several years and there is every reason to believe they will continue to do so.

Another cost of CAPI surveys is for training of interviewers. Interviewer training is also required for PAPI surveys, but CAPI interviewers must also be trained on the use of the tablets. Field supervisors and headquarters staff will also need to be trained on the use of these devices, as well as the differences in the way CAPI surveys are managed. These differences typically result in a more streamlined process which provides field teams with faster, more targeted feedback than with PAPI surveys, and that provides higher data quality.

Another cost relates to communications and connectivity. Depending on availability, the data and quality control information must be exchanged between the field teams and headquarters using either cell phone or Internet connectivity. Whichever is used, additional user accounts will likely be required.

Most CAPI systems now rely on case and data management systems that reside on local, remote, or cloud-based servers. If the server is local this will include the cost of the server hardware and software as well as server maintenance. If the server is hosted remotely or in the cloud, monthly costs may amount to US$150 per server. Each server may host multiple surveys.

**Indirect Costs**

Technical support is the major cost to consider when deciding whether to switch to CAPI. Depending on the CAPI system these costs can total several hundred thousand US dollars and might completely overshadow the direct costs of CAPI equipment and training.

Survey Solutions offers many options for configuring servers. These options are described [here](#).

Survey Solutions is designed to minimize indirect costs. Still, there is likely to be the need for some technical assistance for the first one or two surveys. The extent of this TA will depend on the specific requirements in each country and will be determined after discussions with the Survey Solutions team.
Benefits
There are many benefits to using CAPI to replace PAPI data collection. First and foremost is the improvement of data quality. When data are rapidly available for assessment after they are collected, errors in fieldwork can be corrected and changes in procedures, if needed, can be implemented quickly.

Another benefit is the elimination of paper questionnaires, reducing printing costs and reducing the burden on field teams to ensure they have sufficient numbers of blank forms. Transmission of the results of the interviews is done electronically so a package delivery network is no longer needed.

The day that fieldwork is completed the data are immediately available for review and cleaning without the need to finish entering data from the remaining paper forms. This means that analysis can begin sooner and results will be available earlier.

CAPI Scheduling
A typical CAPI survey requires at least 3 months for fieldwork preparation. This is somewhat longer than for PAPI surveys, due to the need for tablet procurement for initial surveys, and the need for additional staff training. However, this schedule should be reduced after the first or second CAPI survey since most of the tablets will still be available and most of the staff will already be trained. Also, the data will be available for analysis earlier than they would be for a PAPI survey.

An example of a typical CAPI survey is shown in Figure 1. While the schedule of a specific survey may be somewhat different from this example, it will not vary substantially based on the size of the sample. The length of fieldwork may vary considerably based on the ratio of size of the fieldwork team to the size of the sample, and whether listing efforts will take place before or during fieldwork. The example in Figure 1 does not include listing and assumes it is undertaken as a separate task.

1. **Obtain coverage maps for Internet access and from cellular providers:**
   If you are planning to transmit data and monitor fieldwork directly from the data entry tablets, than widely available cellular and/or Internet coverage is essential. On the other hand, if frequent data transmission is not critical, or if fieldwork monitoring is available in the field, the need for widespread connectivity can be somewhat relaxed.

2. **Select sample PSUs and compare with coverage maps:** Once the sample PSUs are selected, coverage maps will help determine the best provider and routing of interviewer teams to optimize their opportunities to transfer data. If necessary use a combination of providers to increase coverage of selected PSUs or develop field work routines to cope with uncovered locations.
3. **Decision to use CAPI**: This should be done in consultation with all stakeholders and with a clear understanding of this new technology. A major part of this decision includes the data network to use, whether it uses the Internet or cell phone infrastructure, or a combination of both. A comparison should be made between the selected primary sampling units (PSUs) and the availability of cell phone and/or Internet connectivity along with the anticipated frequency of remote fieldwork monitoring.

4. **Android tablet procurement**: Procurement of the tablets should be started at an early stage. Check whether they are available in the country, and reconfirm their compatibility with the selected data entry software.

5. **Questionnaire implementation on tablets**: Ideally training on the process of loading your questionnaire onto the interviewers’ tablets should only be required the first time the software is used.

   Some data entry software requires the use of programming languages, and provides extensive flexibility and power for large, complex questionnaires. These packages generally require substantial training of technical staff or the extensive reliance on consulting assistance. Other software packages require little or no programming but are appropriate only for use with simple questionnaires and sample designs. Therefore the selection of survey processing software should take into account the types of surveys that are expected in the future.

6. **In-house questionnaire testing**: An initial test of the tablet data entry system should be conducted within your office, ideally by staff members who are familiar with the questionnaire and subject matter. This will help to resolve the most obvious errors prior to the field pre-test.

7. **Supervisor and Interviewer training**: Supervisors should participate in the same training as interviewers to provide them with a clear understanding of the questionnaire and the use of the tablets.

8. **Field pre-test**: This is the first opportunity to test the tablet-based data entry system under realistic conditions. Locations should be identified that are not included in the survey sample, and interviewers and supervisors should conduct realistic interviews with real households. This provides both a training experience for the field staff and a test of the data entry system.

9. **Modifications to questionnaire based on pre-test**: Based on observations from the pre-test, problems with field procedures and the tablet questionnaire must be resolved. If necessary, changes should be made to the data entry system on the tablets. The result of this task is the final version of the data entry system.
10. **Pilot test**: After the questionnaire is finalized in the previous task, the pilot will test the final questionnaire as it is implemented on the tablets, and will also focus on connectivity and communications. This will test your network infrastructure and ensure that interviewers and supervisors are familiar with data transfer procedures. This will ensure that all of the changes based on the first pre-test have been properly implemented and that the data can be transferred when needed. It will also ensure that all supervisors are able to provide the necessary support to their teams in the event of a technical failure in the field.

11. **Supervisor training**: This training should clearly define the role of the field supervisors, and enable them to use the data and case management tools they will be using in the field. This will enable the supervisors to take advantage of CAPI by checking the quality and completeness of the data as they are collected. In most data collection software, this is done by visually checking the data on the tablets or using separate management software.

12. **Fieldwork**: When considering the trade-off between a larger field team and an extended period of fieldwork, it is generally preferable to have a longer period of fieldwork. This provides the survey team with more time to learn about and respond to problems in the field.

13. **Training on data export, import, and analysis**: This training should only be required the first time *Survey Solutions* is used. *Survey Solutions* provides a straightforward method for exporting its data to tab-delimited files. These files are organized as a relational database, with the ability to easily import and merge them into files that can be used for statistical analysis. We provide the training to export, import, and analyse the *Survey Solutions* data.

14. **Data cleaning and analysis**: This is the final step in the primary use of the survey data. With *Survey Solutions*, data cleaning should be minimal since the data entry and built-in case management systems check for most of the fieldwork problems encountered with other data entry systems.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Obtain coverage maps for Internet access and from cellular providers</td>
<td>7d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td></td>
<td>[ ]</td>
</tr>
<tr>
<td>2</td>
<td>Select sample PSUs and compare with coverage maps</td>
<td>10d</td>
<td>[ ]</td>
<td></td>
<td></td>
<td></td>
<td>[ ]</td>
</tr>
<tr>
<td>3</td>
<td>Decision to use CAPI</td>
<td>3d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Android tablet procurement</td>
<td>15d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td></td>
<td>[ ]</td>
</tr>
<tr>
<td>5</td>
<td>Questionnaire implementation on tablets</td>
<td>14d</td>
<td></td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>6</td>
<td>In-house questionnaire testing</td>
<td>3d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Supervisor and Interviewer training</td>
<td>6d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Field pre-test</td>
<td>6d</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>9</td>
<td>Modifications to questionnaire based on pre-test</td>
<td>6d</td>
<td>[ ]</td>
<td></td>
<td></td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Pilot test</td>
<td>6d</td>
<td></td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td>[ ]</td>
</tr>
<tr>
<td>11</td>
<td>Field supervisor training</td>
<td>6d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Fieldwork</td>
<td>30d</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Training on data export, import, and analysis</td>
<td>4d</td>
<td>[ ]</td>
<td>[ ]</td>
<td></td>
<td></td>
<td>[ ]</td>
</tr>
<tr>
<td>14</td>
<td>Data cleaning and analysis</td>
<td>14d</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

Figure 1: CAPI Survey Schedule Example