

American Institutes for Research

Academy for Educational Development

Aga Khan Foundation

CARE

*Discovery Channel Global Education
Fund*

Education Development Center

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



***Pilot of Digital Data Collection Options for
Honduras' Education Assessments***



Produced by:

**Mauricio Estrada and Roy Zimmermann
with Jeffrey Lansdale
American Institutes for Research
under the EQUIP1 LWA**

December 2009

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

***PILOT OF DIGITAL DATA COLLECTION OPTIONS FOR
HONDURAS' EDUCATION ASSESSMENTS***

by
Mauricio Estrada and Roy Zimmermann
with Jeffrey Lansdale

Mejorando el Impacto al Desempeño Estudiantil de Honduras /
American Institutes for Research
(MIDEH/AIR)

DECEMBER 2009

TABLE OF CONTENTS

Abstract	1
Contextual Summary	1
Introduction.....	2
Pilot Rationale	3
Pilot study of the ResponseCard XR	5
Key Findings.....	6
Pilot study of the Earthwalk mobile computer labs	6
Key Findings.....	8
Comparing ResponseCard XR vs. Earthwalk mobile labs	8
Conclusions and Recommendations	10

ABSTRACT

In an attempt to identify new, cost effective technologies for data collection that can be applied in the Honduran context to improve sustainability of the national assessment process, the present study evaluates two alternative technologies to collect student responses when it is not possible to use answer sheets in a standardized testing system (e.g. students are too young). The technologies tested are Turning Technologies ResponseCards XR, and Earthwalk mobile computer labs. The purpose of the study is to determine the reliability of the methods, the speed of data entry and the user interaction with both systems. For each technology, a 5-day pilot test was conducted; each pilot test was attended by 16 participants. No significant differences were found regarding data integrity. The major differences are data entry speed, in which Earthwalk mobile labs hold the advantage; and the cost of each system, in which the ResponseCards XR is the most cost effective solution. However, other considerations such as transportation, power, safety and technical skills were also taken into account.

CONTEXTUAL SUMMARY

In Honduras, the American Institutes for Research (AIR) with the Honduran Ministry of Education and other partners has implemented a series of projects called *Mejorando el Impacto al Desempeño Estudiantil de Honduras* (MIDEH—Honduras Improving Student Achievement Project). Funded by the US Agency for International Development (USAID), the projects are designed to build a national standards-based reform program, and since 2004, have been able to create a sustainable educational reform structure. The nationwide approach is based on the development of internationally-comparable academic standards, teacher pacing guides for implementing the standards, standardized student assessments (both formative and summative), and in-service training programs.

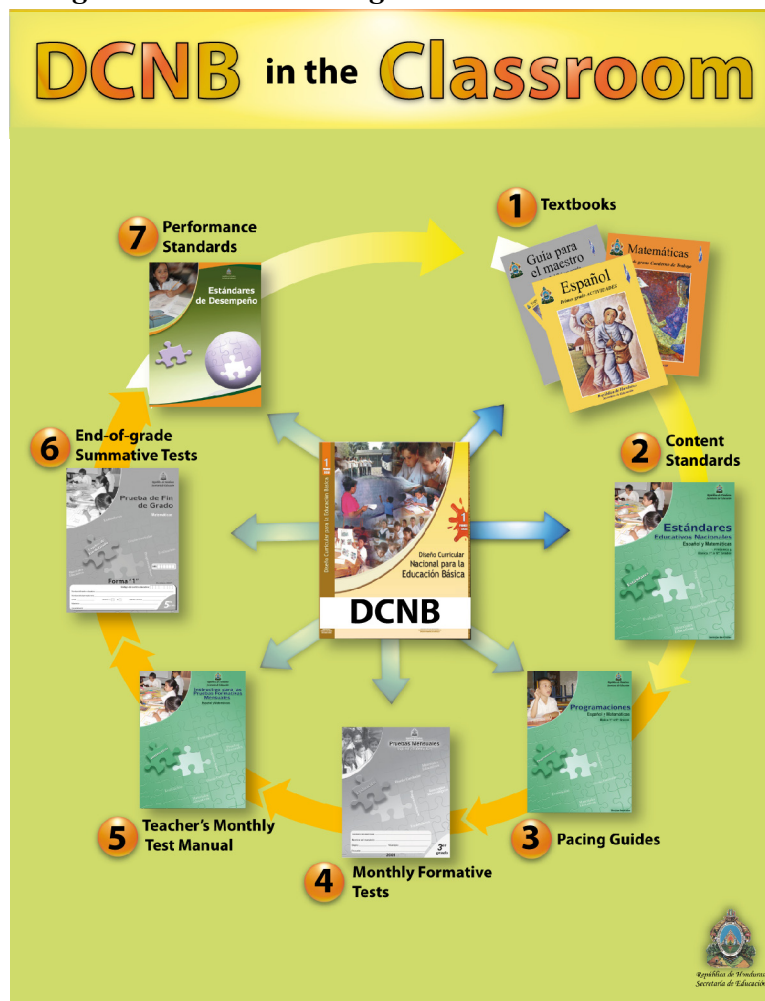
Perhaps the most critical aspects of the education reform structure are the pacing guides and formative assessments, where teachers in grades one through eleven have classroom instructional materials organized by sequenced standards, along with monthly assessments for each of their students. Each student receives individual booklets that have been distributed to all schools. The various components of the system, including textbooks funded by the government and other donors, are standards-based and mutually reinforcing.

A well-known international education expert, Dr. Richard Kraft, visited Honduras in 2009 and stated that he was “delighted to run across a project that appears to have all the components of what I've been preaching for these many years in countries around the world.” Results from tests over the past two years have shown significant gains in two current subject areas, mathematics and Spanish language. Next steps for Honduras will be to expand into more subject areas and improve sustainability through the establishment of a national institute for student assessment.

INTRODUCTION

In 2002, the Honduran Ministry of Education launched a new curriculum for grades K-9, called the DCNB, (Diseño Curricular Nacional Básico). To support the implementation of the new curriculum in the classroom throughout the country, the USAID funded MIDEH/AIR Project (Mejorando el Impacto al Desempeño Estudiantil de Honduras / American Institutes for Research), in close coordination with the Ministry of Education and education leaders from every department of the country, developed a national testing system based on content standards in Math and Spanish language. Materials developed by MIDEH/AIR include: content standards that define what should be taught and learned in grades K-11 in Math and Spanish language; pacing guides that organize the standards to be taught and mastered month-by-month; monthly formative tests aligned to the standards, with a teacher's manual describing how to administer and score the tests and interpret the results; and a software package that assists teachers and school directors to record formative test scores on a monthly basis. These materials are all aligned to the new curriculum and the new official Ministry textbooks for Math and Spanish (See Figure 1).

Figure 1: DCNB and Aligned Educational Materials



MIDEH/AIR developed end-of-grade summative tests, which were administered to a national sample of approximately 400 schools in 2007, grades 1-6. In 2008, the end-of-grade tests were administered to a national and departmental sample of 900 schools, reaching 105,000 students in grades 1-9, as well as to a sample of students enrolled in alternative education programs. Using the test results and input of teachers from across the country with subject area and grade level specialty, MIDEH/AIR facilitated the definition of performance standards in Math and Spanish language.

Analysis of the results from the end-of-grade summative tests has been used by Ministry of Education senior authorities to guide policy adjustments and fine-tune teacher-training initiatives in response to weaknesses identified by test results.

PILOT RATIONALE



One of the many challenges of such a large scale effort in test administration is the management of all the paper testing materials, including test booklets and answer sheets. These materials are transported to a central location in the capital city of Tegucigalpa from all 18 departments. Once in Tegucigalpa, testing materials have to be unpacked, collated, quality controlled and then scanned into an optical scanner for digitizing purposes. Further exacerbating the problem is the need to transfer student responses from test booklets to answer sheets for grades 1-3 where students enter their responses directly into the test booklets. In 2008, this process took nearly three months which delayed the presentation of results and report findings. Both central and local education authorities need to have these results prior to the subsequent school year in order to be responsive to the learning needs of students and training needs of teachers.

Due to the ever increasing number of schools contemplated in future samples, a new strategy for collecting and digitizing student test data is needed. MIDEH/AIR, in coordination with the Ministry of Education proposes decentralizing to six regional field offices, as opposed to performing all the data digitization in one central location. The new approach would serve many purposes, including decreasing the time between test administration and digitization as well as reducing paper loads and transportation logistics. Further, this proposed technology platform would require less overall data manipulation by having student data files transferred electronically to the testing database. It would eliminate the step of transferring student responses for grades 1-3 from test booklets to answer sheets and would allow for quality control measures to be implemented during digitization rather than waiting for all data to be scanned before reviewing for problems. Finally, the implementation of a decentralized approach allows for faster problem solving strategies during test administration and processing results.

To identify state-of-the-art, cost efficient and reliable data collection technologies, two pilot studies were organized by MIDEH/AIR in Tegucigalpa with some additional funding support from the Education Quality Improvement Program (EQUIP1) Leader Award. The pilot studies were to determine which technology can be used in the Honduran context; both options were recommended by the AIR home office in Washington, DC. The first company, Turning Technologies offers Audience Response Systems (ARS), a technology that allows recording real-time responses of an audience to a series of questions. The answers are recorded by remote controls in the hands of the participants and are electronically connected to a computer with a radio receiver and software to process the responses. The ARS can be

used with different handheld devices called ResponseCards. Two models are available, the ResponseCard RF (RF keypad) and the ResponseCard XR (XR keypad). The model used in the pilot study was the ResponseCard XR. Turning Technologies provided (on loan) all the ResponseCards needed for the pilot test. Detailed information on the selection of the ARS and the characteristics of each can be found in Table 1.

Table 1. Comparison between ResponseCard RF and ResponseCard XR

Specifications	 ResponseCard RF	 ResponseCard XR
LCD Screen	No	Yes
Self-Paced data entry	No	Yes
Batteries	Coin cell	AAA
Languages	No	Language pack via updates
Memory	No	Stores up to 5, 100-question test
Alphanumerical responses	No	Yes
Stores user data	No	Yes
Allows to check responses before sending	No	Yes
Successful transmission indicator	Yes, via blinking LED	Yes, on screen message
Multiple digit responses	No	Yes

In addition to the ResponseCards pilot study, MIDEH/AIR conducted a pilot study for Earthwalk mobile computer labs. These computer labs allow the use of laptop computers with long-life external batteries, and also provide the possibility to setup a wireless network to exchange data and share an internet connection. The results of the pilot studies will help inform MIDEH/AIR which technology best suits the needs of test administration and data processing in Honduras.

The ResponseCard pilot study was conducted in 3 phases. In Phase 1 a proof of concept workshop was conducted for the ResponseCard RF ARS. The main objective was to determine if the ResponseCard devices could be used for data collection procedures. In Phase 2, a different ResponseCard device was piloted, the ResponseCard XR to see which ResponseCard was more appropriate. In Phase 3, the chosen ResponseCard, XR was tested in a near to reality pilot study, which is the focus of this paper. This phase was supported by funds from EQUIP1 Leader Award. The time frame for each pilot test was 1 week, with the first day devoted to training participants in the use of the ARS, and 4 days to data entry and quality control (QC) procedures.

PILOT STUDY OF THE RESPONSECARD XR

A week-long pilot test was organized to determine if the ResponseCards XR could be used effectively for data collection purposes. For this pilot study 16 data entry participants and 2 facilitators were contacted to participate in the study. (The data entry participants were offered lunch and a stipend to cover local transportation costs). The facilitators were trained in the use of the Turning Test Administrator software, ResponseCard XR use, and ResponseCard Programmer. Data entry participants were trained in the use of the ResponseCard XR devices. The pilot study lasted 5 days, working 8 hours a day.

During the first day, all the participants were given an introduction to MIDEH/AIR work in Honduras and trained in the use of the ResponseCards XR. None of the participants had prior experience in standardized testing procedures. The participants did not receive training manuals because this material often confused people during the previous proof of concept tests. This strategy led to the desired outcome. When participants had questions, they were addressed as examples for the whole group, instead of being addressed individually. This produced useful feedback and prevented questions and interruptions during the data entry process. Another important finding during the first day was that the group needed to be divided in two sections because of space constraints. Each participant was given a box containing the tests to be digitized, approximately 200 tests per box. They were instructed to enter the tests in groups of 5, and then transmit the data to the computer. Also, they stored their user data in the device, so that this information could be used for QC purposes. Because only one computer was being used to collect data from the ResponseCards XR, the process was slow, and led to a bottleneck causing people to be idle for roughly 15 minutes per round.

On the second day of the pilot, the procedures from the first day were modified slightly. In the new procedure, two computers were set up to collect data from the ResponseCards XR: one of the receivers was set to channel 31, and the other receiver was set to channel 41. Each participant was given two ResponseCards XR. The idea was to take advantage of the ResponseCard storage memory, which is capable of storing up to five tests, with a maximum of 100 questions per test and to use the ID change capabilities of the ResponseCards. Each ResponseCard has a unique code (ID) that avoids interference between the devices when data is being transmitted via radio waves. This security measure also is a downside for the ARS because a participant cannot send more than one test per administration session. To overcome this, Turning Technologies offers a device called ResponseCard Programmer, which allows the user to change the unique code of the ResponseCards XR in real time. This feature allows the participant to send more than one test per administration session, after changing the unique code of the device. In this process the participants entered five tests and when they finished, they had to hand in the ResponseCard to the persons in charge of changing the device ID and transmitting data to the computer. While the data was being transmitted to the computer, the participants took the second ResponseCard and entered another five tests. When they finished entering the second set of tests, they handed in the second ResponseCard and retrieved the ResponseCard they previously used. They repeated this process, which avoided the previously cited bottleneck. This proved to be an efficient strategy for a constant flow of information. To improve QC procedures, the facilitators conducted spot checks to verify that the data was being inputted correctly. There was significant improvement in the process when two participants were selected to serve as assistants to the facilitators. During the first

day of the pilot, it was very difficult for the facilitators to supervise the process, monitor the transmission of data to the computer, answer questions, change the ResponseCards ID's and do QC procedures, especially because transmitting data to the computer and changing the ResponseCard ID's were the most time-consuming procedures of these activities. Arranging for 14 data entry staff, 2 facilitators and 2 assistants for transmitting data to the computer proved to be the most efficient strategy.

It was important to monitor the mood and energy level of the participants. Even after four hours of work they maintained a constant pace for the data entry process. There were some complaints about the physical position they had to adopt while working, and some of them suggested the use of copy holders for the test booklets to make their work more comfortable. During four days of implementing the 14-2-2 setup, no significant problems were presented. By the end of the week, the student responses in 3,751 booklets had been stored in the data base. When the data in the database were compared to the data used for reporting, there was 98% reliability.

Key Findings

1. Participants became familiar with the new technology very quickly. Even participants with little computer experience felt comfortable with the ResponseCards XR.
2. The ResponseCard XR ARS can be used as a data entry solution if the procedures include a plan to maintain a constant flow of data.
3. A user manual is not needed if participants receive a full day of training prior to the actual activity. Not using a manual helps ensure all participants in the group have an equal understanding.
4. Sufficient space and adequate office furniture must be provided for the data entry staff. Some participants complained of back pain because of positions adopted during the activity.
5. Each data entry person should work with one box of booklets at a time, and when they finish they should return the box and be given a new box.
6. QC procedures allow checking a sample of 60 to 80 booklets a day, with 2 people devoted to this activity.
7. When the ResponseCard XR is storing the responses, the user has to wait a few seconds before continuing, this may delay the speed a bit, but allows the user to take the time to check if the responses entered are correct.

PILOT STUDY OF THE EARTHWALK MOBILE COMPUTER LABS

As an alternate solution to data entry procedures, a mobile computer lab offered by Earthwalk was also tested. This company produces mobile labs, which consist of laptop computers and a mobile storage that charges each laptop and contains long life batteries that can be plugged to each laptop. This alternative was tested to see if working with computers would speed up the process, and balance out the additional cost of the computers.

To conduct the pilot study, contact was established with Grupo POPA; they distribute HP computers and Earthwalk products. They provided (on loan) a mobile lab with 16 computers (see Figure 2). This mobile lab consisted of a metal cabinet where computers are stored and charged overnight, 16 long life batteries, a wireless router, and 16 HP laptop computers. The mobile lab allows users to share an internet connection or files over a wireless network. As for software, a 30-day demo version of SPSS

Data Entry 4.0 was used. This software allows the user to design a data entry form, validate the fields, and also provides double entry capabilities to check the data accuracy.

Figure 2: Earthwalk mobile computer labs.



For the Earthwalk mobile lab pilot, a 5 day activity was planned, to follow the steps defined for the ARS pilot. Unfortunately, SPSS Data Entry 4.0 is not compatible with Windows Vista, and 8 computers used Vista. The 8 laptops were substituted with desktops in order to start the pilot test. A group of 16 data entry participants, different from the ARS group, was identified. The same 2 facilitators participated in this pilot test as well. During the first day, participants were introduced to MIDEH/AIR work in Honduras, as well as how to use the software. For accountability purposes, each computer was assigned to a user for the week. It is important to note that participants with little computer expertise felt very anxious when they were being trained. Some even closed the program without saving the information because they were nervous or lacked computer experience.

During the training, some participants did not following instructions or simply got lost. This delayed the process. USB keyboards were used, but not all of them worked with the laptops. Some participants had to use the laptop keyboard for data entry, which was uncomfortable compared to the numeric keypad of a desktop keyboard. The lack of a mouse also delayed the data entry process. Office space was limited, even using laptops. This also slowed the data entry process.

Regarding QC procedures, the SPSS Data Entry form was designed so the user had to double enter all the responses. If the first entry and the second entry didn't match, the program would signal the user to correct the mistake. The user had to look in the booklet for the correct answer and correct it in the form. There were some challenges. Some participants struggled to follow instructions, and given the double entry programming requirement, some users who tried to only enter the scores once were

unable to save their work. This meant they had to re-enter their data, which took additional time . It took about 2 days for the inexperienced user to become more comfortable and confident. It is important to mention that participants with computer experience were able to work quickly, but users with less experience felt nervous and insecure when using the computer, resulting in differences in the number of booklets entered per person per day. Another measure to avoid data loss was to transfer the data files from the computers to a file server. The first approach was to take advantage of the mobile lab wireless networking capabilities, but this procedure had to be abandoned because the users who were not familiar with the use of network resources struggled with file management. The next approach was to use USB flash drives to exchange data between computers. This process yielded 32 files a day, 16 in the morning and 16 in the afternoon. Ensuring that all files were collected and no data file lost was a priority and required careful tracking of information flows.

Two computers collapsed during the pilot test. Some of the data could not be retrieved from the laptops. A data recovery plan is critical during actual implementation to minimize data loss. Besides this incident, the flow of data was constant during the 4 days of data entry. By the end of the week, the student responses in 5,265 booklets were entered, with a reliability of 97.5%.

Key Findings

1. The use of laptops speeds up the process, but sufficient space is needed to ensure a good working environment.
2. If the people hired for data entry don't have much experience with computers or data entry procedures, the process is more complicated and slower.
3. People with minimal computer experience felt anxious while using the computers and tended not to ask questions, which caused loss of data.
4. Having many files to collect during the day may be a problem at the time of consolidating all the data in a single data file for analysis.

COMPARING RESPONSECARD XR VS. EARTHWALK MOBILE LABS

The results of the pilot tests conducted using the ResponseCard XR ARS and Earthwalk Mobile Computer labs are recorded in a table. The results and comparisons are based on the data collected during 4 days, and the reliability percentage was calculated by comparing the data entered by the participants and the actual data stored in the MIDEH/AIR databases. It should be noted that, during the ResponseCard XR ARS pilot test, there were 14 data entry staff, 2 facilitators and 2 data sending and programming staff; for the Earthwalk pilot where there were 16 data entry staff and 2 facilitators.

The results demonstrate that both technologies are reliable. The computers are faster in processing test results. However, the computers are more expensive, and require more highly trained participants. The proper care, usage, and storage of the computers present complex implications as compared to the ResponseCards which can be stored and transported safely in a shoulder carrying case. More specific information related to the results and differences between the two technologies is presented in Table 2.

Table 2: Differences between ResponseCard XR ARS and Earthwalk mobile labs

Specification	ResponseCard XR ARS	Earthwalk mobile labs
Data entry users	14	16
Data entry days	4	4
Working hours	8	8
Total booklets entered	3,751	5,265
Excluded booklets (because of students mistakes)	266	539
Booklets used to compare	3,485	4,726
Average reliability	98%	97.5%
Participant attitude	Good reception, all felt on the same level	Some showed anxiety when using the computer
Data entry speed (booklets per user per hour)	7.33	10.28
QC procedures	Needs extra staff	Can be self checked, but the participant can change the value as he/she feels fit.
Ease of use	The ResponseCards don't require much knowledge.	Requires more than basic computer skills
Data storage and conversion	Needs 1 person for every 8 users to send data to the computer, also takes time to format the files to comply with the evaluation system.	Users have to read the messages displayed to avoid losing data, files already in a format that complies with the evaluation system.
Office space	Doesn't take much space	Takes a lot of space
Transportation and storage	Portable	Bulky
Cost per kit (8-device kit)	\$1,708.00	\$8,798.90 (8 laptop mobile lab)

Under the data collection scenario used, it would take 53 days to process 300,000 test using ResponseCard and 38 days using Earthwalk mobile labs. For more precise information related to how long it would take to process 300,000 test booklets, the calculations for each system are presented in the Table 3.

Table 3: Processing time for ResponseCard vs Earthwalk mobile labs

ResponseCard XR		Earthwalk mobile labs	
Test booklets in sample	300,000	Test booklets in sample	300,000
Test booklets per user per hour	7.33	Test booklets per user per hour	10.28
Data entry users	16	Data entry users	16
Working hours	8	Working hours	8
Data entry centers	6	Data entry centers	6
Data entry days	53	Data entry days	38

The total data entry days were calculated using the following expression:

$$\text{Data entry days} = \frac{S}{B * U * W * C}$$

Where

S: Test booklets in sample

B: Test booklets per user per hour

U: Data entry users (we assume the same number of users in each region)

W: Working hours

C: Data entry centers

Based on the results of the expressions, the shortest processing time is obtained by the implementation of the Earthwalk mobile labs. However, this is not the only consideration. There are other issues that need to be considered, such as transportation of the equipment, availability of adequate power sources and office equipment, safeguard from theft, technical support, alternative uses for the system when not needed for test processing, life expectancy of the equipment and operation costs. These are discussed in the next section.

Conclusions and Recommendations

Prior to drawing conclusions related to this pilot study, it is important to take into consideration general contextual conditions in Honduras. Honduras is a poor country, and more than 70% of the public schools are located in rural areas with difficult access. The administration of external evaluations is complicated by constraints such as in the shortage of qualified and trustworthy personnel in rural regions to assume the responsibility of administering tests and processing results. It is not uncommon to have schools in the sample which are accessible only by horseback, or on foot. The end of the school year (October/November) coincides with the beginning of the rainy season in the northern half of the country, complicating access to schools as well as causing attendance problems when students and/or teachers do not show up. Power outages are not uncommon.

The process of deciding which technology for test processing might be more appropriate for Honduras should take these contextual constraints into consideration. As external evaluation in Honduras becomes more ambitious, with more than 100,000 students evaluated in 2008, and more than 200,000 tests printed and distributed to the national sample of schools, processing results efficiently and effectively in a centralized location (Tegucigalpa) becomes increasingly challenging. Strategies have been developed to decentralize the processing of results in six regional centers. The technology which will be used to process the test results needs to be transported to these regional centers. The technology needs to be safeguarded from theft. The technology will need to be operational even when there are power outages. The technical support in the regional centers will be limited.

In spite of the fact that computers can process test results faster than the ResponseCards, based on the contextual considerations cited, it is recommended that ResponseCards be incorporated in the processing of test results. The main reasons for this recommendation are:

- As a technology, ResponseCards are less expensive than computers.
- ResponseCards are easier to use.
- ResponseCards are easier and less expensive to maintain.
- ResponseCards will function on batteries for a longer period of time
- ResponseCards are less attractive or tempting from the perspective of theft.
- ResponseCards are easier and less expensive to transport to regional centers.
- When not being used for summative test processing, ResponseCards can be used for other activities, such as processing diagnostic test results, processing formative test results, and recording student enrollment by district supervisors.

There are other issues that need to be considered, such as transportation of the equipment, availability of adequate power sources and office equipment, safeguard from theft, technical support, alternative uses for the system when not needed for test processing, life expectancy of the equipment and operation costs.