

American Institutes for Research

Academy for Educational Development

Aga Khan Foundation

CARE

Discovery Channel Global Education Fund

Education Development Center

Howard University

International Reading Association

The Joseph P. Kennedy, Jr. Foundation

Juárez and Associates, Inc.

Michigan State University

Sesame Workshop

Save the Children Federation, USA

University of Pittsburgh

World Education



USAID
FROM THE AMERICAN PEOPLE



Digital Data Collection Demonstration White Paper

A Comparison of Two Methodologies: Digital and Paper-based



Produced by:
American Institutes for Research
under the EQUIP1 LWA

Prepared by:
Dr. Roy Zimmermann
American Institutes for Research
rzimmermann@air.org

November 2008

U.S. Agency for International Development
Cooperative Agreement No. GDG-A-00-03-00006-00

This study was made possible through coordination and support from
UNICEF, USAID and Global Relief Technologies

In an effort to explore various options for data collection at the field level, the American Institutes for Research (AIR), with funding from the U.S. Agency for International Development (USAID) through the Educational Quality Improvement Program 1 (EQUIP1), launched a comparative study of paper versus digital education data collection methodologies. This study ran parallel to an existing evaluation that AIR was conducting on behalf of UNICEF to evaluate the Child Friendly Schools (CFS) model. The study was conducted in Nicaragua and used a deployment of Global Relief Technologies' (GRT) digital data collection approach to compare this data collection methodology with the paper-based approach that was being used on the CFS evaluation. The GRT model used hand-held personal digital assistants (PDAs) for capturing observable, quantifiable data, a portable scanner and portable satellite transmitter for capturing and communicating data.

The comparison considered process efficiencies, implementation flexibility, and cost. This pilot was a separate exercise from the UNICEF CFS evaluation study and was deliberately designed not to interfere with the actual data collection in any way. This paper provides the findings of the comparison of the two processes of collecting data as experienced in Nicaragua: digital and paper-based.

BACKGROUND

In January of 2008, the American Institutes for Research (AIR) was awarded a contract from UNICEF to conduct a global evaluation of the UNICEF Child Friendly Schools (CFS) program. AIR's evaluation was structured to sample two countries in each of the three major regions that CFS is being implemented: Africa, Asia, and Latin America/Caribbean. The six country case studies were conducted in Nigeria, South Africa, Indonesia, Thailand, Guyana, and Nicaragua. In



Students at school in Nicaragua

each country, AIR collected data from 25 schools and interviewed staff at the MOE, UNICEF, and advocacy groups.

The overarching goal of the evaluation was to determine the impact the CFS model is having in schools around the world. The CFS evaluation was to determine effectiveness of the CFS model and produce recommendations for future CFS investments. The AIR approach to the evaluation utilized both primary and

secondary data sources and focused on the six key elements of CFS: Architecture and Services, Pedagogy, Inclusiveness, Participation and Governance, Systemic Management, and Cost.

TWO PROCESSES

AIR developed and prepared 14 Master Research Instruments ranging from student surveys to school cost data forms to parent focus group protocols. Each teacher- and student-directed instrument was sent to the local country UNICEF office where it was first adapted to the local lexicon and then reviewed and approved by the AIR home office. When necessary, instruments were returned to UNICEF country offices for translation, while others were administered in English.

Two AIR research specialists were deployed to each country where they trained and worked with 4 to 8 locally hired, experienced data collectors. Qualitative and quantitative data were gathered from policy makers, school directors, teachers, parents, students and community organizations. Data collectors for the CFS evaluation used paper-based surveys and recorded interviews and focus groups using digital voice recorders, which allowed data collectors to transcribe each session. Photos and videos were captured using digital cameras to provide visual media to illustrate findings and evaluation activities. In Nicaragua, AIR deployed additional data collectors to shadow those conducting the CFS evaluation. The AIR data collectors used GRT PDAs to collect the same data from the same groups. Below the two processes are described in more detail.

Paper-based Methodology

After being translated, all instruments were photocopied in-country at least 4 days prior to deployment. On average, 20,000 photocopies were produced in each of the six targeted countries which also required 3,000 sharpened pencils. Upon arrival in country, the first task was to verify that all copies had been made and that print quality was acceptable. In a few instances additional copies were needed along the way. Next, all photocopies were collated so that the correct number of instruments and copies was set aside for each sample school.

Each locally hired data collector was trained in each instrument, reviewing each question together with the AIR research specialists for common



Students completing paper based survey

understanding of terms and the goals of each question. On occasion, adjustments needed to be made to the terminology used. For example, it was determined in one country that the term “minority group” used on an interview protocol was not an appropriate term for the context and it was changed to “ethnic group.” Local data collectors took notes of these changes and were instructed to incorporate these changes when administering the instruments, however at this point it was too late to change each printed instrument. In addition, local data collectors were required to capture visual media to further document the evaluation and were dispatched with digital cameras, digital video cameras and digital audio recorders for focus groups and interviews. During the data collection process, local data collectors sometimes wrote notes, generally in their mother tongue, in margins of some of the instruments, to document unexpected observations or situations.

The large volume of paper required, in several instances, additional suitcases or water proof containers be purchased in country to allow for local transport and security. Completed surveys were either shipped via Federal Express or hand carried back to the AIR home office in DC. Once received by AIR/DC, they were manually catalogued through an intake process that organized hardcopies, electronic copies, and digital media into a tracking log. After this step was completed, hard copies of the surveys were shipped to a scanning service in Chicago where they were scanned into a digital format and sent back to AIR in electronic data sets for analysis.

Digital methodology

GRT required receipt of final instruments one week prior to deployment so that each instrument could be formatted into a digital format onto PDAs. The instruments were digitized in a format that exactly mimicked the paper based version, asking identical questions in identical order with identical response options (whenever multiple choice applied). Once in country, any changes required of the instruments were made to the digital format “on the fly” and all instruments were remotely updated to reflect the revisions. This process took, on average, one hour to complete.



**The PDA running
GRT’s Rapid Data
Management System**

When visiting schools, the data collectors using the PDAs shadowed local data collectors deploying the paper-based instruments and sought to capture the same information. At the end of each day, all data captured through the PDAs was uploaded using a portable satellite transmitter to GRT’s Virtual Network Operations Center (VNOC) that automatically catalogued and organized data into a database by school and instrument. Additionally, student surveys that had been captured via paper were scanned into a dual-sided portable scanner and digitized. The approximately 100 surveys from each school required 45 minutes per school to scan. Once the student surveys collected that day had been scanned they were also uploaded to the VNOC and

automatically catalogued and organized into the database.

The PDAs came equipped with a bundle of other relevant features including GPS, cameras, audio/video recorders and cellular data/voice capability. The PDA camera and audio recorders were used to capture the same information that CFS evaluators were capturing with their digital cameras and voice recorders. Additionally, the GPS feature allowed each school to be plotted on a map. This was information that was not available to the data collection team or UNICEF prior to the evaluation. This feature, while not required for the CFS evaluation, enabled a home office to verify that data was actually collected in various sites, providing confirmation that data collectors were not derelict in their duties, for instance, sitting in one spot inputting artificial information.

COMPARISONS

The section below highlights the main differences that AIR encountered when conducting its paper- and digital-based data collection methodologies in Nicaragua.

Flexibility

Flexibility is extremely important when conducting such a large scale evaluation. Because of the low availability of resources in the countries that AIR studied, coupled with the sheer number of instruments and respondents, it was often burdensome to make changes at the last minute. Both the paper and digital methodologies are flexible in different ways. The paper-based method allows flexibility for data collectors to quickly jot down notes (in margins, on flip side of paper, or in separate data logs) that can be referred to at a later time. However, because photocopies of the instruments had to be made in advance, the team never knew exactly how many copies were required at each school since exact student numbers were not known in advance of site visits. In order to circumvent any major issues, AIR always photocopied the maximum number of respondents per school, even though this number of students was not always present. This occasionally led to an inefficient use of resources.

The digital approach to the data collection provided a level of flexibility not possible with the hard copy approach. For example, revisions to any and all instruments could be made on the fly within an hour or less. While capturing text was difficult with the PDA format, any notes collected were typed, removing the difficulties that would normally occur due to illegible handwriting.

Ease of data entry and collection

With such a complex logistical process in place, it is also extremely beneficial for data collection methodologies to be appropriate for the environment in which they take place. Easing the data entry and collection processes can reduce the burden on data collectors and increase the accuracy of the data collected. In order for hard copies to be “scannable” they needed to be relatively clean and wrinkle-free, as well as legibly completed on 60 lb paper for effective scanning. This was

Assessment Details: AIR – Nicaragua / Child Friendly Schools / Parent Interview Protocol

The screenshot displays the VNOC assessment details for 'Parent Interview Protocol' in Nicaragua. The page is divided into several sections:

- Assessment Information:** A table listing key details such as Staff (Device), School, Subject, City, Lat, Lon, mgrs, Assessed, Posted, Last Reviewed, Last Edited, and Last Shared.
- Assessment Details:** A section for 'Tracking Information' with fields for Country ID, School ID, and Locality.
- Category: CFS:** A section for 'What does CFS' with a link to 'Its a school unere (hildt Art)'. Below this is a photo of a group of people sitting around a table.
- Audio Player:** A video player interface showing a green bar chart and a play button. A green arrow points from the 'Click here to listen' link to the audio player.

An interview assessment in the VNOC tagged with photo and audio attachments

sometimes a challenge, particularly when last minute additions or revised copies had to be made in the field. The matter was complicated by the amount of traveling the data collectors did – transporting 20,000 hardcopies of instruments over long distances made it difficult to ensure they were kept scannable. Handwritten notes provided by local data collectors or respondents on the data collection instruments were sometimes difficult to decipher and in a few instances questions went unanswered. Any changes made to instrument questions during the training period after they had already been photocopied had to be recorded and remembered by local data collectors over the course of the two week site visit.

The digital data collection model required much less paper, but did not remove the need for paper completely. The paper that was used (e.g. surveys of students, teachers and principals) was scanned and uploaded on a nightly basis and therefore easier to keep clean than bundles of paper that were transported around a country for up to two weeks before being shipped or transported to the U.S. Furthermore, digital questionnaires can be set up to require an answer to each question before allowing the collector to ask subsequent questions, avoiding unanswered fields. Media collected, including photos and audio were automatically tagged to specific correlating questions and/or instruments as part of the automatic cataloguing system. While the GRT system was clearly more conducive to quantifiable, close-ended questions than supporting interviews or focus groups. Multiple choice or number formats worked best, for example capturing the number of students in a classroom or observing conditions of school campus. The number of respondents

was limited by the number of PDAs. Surveys of large numbers of respondents were not conducive to the PDA format and, at least in the current iteration, would still need to be captured via hard copy, which were scanned/digitized in the field and uploaded on a daily basis. There is also a comfort and learning curve associated with working with the PDAs, though that seemed to be overcome within an hour or two of use.

Materials

Along with the paper instruments, 135 sharpened pencils were required for each school, 3,000 per country. Collecting these back for re-use quickly became a non-starter as students and teachers in some locations came to covet the pencils. In addition to the papers and pencils, data collectors carried digital cameras, video cameras and digital voice recorders. The PDAs combined six devices in one: GPS for tagging exact locations of schools, audio recording, cellular, video and digital photo capabilities as well as automatic intake cataloguing of media files as they are created.

Time

Once received, intake time for the paper-based data before documents could then be shipped out to the scanning service was consuming. Ultimately, the entire process from data collection to digitization took approximately four weeks, before analysis could take place. The digital format allowed for near real-time reporting. As data was transferred daily into the VNOC it could be viewed by anyone with internet access and appropriate privileges for quality control or to see early data patterns. It would be possible to red flag questions or even data collectors themselves that were not delivering desired inputs. In this regard, adjustments to specific questions or elements of the research methodology could be made while still in the field. The nearly instantaneous digitization as well as the top-level analysis and reporting provided through the database system greatly reduced the overall duration of the evaluation.

Accuracy

While this demonstration was not designed to measure the accuracy of either the digital or paper-based methodology, a sample of schools and student surveys indicated 98.86% accuracy when comparing the digital data to the paper-based data.

Transportation & Security

Having findings recorded on paper provides a sense of security that data is in hand and can be referred to on an as needed basis. However, issues can emerge around transport of 20,000 sheets of paper from a central location to schools often in remote places only accessible via four-wheel drive vehicles or, in the case of Guyana, via boat. Keeping the documents clean and dry was another requirement that data collectors were tasked with. Shipping of completed instruments back to the home office on an on-going basis during the field-work period was not always

possible given the remoteness of some of the targeted schools and, in some cases, when it was possible, it was very costly.

Another concern, though not one that was an issue in this activity, was that of data security and vulnerability. If the completed forms went missing either during data collection or shipping back to the home office, there was no form of back up. Data was secured in digital format upon transmittal so there was no need to worry about data lost during transport or shipping. It was also backed up through a redundant server security program through GRT's own security systems. Data on the VNOC was password protected.

Cost

This comparison exercise sought to look at two processes for comparing data on a side by side basis. It was not set up to do a detailed cost comparison. However, there are some conclusions that can be drawn about the cost. The majority of cost in the paper-based approach came from labor categories, primarily in the analysis and what happens with the data after it is collected in the field. This process on average took about four weeks before data in an analyzable format was received at the home office. The primary cost associated with the digital format was in renting the equipment and services of GRT during the two week site visit. Moreover, the paper based approach required additional equipment to be purchased such as cameras and voice recorders that were a part of the PDA platform. An initial and high level look at both costs suggest that, at the very least, the savings in labor realized through the digital approach make up for the added costs for equipment and services to complete the project, and at best even have the potential to reduce overall costs as well as provide the other benefits described above around data security and implementation/process efficiencies. Furthermore, paper and digital materials used in the study represent a relatively very small percentage of the overall project budget. The materials costs for the paper version for Nicaragua represent less than one percent of the total budget and the costs for the digital version represent just over 1 percent. Much of the costs associate with digital data collection are attributable to equipment. If equipment is purchased and those costs amortized over several projects then costs come down even further. This makes the cost/benefit ratio of digital very attractive compared to that of paper, particularly in light of lower vulnerability to loss which would be at significant cost.

CONCLUSIONS

The paper-based approach is the methodology that has been used for the last several decades for international evaluations – especially those that take place in developing countries. It is familiar to both data collectors and designers of such evaluations. However, as the international development field moves forward and information and communication technologies become cheaper and more readily available, it is important to look to the future and experiment with methods that while unfamiliar and require a learning curve, may ultimately prove to be a time-efficient and cost-effective methodology.

Further research would help to validate these conclusions and expand understanding around the possible efficiencies to be realized in using more digital data collection in school and education related studies. For instance, student surveys which were essentially a hybrid of paper (for collecting data) and digital (for analysis) might be reconsidered in a digital only format to skip the inefficiencies and vulnerabilities in using paper. Another opportunity for further research would be to continue to validate the accuracy of the digital methodology when compared to paper-based data and to determine which has a higher accuracy rate. In this comparison of digital versus paper-based data collection methodologies, AIR hopes to generate further discussion around the benefits and obstacles related to using digital data collection methodologies in large scale global evaluations.