



**USAID**  
FROM THE AMERICAN PEOPLE



## PERFORMANCE & IMPACT EVALUATION (P&IE)

### Semi-Annual Report, October 2015

**31 OCTOBER 2015**

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by NORC at the University of Chicago.

# PERFORMANCE & IMPACT EVALUATION (P&IE) SEMI-ANNUAL REPORT, APRIL 2015

31 OCTOBER 2015

PN 7384; USAID Contract N0: AID-617-C-12-00006

**PRESENTED TO:**

USAID/Uganda

Joseph Mwangi

**PRESENTED BY:**

NORC at the University of Chicago

Jeffrey Telgarsky

Executive Vice President, Research

4350 East-West Highway, 8th Floor

Bethesda, MD 20814

Telephone: (301) 634-9413

Fax: (301) 634-9301

**DISCLAIMER**

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

# CONTENTS

- HIGHLIGHTS ..... 3
- DESCRIPTION OF ACTIVITIES ..... 4
  - A. Impact Evaluation Activities..... 4
    - Data Quality Assessment..... 4
  - B. Performance Evaluation and CLA Activities ..... 7
    - Systematic observation and documentation of project implementation activities as inputs for the performance evaluations..... 7
    - Collaboration, Learning and Adaptation (CLA) Activities ..... 8
    - Final Performance Evaluation Planning ..... 9
  - C. Results Dissemination Workshop for School Health and Reading Program ..... 9
- ANNEX 1: FEEDBACK from OBSERVATION OF ENUMERATOR TRAINING AND PILOT TEST FOR C1-Round 4, C2-Round 3 and C3-Round 2 EGRA DATA COLLECTION in october 2015..... 11
- ANNEX 2: OBSERVATIONS FROM FIELD VISITS DURING C1-Round 4, C2-Round 3 and C3-Round 2 DATA COLLECTION..... 22
- ANNEX 3: RESULTS DISSEMINATION WORKSHOP (JUNE 17, 2015) - PROCEEDINGS AND PARTICIPANT LIST ..... 25
- ANNEX 4: CHALLENGES TO THE IMPACT EVALUATION, AS PRESENTED IN SEMI-ANNUAL REPORTS PRESENTED IN JUNE 2013, OCTOBER 2013, APRIL 2014, OCTOBER 2014 ..... 26

NORC at the University of Chicago (NORC), in partnership with the Panagora Group, is pleased to submit to USAID/Uganda this Semi-Annual Report recording progress made on the P&IE project between May 1, 2015 and October 31, 2015.

## HIGHLIGHTS

During this reporting period, the NORC/Panagora team:

- Submitted a final version of the Midterm Performance Evaluation Report to USAID in May 2015. This version addressed a final set of comments from Joseph Mwangi to a revised version of the report, which incorporated a larger set of USAID comments. USAID approved the report on May 14, 2015.
- Travelled to Kampala to present the findings of the Midterm Performance Evaluation and the Second Annual Impact Analysis for Result 1, which used the baseline and Round 3 data from Cluster 1 schools and baseline and Round 2 data from Cluster 2 schools to measure the impact of SHRP intervention on reading outcomes. The dissemination workshop took place on June 17, 2015, with attendance from the Ministry of Education (national and district level officials), USAID, implementing partners, and other stakeholders.
- Conducted an initial data quality review of the EGRA Cluster 3 Round 1 dataset. This data, which will be used, along with Cluster 1 and Cluster 2 baseline and follow-up data for each round data, to measure the impact of SHRP's Reading (Result 1) interventions was collected by the IP in February 2015. NORC received the dataset in March 2015, and reviewed it for quality and consistency.
- Conducted a review of the EGRA tool equating process. We found that overall the methodology used by RTI is satisfactory.
- Conducted data quality review for the KAP survey Round 2 data collection, which will serve as the endline for the impact evaluation of Result 2. We discovered significant problems with this dataset, and have described them below.
- Conducted numerous data quality assessment (DQA) tasks related to the EGRA data collection which covered C1-Round 4, C2-Round 3 and C3-Round 2, which took place in October 2015:
  - ▶ In September 2015, reviewed EGRA instruments and provided feedback to the IP. Given that these instruments have now undergone multiple rounds of review, and have been tested/applied in the field 4 times, our comments largely consisted of edits and comments on problematic skip patterns in Tangerine.
  - ▶ Participated in the New Assessors training and Field Practice for C1, C2, C3 EGRA assessors (September 16-18), and Refresher Training for all assessors engaged in C3 EGRA (September 28-October 2).
  - ▶ Travelled to 4 primary schools in Wakiso and Kasese districts covering four language (Lukhondo, Luganda, Ateso, and Lumasaba) to observe EGRA data collection activities (October 6-23).
- Continued the ongoing process of third party monitoring and performance feedback with meetings and activities, based on a monthly calendar and an events and assignments tracker to ensure comprehensive monitoring.

- ▶ Continued to observe RTI/SHRP implementation activities, including leadership and management trainings, school support supervision exercises, and EGRA Assessors trainings. During this reporting period 18 events were observed.
  - ▶ Continued to record, using our observation tools, information and observations on each activity for the upcoming final performance evaluation; and noted appreciative and constructive comments to provide as monthly performance feedback to RTI within the context of the CLA process.
  - ▶ Continued to conduct performance feedback meetings periodically with SHRP senior management based on observations from SHRP events and activities.
  - ▶ Continued monthly meetings between in-country staff and SHRP R1 Reading and R2 Health leads to allow for a more holistic understanding and broader view of overall SHRP progress and achievements.
- Submitted a case study on P&IE's CLA work to USAID's CLA Case Competition in August 2015, entitled "Continuous Evaluation with Real-Time Feedback Fosters Adaptive Program Management." In October, USAID notified the team that, while not an immediate winner, "Continuous Evaluation" was one of the next Top 25 Cases. Cases will be voted on in "Fan Favorite Voting" from October 1 – November 16, with the winner being announced on November 17.

## DESCRIPTION OF ACTIVITIES

### A. IMPACT EVALUATION ACTIVITIES

#### Data Quality Assessment

During this reporting period, P&IE staff engaged in numerous data quality assessment (DQA) tasks related to the EGRA data collection which covered C1-Round 4, C2-Round 3 and C3-Round 2, which took place in October 2015. They included the following:

- Conducted a data quality review of the EGRA Cluster 3 Round 1 dataset. This data, which will be used, along with Cluster 1 and Cluster 2 baseline and follow-up data for each round data, to measure the impact of SHRP's Reading (Result 1) interventions was collected by the IP in February 2015. NORC received the dataset in March 2015, and reviewed it for quality and consistency.

Overall the quality of data was quite good. NORC's initial DQR revealed minimal problematic findings, a majority of which were easily rectified. The NORC DQR noted several instances of internal data inconsistencies throughout the dataset. This includes contradictory data as well as instances where skip patterns should have been observed and weren't. After discussing with RTI, the team decided to drop these inconsistent observations. The DQR noted that several open ended questions yielded responses which were closely aligned but not necessarily the same. The team decided to leave these issues alone given the infrequency and insignificance relative to overall response volume for the questions. There were also a few instances of student scores being improperly aggregated for various learner tests. These issues were flagged, assessed, and subsequently corrected.

Relative to the previous datasets, approximately 25% of learners reportedly didn't know how old they were. This figure was higher than the previous rounds of data collection and RTI was unsure as to why, though suggested that it may have something to do with the students being almost

exclusively enrolled in PI. NORC and RTI needed to revisit the methodology behind the calculation behind one of the score aggregations, as it was believed that only integer outcomes were possible. The team revisited and determined that non-integer responses were also possible. The initial DQR also noted that several modules from the questionnaires were missing. These were since incorporated however several are still missing.

- Conducted a review of the EGRA tool equating process. NORC's subcontractor, School-to-School International communicated with RTI's Research Statistician, Simon King, to better understand the test equating process. We understand that the process involves the following: For the grid tasks (letter sound, non-word), RTI shuffles items within each row, thus keeping the same first 10 items together so that if students miss the first ten and trigger the auto-stop, they are consistently missing the same first ten items. Consistency within the first ten items provides a common set of items that can serve as a basis of comparison between assessments. For the oral reading fluency subtask, RTI makes minor changes to the story – e.g., switching one character name for another, switching mangos for oranges or switching verbs - in an effort to keep the texts as similar as possible. Once piloted, the mean of the new story is compared to the mean of previous stories to create a “means ratio.” The objective is to achieve a means ratio as close to 1 as possible in order to avoid the need to conduct other equating methods. Once the means ratio of pilot tools is brought close to 1, a comparison of the pilot distributions of the two tests is compared to the main study distribution to ensure they are linear (i.e., the distributions “match up” – close enough to be considered comparable, thus excluding the need for equating). If this comparison of distributions between pilot and main study distributions were *not* linear, another method of equating would need to be introduced, but until now, this has not occurred. For the third follow-up (e.g. Cluster 1's third follow-up in the fall of 2015), Mr. Simon recommended that SHRP use the baseline EGRA tool exactly as it is, since enough time has passed that tools could be re-used without significantly impacting analysis, thus eliminating the need for equating. STS and NORC are satisfied with RTI's approach and that it is applying it in a consistent manner.
- Conducted a data quality review of the KAP survey Round 2 data, which will serve as the endline for the impact evaluation of Result 2. NORC staff discovered major problems with the KAP dataset, as described below:
  - Inconsistency between the learner, school and teacher-level datasets whereby some schools appeared in one dataset but not another, resulting in an inconsistent number of schools represented in the three datasets (this problem affects about 7 schools in total). RTI could not explain the source of these inconsistencies due to a lack of field documentation but indicated that possible explanations included refusal from teachers to participate and schools with no eligible learners due to lack of parental consent.
  - Mismatch of school codes between the baseline datasets and the Round 2 datasets prevented a clean merge of the two rounds of data. Feedback was sent to RTI and corrected datasets were sent back to NORC.
  - Lack of field documentation and of a well-organized complete field report. In some schools, the number of learners is well below the target of 30. This is due to a number of reasons including high absenteeism and lack of returned consent forms; however the field report is incomplete (information of Kabale, Apac, Wakiso districts is missing) and does not provide a complete picture of field issues. Rather than providing a consolidated table with total number of learners surveyed and overall response rates as well as reasons for not reaching

the quotas for each school, the report compiles individual field reports from supervisors which provide information in different formats.

- Reviewed and provided feedback on all data collection instruments (EGRA and learner context instruments, teacher/head teacher survey, classroom observation tool, school inventory) for the EGRA. Given that these instruments have now undergone multiple rounds of review, and have been tested/applied in the field 4 times, the main problems noted were with skip patterns, formatting and editorial issues, rather than problems with substantive content.
- Participated in two training workshops conducted by SHRP for assessors. These included a New Assessors training for C1, C2, C3 EGRA assessors (September 16-18), and refresher training for all assessors engaged in C3 EGRA (September 28-October 2).

The P&IE team’s Literacy Experts, Dr. Mark Lynd and Laura Harrington of School-to-School International, travelled to Uganda to participate in the new assessor training, the goal of which was to train new assessors from all three language clusters (C1, C2 and C3) on the administration of the EGRA and accompanying surveys. This was an introductory 3-day training that included a field practice, which took place on September 16-18. The training, which was facilitated by trainers who had participated in previous EGRA trainings, included approximately 25 new assessors, as indicated in the table below:

| Cluster      | Language          | Number of assessors | Region       | District                        |
|--------------|-------------------|---------------------|--------------|---------------------------------|
| 1            | Ateso             | 2                   | Eastern      | Kumi, Katakwi, and Serere       |
|              | Leblango          | 1                   | Northern     | Apac, Lira, and Kole            |
|              | Luganda           | 3                   | Central      | Wakiso and Gomba                |
|              | Runyankore/Rukiga | 6                   | Western      | Kiruhura, Bushenyi, and Kabale  |
| 2            | Runyoro/Tutoro    | 2                   | Western      | Masindi, Kyenjojo, and Kabarole |
|              | Acholi            | 1                   | Northern     | Gulu, Pader, and Kitgum         |
|              | Lugbarati         | 2                   | North West   | Arua                            |
|              | Lumasaba          | 1                   | Eastern      | Mbale, Sironko, and Manafwa     |
| 3            | Lugwere           | 2                   | Mid -Eastern | Budaka, Pallisa, Kibuku         |
|              | Ngakarimojong     | 1                   | North East   | Nakapiripirit, Napak, Moroto    |
|              | Lukhonzon         | 1                   | Mid-Eastern  | Kasese                          |
|              | Lusoga            | 3                   | East Central | Iganga and Kamuli               |
| <b>Total</b> |                   | <b>25</b>           |              |                                 |

NORC’s Resident Evaluation Manager, subsequently participated in a refresher training workshop to train C3 Assessors on the administration of EGRA and accompanying surveys in C3 local languages, Lusoga, Lugwere, Lhukonzon, and Ngakarimajong. The five-day training workshop, which took place on September 28-October 2 included a one day supervisor training. The refresher training, which was also facilitated by trainers who had participated in previous EGRA trainings, included 48 participants from four language communities. Each local language group was paired with a

supervisor, who play a Data Quality Assurance (DQA) role for the group during the field data collection. Trainers used engaging methods such as PowerPoint presentations, demonstrations of EGRA tasks, video clips, group practices, and video Inter Rater Reliability tests.

During these two trainings, NORC team members observed that the IP had implemented some of the recommendations we had made in previous rounds, such as use of videos of simulated EGRA tasks for IRR activities and immediate debriefings after the pilot exercise. However, we also noted that some concerns we had previously alerted the IP to were still present, such as the use of clipped sounds and only accepting one sound per letter, even though some letters clearly have several associated sounds and the lack of clearly defined learning objectives for each training session. We also noted some new concerns such as the fact that some PowerPoint presentations were outdated and the confusion over some IRR procedures. For issues/problems with quick fixes, which in our opinion would contribute to the quality of the instruments, we provided immediate verbal feedback to the IP. Other (bigger) issues that require more complex solutions have been documented in Annex 1 of this report and will be shared with the IP through the monthly Performance Feedback memo well, so they can be taken into consideration for the next round of EGRA data collection.

- Conducted field observations in the Kasese and Wakisodistricts between October 6 and 23, 2015. NORC's Resident Evaluation Manager travelled to the field to observe field work for the EGRA data collection in 4 primary schools in these districts. Observations resulting from these field visits are presented in Annex 2; these observations are being shared with the IP in NORC/Panagora's September-October 2015 Feedback Memo to RTI.

## **B. PERFORMANCE EVALUATION AND CLA ACTIVITIES**

### **Systematic observation and documentation of project implementation activities as inputs for the performance evaluations.**

The P&IE team, led by subcontractor Panagora Group, continued to implement our processes for systematically monitoring and documenting SHRP implementation activities, and our system for carrying out the CLA elements of our contract and providing RTI with performance feedback.

During this period, we continued to follow a sequence of information collection, reporting, review, and feedback that includes the following:

- A monthly meeting between in-country P&IE staff (Resident Evaluation Manager and Sr. HIV/AIDS Specialist) to review the prior month's work and determine content of the monthly report and the performance feedback memo to RTI based on information from the observation reports
- Preparation and submission of a monthly activities report and draft SHRP performance feedback memo by P&IE country staff to U.S.-based team
- Full P&IE team meeting to discuss performance feedback memo and monthly in-country activities (U.S.-based and in-country staff)
- Periodic performance feedback meeting with SHRP project staff
- Performance Evaluation/CLA coordination meeting (Panagora, in-country staff)

The memos, meetings and feedback activities listed above have the end goal of accurately documenting the P&IE team's observations of implementation activities and providing appreciative and constructive feedback to the SHRP team, so they can use our observations and suggestions to improve implementation in real time. The memos will also serve as input into the Final Performance Evaluation.

During this reporting period, our in-country staff attended and observed the following meetings, events, and trainings, and prepared a report on each of them using the appropriate observation and monitoring tool. Each of the reports provided appreciative and constructive observations that were collated and shared with RTI in monthly performance feedback memos.

- EGRA/NAPE Dissemination Workshop, observed April 22, 2015
- CI Consultative Meetings with LLB members on Orthography, observed April 28-29, 2015, Masaka District, Kamapala
- Early Grade Reading Refresher Training of Trainers for C3 local language regions (Lusoga, Lugwere, Lukhonzon and Ngakarimojong), observed May 5, 2015 at Kabwangasi PTC in Pallisa District
- Early Grade Reading Refresher Training for C3 teachers and head teachers on using SHRP PI literacy Instructional materials, observed May 6, 2015 at Kabwangasi PTC in Pallisa District
- SHRP Stakeholders' Engagement Meeting for MoES, District/Municipal, and College Officials, observed May 12, 2015 in Kampala
- Monthly Coordination Meeting with SHRP Result Managers on May 13, 2015 at the SHRP offices
- Early Grade Reading Training for Pre-Service Tutors, observed May 20, 2015 at Shimon CPTC
- Material Development for C2 and C3 Primary Three, observed May 27, 2015 at the National Curriculum Development Center offices
- KAP Mid-Term Survey Data Collectors and Field Supervisors Training Workshop and Field Practice, observed June 8-13, 2015 in Kampala
- Results Dissemination Workshop for School Health and Reading Program organized by NORC and Panagora Group on June 17, 2015 in Kampala
- KAP Mid-Term Survey Field Data Collection, June 23-26, 2015 at Bushenyi, Rukungiri, Kiruhura and Wakiso
- District Planning Meeting, observed on 21 July 2015 at District Education Offices, Kitgum district
- Field Monitoring and Support Supervision, observed from July 22-23 2015, Kitgum District in Kalabong and Pajimo Agweng Primary Schools
- SMC/PTA Orientation Meetings, observed on July 22-23 2015, Kitgum district in Kalabong and Pajimo Agweng Primary Schools.
- End of Field Support Supervision Feedback Session, observed on 23 July 2015, organized at Kitgum CPTC.
- Eastern Region SHRP Stakeholders Meeting, observed on 27 August 2015, Crown Suites Hotel, Mbale district.
- EGRA New Data Collectors Training, observed on September 16-18, 2015 at SHRP RTI offices
- EGRA C3 Data Collectors and Supervisors Refresher Training, September 28 – October 2, 2015

### **Collaboration, Learning and Adaptation (CLA) Activities**

To implement the CLA component of the P&IE contract, we continued to provide performance feedback to RTI, with both appreciative and constructive feedback, focusing on elements of performance where real-time feedback will help to strengthen performance and lead to optimal outcomes.

The performance feedback continues to be drawn directly from the reports of meetings, events, and activities observed by P&IE in-country staff. Each observation tool includes a section to note and record both appreciative and constructive feedback. Our Resident Evaluation Specialist and Senior HIV/AIDS Evaluator continued to collate these comments into a monthly performance feedback memo. Panagora

continued to finalize the memo, integrating comments from the full P&IE team, provide the memo in advance to RTI, and lead the feedback session with RTI's leadership and the full P&IE team participating.

Six months of performance were covered during the reporting period; April, May, June, July, August, and September.

### **Final Performance Evaluation Planning**

During this reporting period, the NORC/Panagora team met for two half-day meetings during which we brainstormed about key evaluation questions and data collection methodologies to be utilized in the Final Performance Evaluation. Based on our experience and lessons learned about effective practices/approaches during the midterm PE, we have developed and are submitting to USAID, as part of our October 2015 deliverable package, a Final Performance Evaluation Implementation Plan.

### **C. RESULTS DISSEMINATION WORKSHOP FOR SCHOOL HEALTH AND READING PROGRAM**

The NORC team organized a Results Dissemination Workshop to share IE and PE findings with stakeholders in Kampala on June 17. Thirty-eight participants attended, including officers from USAID, MoESTS central office, MoESTS partners (DES, NCDC, NAPE), RTI staff and implementing partners. During this one-day meeting, NORC team members presented key findings from the Midterm Performance Evaluation, and Second Annual Impact Evaluation. The workshop agenda is presented below. Annex 3 contains the participant list and workshop proceedings.

**PERFORMANCE AND IMPACT EVALUATION OF THE USAID/UGANDA SCHOOL  
HEALTH AND READING PROGRAM**

**RESULTS DISSEMINATION WORKSHOP AGENDA**

**Protea Hotel**

**JUNE 17, 2015**

|                      |  |
|----------------------|--|
| 09:00-09:30          | Arrival and registration of participants<br>Coffee   |
| 09:30-09:45          | Welcome Remarks, Mr. Martin Omagor, Commissioner of Special Needs Education, Ministry of Education, Science, Technology and Sports |
| 09:45-10:00<br>USAID | Welcome Remarks, Mr. Mark Meassick, Deputy Mission Director,   |
| 10:00-10:45          | Presentation of Performance Evaluation Results, Ms. Betsy Bassan, Panagora Group   |
| 10:45-11:15          | Q&A  |
| 11:15-11:45          | Coffee Break   |
| 11:45-12:30          | Presentation of Impact Evaluation Results, Ms. Alicia Menendez, NORC   |
| 12:30-01:00          | Q&A  |
| 01:00-01:15          | Closing Remarks, Mr. Joseph Mwangi, Senior Strategic Information Advisor, USAID  |
| 01:15-02:00          | Lunch  |
| 02:00-03:00          | Post-lunch Discussion (Optional)   |

# **ANNEX 1: FEEDBACK FROM OBSERVATION OF ENUMERATOR TRAINING AND PILOT TEST FOR C1-ROUND 4, C2-ROUND 3 AND C3-ROUND 2 EGRA DATA COLLECTION IN OCTOBER 2015**

## **INTRODUCTION**

The School Health and Reading Program (SHRP) funded by USAID/Uganda conducted two assessor training workshops for the October 2015 EGRA: a 3-day new assessor training on September 16-18, and a 5-day refresher training on September 28-October 2.

This report presents a summary of that workshop, including effective practices observed during the training as well as questions, concerns, and recommendations.

## **OVERVIEW OF THE WORKSHOP**

The workshop was conducted from Wednesday September 16 to Friday September 18. Dr. Mark Lynd and Laura Harrington from School-to-School International and Ms. Evelyn Namubiru, Evaluation Specialist for NORC, attended the workshop in the capacity of Quality Assurance monitors with the purpose of observing and providing feedback on the quality of the training and tools. Dr. Lynd, Ms. Harrington, and Ms. Namubiru observed the workshop from September 16-18. Ms. Namubiru was the sole Quality Assurance monitor during the Cluster 3 Assessor and Supervisor training. The workshop was facilitated by the following:

- Tracy Brunette, M&E Director, SHRP
- Rehemah Nabachwa, M&E Specialist, SHRP
- Stella Kambugu, CSR Trainer
- Lydia Nakijjoba, CSR Trainer
- Deborah Nakyejwe, CSR Trainer

Also attending at various points were three officers from UNEB.

The room was arranged with approximately 25 participants (some new assessors arrived late and are not included in the tally below) grouped by language clusters at 2 long tables. The languages and locations represented were as follows:

| Cluster      | Language          | Number of assessors | Region       | District                        |
|--------------|-------------------|---------------------|--------------|---------------------------------|
| 1            | Ateso             | 2                   | Eastern      | Kumi, Katakwi, and Serere       |
|              | Leblango          | 1                   | Northern     | Apac, Lira, and Kole            |
|              | Luganda           | 3                   | Central      | Wakiso and Gomba                |
|              | Runyankore/Rukiga | 6                   | Western      | Kiruhura, Bushenyi, and Kabale  |
| 2            | Runyoro/Tutoro    | 2                   | Western      | Masindi, Kyenjojo, and Kabarole |
|              | Acholi            | 1                   | Northern     | Gulu, Pader, and Kitgum         |
|              | Lugarati          | 2                   | North West   | Arua                            |
|              | Lumasaba          | 1                   | Eastern      | Mbale, Sironko, and Manafwa     |
| 3            | Lugwere           | 2                   | Mid -Eastern | Budaka, Pallisa, Kibuku         |
|              | Ngakarimojong     | 1                   | North East   | Nakapiripirit, Napak, Moroto    |
|              | Lukhonso          | 1                   | Mid-Eastern  | Kasese                          |
|              | Lusoga            | 3                   | East Central | Iganga and Kamuli               |
| <b>Total</b> |                   | <b>25</b>           |              |                                 |

A summary of training activities and procedures used in the workshop can be found in Annex A.

The following is a summary of the workshop, including effective practices observed during the training as well as questions, concerns, and recommendations.

## GOOD PRACTICES

### Effective workshop design.

On balance, workshop content and process were very good.

- Adequate overview of EGRA as an introduction to EGRA administration (3 days total, including 1 day school practice) for new assessors who will join other experienced assessors for an additional week-long assessor training.
- Effective management and use of groups.
- Knowledgeable, dynamic facilitators. Most of the training was conducted by DQAs, who were extremely knowledgeable in the content of the EGRA instrument and its use, and dynamic in their delivery and facilitation of group sessions.
- Hands-on, active learning. Overall, assessors were highly engaged in practice activities and had substantial time with facilitators, both in pair practice and in guided practice (group) sessions.
- Effective reference materials. The letter sound reference sheet that was included in training packs is good reference for assessors to remember the correct letter-sound correspondence. Additionally, sound-letter audio files were made available to assessors upon request. These allowed assessors to continue listening to standardized pronunciations.
- Cheat sheet. An EGRA Administration “Cheat Sheet” was provided to participants summarizing key instructions for the administration of each task. This tool proved helpful in remembering and applying key aspects and rules of EGRA administration.
- Observation checklist. An EGRA Assessor Observation Checklist was used by observers during the practice sessions of the training and in-school practice. This sheet summarized steps to be

taken by EGRA assessors, allowing observers to “sign off” on sections for each assessor. These sheets were effective in helping assessors grasp key aspects of EGRA administration as well as providing trainers with a way to measure the ability of assessors and to make decisions on who to retain for data collection.

- h. Visual support. Facilitators made effective use of the video projector for PowerPoints and projection of Tangerine screens to highlight how to mark responses. Facilitators also made effective use of video-recorded presentations of simulations of EGRA subtasks and a video of Lydia, DQA, trainer, and literacy expert, pronouncing each letter of the alphabet while displaying a the corresponding letter in writing.
- i. Videos of simulated EGRA tasks were used to facilitate IRR activities to be used during the full assessor training – following up on a recommendation made in a previous NORC DQA visit.

### **Effective field practice design**

- a. Well planned field practice. Each assessor had the opportunity to practice conducting an EGRA twice with P1-P3 learners in real conditions.
- b. Effective debriefing. Immediately after the field practice, the group discussed their experience. This approach of discussing experience immediately after the experience while it is still fresh leads to pertinent and lively discussion and should be continued. This is a recommendation from the March 2015 report that was taken and applied by RTI. (Additional structure to the discussion can and should be provided – see point 5.a in Concerns and Recommendations below).

### **C3 Assessors and Supervisors EGRA refresher training**

- a. Effective use of language groups for practice activities. After each training session in which participants focused on administering a new task or tool, participants were asked to work in pairs or groups of three in their respective language groups to practice this new skill. The DQAs and trainers made themselves available to support participants during practice activities, and used Assessor Observation Checklist forms to assess skills of participants they observed during practice sessions. Using this approach, facilitators were able to assess participants’ progress in each language group.
- b. Interactive methods. Generally, the training used interactive approaches such as practice activities and experience sharing that offered active learning. Several practice activities and experience sharing sessions were included on the agenda. Participants shared past EGRA challenges with the DQAs and trainers.
- c. Time management. RTI effectively managed time during the entire training. The majority of the sessions were adequately timed, although there were moments where IRR discussion sessions were protracted (see Concerns and Recommendations 6.d & e below).
- d. Training room logistics. The training room was adequate to accommodate four language groups. The room was well organized using tables where each language group was allocated space. During practice activities, groups did not interrupt each other.
- e. IRR group discussion facilitation of results. Each language group discussed IRR results separately. The respective DQAs were present in the discussions to note group and individual challenges to follow up. The discussions tried to resolve all IRR items in question.
- f. Quiz. In order to better gauge comprehension and ability to administer EGRA, three questions were added to RTI’s quiz for a total of eight questions. This allowed for the inclusion of questions related two new subtasks: letter writing and orientation to print. In addition, the quiz included questions on subtasks not covered by video IRR tests. (Video IRR assessed only English sound knowledge, local language sound knowledge and local language oral reading fluency.)
- g. Engaged DQAs and trainers. DQAs were always prepared to address difficulties experienced by the participants. They created time to allow participants to continue practicing problematic

areas with DQAs. Moreover, trainers would highlight and discuss and clarify problematic areas, identified by the DQAs, to the entire group to ensure standard facilitation and common understanding of correct answers.

## QUESTIONS

### Questions shared with RTI during the workshop

- a. Incorrect vs. no response. Some items require the assessor to choose between incorrect and no response; other items combine these two. The distinction was highlighted as important in the training, and was included as one of the items on the quiz on Wednesday. Yet to date, the responses have not been analyzed separately (no answer is counted as incorrect), so it is unclear why this distinction is being highlighted. RTI indicated that they will review whether this distinction should be emphasized in future training.
- b. Sound-letter recognition. Only one correct form of letter pronunciation is accepted as “correct,” yet some letters have more than one sound. This problem was noted in previous DQA visits.
- c. Lack of lower case “L’s”. RTI was asked why no lower-case “L’s” were used in the English Sound-Letter correspondence. They responded that lower case L’s could be confused with upper case “I” so were omitted to avoid confusing pupils. We think this is an appropriate response.
- d. True Random app. We discussed the possible use of an application called “True Random” – a free download that can be used with Tangerine to facilitate random selection of pupils. RTI has been using interval sampling and felt that since assessors were comfortable with this method, they would retain it. We feel that this response is appropriate.
- e. Noting pupil identifiers. During field practice a tablet crashed. STS shared its practice of assessors noting the random identifier on a piece of paper once it is generated to facilitate data cleaning should tablets crash and data sets need to be merged later. RTI indicated that in the past they had experienced so few tablets crashing that they were not sure it merited being an added step in the assessor protocol.
- f. IRB. What is the IRB status for the SHRP EGRA assessments? RTI provided a copy of the letter stating official approval of EGRA assessments in Uganda (see Annex C) This is obtained on an annual basis. RTI received an IRB exemption for on March 17, 2011 (IRB# 11779).
- g. Equating. Test equating is a procedure used to ensure that assessments of similar populations administered at different time points are of comparable levels of difficulty so that results can be compared from one administration to the next. We asked whether RTI had equated their EGRAs for comparison across administrations. They said that EGRAs had been equated in previous rounds “using a means ratio technique.” We are not yet clear on what this means and are still investigating. RTI also said that “moving forward, EGRA scores will be equated after test administration during the analysis phase.” We presume that this means that differences in difficulty will be statistically adjusted in order to make it possible to compare from one EGRA to the next. We are also still investigating the procedure they are using to accomplish this.

## CONCERNS & RECOMMENDATIONS

Note that many of these recommendations may apply to future EGRAs but in some instances cannot be adopted (thus changing the EGRA) at this point since EGRAs should remain consistent from one administration to the next in order to ensure comparability. These recommendations are made in the spirit of improving practice either in current or future iterations of EGRA in Uganda.

## Workshop process

- a. Organization/logistics. The main training room was tight for such a large group of assessors, with insufficient table space for all DQA assessors to work (one had to sit in a chair and use his computer on their lap). Recommend: Use a larger space that allows for the whole group to break out and practice more easily, and to provide sufficient room for DQA observers to work comfortably.
- b. Instructional approach.
  - i. EGRA orientation: Start with paper or tablets? Much of the first day was spent on orienting assessors to paper EGRA administration. With such a short period to get new assessors up to speed, time could be gained by training the new assessors directly on the tablets. (When we raised this with RTI, they said they would consider it, noting that since the beginning of the project, they had never had problems with the tablets that forced them to use paper EGRAs.) Recommend: Begin training directly with tablets for new assessors. A cohort of more experienced assessors could then be trained in the use of paper EGRA administration allowing teams to have at least one member per team skilled in paper administration should there be multiple tablet malfunctions.
  - ii. Difficulty of sounds. As mentioned in previous visits, consistent marking of sound production was one of the biggest challenges noted in this workshop. Pronunciations can vary by language group and region, and sometimes conditions (e.g., room set up) made it difficult to clearly discern sounds. Reviewing sound-letter correspondence in a large group makes it difficult to ensure assessors are able to produce and assess the correct sound. While this training was a boot camp with several language groups, a more in depth review of sound-letter correspondence was scheduled for the following weeks. Nevertheless, steps could be taken to ensure consistent acquisition of this assessment skill. Recommend: Present English letter sounds, then ask each assessor to produce the sound individually. When time is short (as in boot camp), focus on the more difficult sounds.

The following recommendations were made during the last DQA visit and are presented here again.

- Video simulations. The video presentations of subtasks typically consisted of simulations with facilitators, never with real children in real contexts. While there is value in showing simulations with facilitators and peers, additional value could be gained by showing videos of real-life situations. Recommend: Use real-life contexts in videos.
- The effectiveness of IRR could be improved. We recognize this was a boot camp and more in depth discussion on IRR was going to be covered during the week-long assessor training. However, subtask simulations for the IRR activity were done in quick succession. Additional benefit could be reaped with the use of video. Recommend: Use videos for IRR activity wherever possible, coupled with discussion of results immediately following each subtask using stop-action review of scoring.
- Vary presentation style. Most presentations consisted of PowerPoint presentations with long descriptions of how to administer EGRA. Sometimes, support was provided via video or simulations, but most of the time, facilitators talked at length about procedures, quickly flipping from one slide to the next. Recommend: Vary the style of presentation – e.g., instead of telling participants how something works, simulate it, then elicit the steps from them. This is more meaningful to adult learners than being told how to proceed.
- Present sessions are objective-based learning units. Though agendas were provided at the beginning of each day, no learning objectives were presented for any session (e.g., by the end of this session, assessors will be able to...), nor were assessors given opportunities for demonstrate of new knowledge or skills to show that the session objectives had been met. Recommend: Begin sessions by explicitly introducing the learning objectives for each exercise;

by ensuring that assessors understand the objectives and targeted learning outcomes (content or skill) of each practice session; and by giving assessors the opportunity to evaluate whether the learning objectives had been met. Recommend: Make instructional goals explicit.

### EGRA tool

- a. Skip patterns. Several irregularities in skip patterns were found and sent to SHRP M&E Director on September 18, 2015. Recommend: Correct skip patterns to avoid data quality issues and additional data cleaning. A list of these can be found in Annex D.
- b. Unclear instructions. On the vocabulary task/body parts task, it is not clear that the assessors should correct pupils if they get the second example incorrect. Recommend: Ensure that the instructions for all tasks specify that the assessor should correct the pupil if needed when presented with examples.
- c. Letter-sound task:
  - i. Making clipped sounds only as correct. For example, when asked to pronounce the sound of the letter “b,” responses are marked correct if pupils give the clipped version - /b/ - and incorrect if they aspirate – e.g., /buh/. This assessment practice biases the results in favor of treatment schools, where teachers are trained in the “correct” way to assess this skill, even if /buh/ also shows that the child understands the letter sound (as opposed to the letter name for example). Recommend: As one facilitator said, most children will put some vowel sound after letters in order to be heard. This should be accepted as correct. This recommendation was made previously; we are repeating this observation because we believe this practice poses a threat to the validity of the test that should be mentioned in the report.
  - ii. The use of only one sound per letter. On the letter sound subtask, only one pronunciation for each letter is still accepted, even though some letters have two or more correct pronunciations. Recommend: Accept as correct all correct pronunciations of each letter. We are keeping this observation here because the practice continues and thus remains a threat to validity.
- d. Comprehension task. Following the timed reading task, there is only one inferential question in the listening comprehension task in the Lugwere EGRA (we understand all versions are comparable in structure). Per RTI/IRC’s *Guidance Notes for Planning and Implementing EGRA*, comprehension questions following the timed reading task should include two inferential questions. Recommend: Include two inferential questions in this subtask.
- e. Unclear position of the pencil. According to the EGRA instructions on the oral vocabulary task, the assessor gives the pupil a pencil, then asks to put it in front of them. This can lead to confusion since, when giving the pupil the pencil, it is in front of them already. Recommend: Pupils keep the pencil throughout the task.

### Field practice

- a. Assessor marking not always noted. During school practice, sometimes observers fail to note what the assessor is marking. Recommend: Ensure that DQAs observe not only the process that is followed by the assessor but that the assessors are marking the answers correctly.
- b. Limited practice with instruments in cluster language. Due to limited time during the boot camp and the availability of different languages in greater Kampala, field practice was only conducted in English and Luganda, resulting in difficult training conditions for non-Luganda speakers. Recommend: Ensure adequate field practice in languages to be used by assessors.

## Debrief after field practice

- a. Debrief session structure. Relying solely on topics assessors experienced and bring up can lead the discussion in a random manner and result in generic advice (e.g. “it’s important to read the instructions” – this point was made multiple times). Consequently assessors were not left with key points to remember from their field practice. Previous recommendations from the February 2015 report were taken into consideration (e.g. inquiring about assessor experience first) during this training and recognized; however, further improvement could be made that would improve assessor ability to conduct assessments reliably. Recommendations for debrief sessions:
  - (i) Begin the session by posting learning objectives – i.e., what the participants will learn or be able to do by the end of the session – and refer back to these objectives at the end of the session.
  - (ii) Throughout the session, highlight key issues.
  - (iii) Give specific examples: the session was more focused when facilitators cited specific, critical incidents and elicited lessons learned.
  - (iv) Provide visual support – e.g., through projection of questions and advice/answers during the discussion, then print for participants’ future reference.
  - (v) Assess participants’ retention or understanding of key lessons learned during or at the end of the session.

## I. C3 ASSESSORS AND SUPERVISORS TRAINING

- a. PowerPoint presentations were sometimes outdated. It was observed that some PowerPoint slide presentations were outdated and added to assessor confusion. Recommend: Review and updates PowerPoint slides to improve clarity, in particular:
  - i. Focus on Tangerine. The English vocabulary subtask PowerPoint presentation primarily used the paper version of instructions, yet assessors were to conduct EGRA in Tangerine. Recommend: Based all presentations on Tangerine, then hold a 1-day session at the end orienting assessors to paper versions for emergency purposes.
  - ii. Inconsistent instructions. The letter writing PowerPoint presentation stated that letters that are sideways or upside down are incorrect, then at the end of the presentation, this was corrected – that letter images were correct, whether mirror/upside down or sideways as long as they show all parts of the letter. Recommend: Show examples of correct images and present them as such.
- b. Multiple use of the same passage. The orientation to print task, a new task and unique to C3 EGRA administration, uses the same text as the ORF subtask. Consequently pupils are exposed to the text prior to being asked to read it for the ORF subtask, introducing the possibility of the “test-retest effect.” Recommend: Use different texts for each task.
- c. Insufficient information in the Supervisors’ Field Manual. Supervisors do not receive documentation that provides detailed guidance on all subtasks as a reference when in the field. For example, the English Vocabulary subtask contains insufficient notes about administration instructions. Recommend: Include more detailed guidance in the field manual given to supervisors that outlines procedures and key elements of administering each subtask so they have a point of reference while in the field.
- d. English IRR room set-up. The English IRR activity was administered to a large group of assessors. Some participants were seated and others were standing. The seating arrangement was such that it was possible for participants to read from neighbors’ tablets. Those seated were very close to each other and those standing had a good view of participants’ tablets seated in front of them. Recommend: Set-up IRR activity so that participants cannot see each other screens in order to obtain a more accurate IRR score.

- e. Confusion over IRR procedures. During the IRR activity there were instances where the IRR videos varied with scripts or Tangerine examples, making it difficult to explain results. Sometimes participants disagreed with what was being proposed as the correct answer. There were also scenarios when the Tangerine was marked incorrectly by DQAs, trainers or experienced assessors. Moreover, videos were shot from different angles, making it difficult at times to clearly hear the “pupil’s” response. Finally, there was a day lag-time in review of IRR results, making it difficult for participants to recall what happened and why they selected their answer. The majority of IRR discussion revolved around trying to figure out explanations for items where majority of participants disagreed with one another. This was particularly an issue in the local language IRR activities. Recommendations:
  - i. Reshoot videos ensuring clear and auditable capture of “pupil” responses. Clear video footage would reduce possible confusion among participants.
  - ii. If re-shooting videos is not possible, harmonize the paper script and tangerine with what appears in the video.
  - iii. Try to discuss the IRR results of each task as soon as possible after each task. This will help participants remember why they chose the answer they did, participate in the discussion and learn to hear and differentiate the correct answer from the “pupil’s” response.
- f. Pair rotation. During the C3 assessor training participants remained in the same pair throughout. After several days with the same partner some participants became bored with practicing and receiving feedback from the same person. Recommend: Rotate participant pairing every day or two. While new pairs could be chosen by participants, STS has found it beneficial for trainers to assign pairs, thus allowing strong assessors to be paired with assessors that may be having difficulty or assessors that are technologically savvy with those who are slow to adapt the use of Tangerine or tablets.

## SESSION PROCEDURES AND COMMENTS

| Session  | Procedure   | Comments  |
|--|---|---|
| <b>Wednesday 16 SEP</b>  |   |   |
| 8:45-9:15 Welcome; introductions; expectations; outcomes (TB)        | PPT   | <i>Fine<br/>There was objectives for the three day training but no training objectives for individual sessions.</i>   |
| School Health and Reading Program, Assessing Early Grad Reading (LN) | PPT   | <i>Fine</i>   |
| Review of EGRA and the subtasks (C1 & 2 10 tasks, C3 8 tasks) (TB)   | PPT, Flip chart, and assessor packet examples   | <i>Fine</i>   |
| Tea  |   |   |
| Learner agreement and rapport (RN)                                   | Form projected (in Word); facilitator explains<br>Video: Simulation with peers<br>Pair discussions  | <i>Facilitators not circulating to listen to pairs examples<br/>No summary or demo in front of the group to highlight what worked well/what to improve<br/>No "wrong" answers in demos – see what's wrong &amp; note it... &amp; not respond</i>  |
| Intro tablet use (RN)  | PPT   | <i>Encourage facilitators to continue circulating to make sure participants are following<br/>Tangerine -2a)- other specify –skip patter for Q 2b</i>   |
| Sound letter knowledge (DN)  | Form projected (in Word) participants have paper protocol & stimuli; facilitator explains<br>Video: Simulation with peers<br>Video: woman modelling pronunciation (showed during a pause)<br>Simulation: facilitator shows stimulus on PPT and has participants say the letters | <i>Letter sound sheet nice aid for enumerators &amp; Video<br/><br/>Word protocol projected very small<br/><br/>Reviewing letter sounds- hard to know what the correct sound was.<br/>Just large group practice- everyone is to receive a copy of audio file<br/>Only one sound is accepted as correct – project decision<br/>I "e"<br/>Y sound<br/>-presentation more focused on paper administration instructions less on Tablet.<br/><br/>"D is guttural?" It is our understanding the sound for D is made in the mouth with the pallet.<br/><br/>Only one pronunciation is accepted. All pronunciations should be accepted for e? c? k?</i> |

PERFORMANCE & IMPACT EVALUATION (P&IE) SEMI-ANNUAL REPORT

| Session  | Procedure  | Comments  |
|--|--|---|
|  |  | <i>It was said that you put your tongue forward to pronounce the letter y. To pronounce the letter y you move the tongue to the middle of the mouth.</i>  |
| Segmenting Discussion and practice in LL C1 & C2 English (drill) (SK)      |  | <i>Happened after lunch</i>   |
| Lunch  |  |   |
| C1/C2 Non word discussion and practice in English and LL (DN)              |  |   |
| C3 Orientation to print, Letter writing (SK)                               |  |   |
| C1/C2 Oral reading and comprehension: discussion /practice in English (SK) | Simulation: participant & trainer  |   |
| C3 Oral Reading Comprehension LL (DN)                                      |  |   |
| Subtask 5: English vocabulary (LN)   | PPT and simulation with participant.   | <i>All responses were correct. Errors should be thrown in so participants can practice catching errors and showing no response.</i>   |
| Guided practice (break out)  |  |   |
| <b>Thursday 16 SEP</b>   |  |   |
| 8:30 Review Day 1  |  |   |
| English Vocabulary   | Form projected<br>Video demo (identifying body parts) done by trainers.<br>Stage demonstration by trainers in Luganda. | <i>Why not have pupils in the videos?<br/><br/>Why not have participants included in the live demo? - All vocab responses by the "pupil" were correct. Pupils will not have correct answers. Enumerators need to practice how to not give involuntary feedback<br/><br/>Pencil recommendation of leaving the pencil in the pupil's hand from previous DQA not taken up.<br/><br/>Identifying body parts has two examples for practice, instructions for first example clear but no instructions given about treatment of second example, e.g. what happens if a learner's response is incorrect, should Assessor not correct but proceed to the assessment?</i> |
| Oral Reading   |  |   |

PERFORMANCE & IMPACT EVALUATION (P&IE) SEMI-ANNUAL REPORT

| Session   | Procedure   | Comments  |
|---|---|---|
| Differences between local language and English instruments (Same rules!) (TB) |   | Fine  |
| Listening Comprehension (Local language)                                      |   |   |
| Break   |   |   |
| Reviewing Cheat Sheet, Administration rules (C3 group only)                   | PPT   | Cheat sheet is a great idea<br><br>C3 Instructions on letter writing not clear especially what is considered correct when it comes to mirror letters and upside down letters.<br><br>*This was corrected during the full C3 training.   |
| Guided practice (pairs, groups of 3)  |   |   |
| Local Language Drilling – Letter sounds, non-words, segmenting                |   |   |
| Lunch   |   |   |
| Learner context interview, discussion and practice (RN)                       | Project Tangerine and walk through each page, asking questions along the way<br>Pair practice with observations |   |
| Pair Practice, one on one observation   |   |   |
| Recap of day, prep for field visit  |   |   |
| <b>Friday 18SEP</b>   |   |   |
| 8:30 Meet for field practice  |   |   |
| Leave for field practice  | Traveled to schools in clusters   |   |
| Field practice, each new assessor practices with 2 learners                   | Field practice with two learners. DQAs roamed to observe assessors  | <i>Due to language limitations field practice was challenging</i>   |
| Lunch   |   |   |
| 2:00 Practice and issues from the field practice                              | Open facilitated discussion on field experience   | <i>A more structured discussion that incorporates participant experiences with key points to keep in mind during administration.</i>  |
| Intro to IRR, practice for full training                                      | Live simulation of subtasks used  | <i>Use videos for IRR that allow for stop-action review of subtasks.<br/><br/>Reviewing results of subtasks before going on to following subtasks would help assessors keep what they scored and why fresh in their minds, rather than having to try to remember a task that took place 10 min ago.</i> |

## ANNEX 2: OBSERVATIONS FROM FIELD VISITS DURING C1-ROUND 4, C2-ROUND 3 AND C3-ROUND 2 DATA COLLECTION

**MEMO TO:** RTI/School Health and Reading Program (SHRP)  
**FROM:** NORC/Panagora Performance and Impact Evaluation (P&IE) Team  
**DATE:** March 12, 2015  
**SUBJECT:** Performance feedback on observed SHRP activities in February 2015

**MEMO TO:** RTI/School Health and Reading Program (SHRP)  
**FROM:** NORC/Panagora Performance and Impact Evaluation (P&IE) Team  
**DATE:** October 20, 2015  
**SUBJECT:** Performance feedback on observed SHRP activities in September and October 2015

### EXCERPT FROM MEMO RELEVANT TO OBSERVATION OF FIELD WORK

**C1-Round 4, C2-Round 3 and C3-Round 2 Data Collection Observed Lukhonzo group Oct 6-7, Luganda group Oct 15-16, Ateso group Oct 20-21 and Lumasaba Oct 22-23, 2015.**

**Purpose:** The main activity was collecting EGRA data from learners along with accompanying information from teachers and school administrators.

**Content/Approach:** Assessors conducted EGRAs with students, along with a learner context interview, teacher and head teacher interview, and school inventory in each of the schools. Classroom lesson observations were also conducted in a subsample of schools. The DQA observed classroom lessons and the rest of the data collection was performed by Assessors and Team Supervisors. Each Local Language group had three to four teams of three Assessors and one Supervisor (a few teams were composed of three), who reported to the DQA; each team conducted EGRA in one school per day. The DQA was the overall supervisor for the teams, and worked with a different team every day.

**Relationship to work plan:** The outputs for this activity correspond with Intermediate Result 1.5, Programs and Policies informed by data and research.

### Appreciative feedback

- **Teams arrived at schools on time.** Teams arrived at the schools early in the morning, with enough time to prepare the assessment area, sample the learners, before beginning the assessments on time.
- **Teams followed field protocols while at the schools.** As specified in the field protocols, upon arrival at the schools, Assessors first went to the head teacher's office to introduce themselves and request permission to carry out EGRA activities. They also respected the teachers and learners they worked with.
- **Teams established good rapport with students.** Realizing that good rapport with students made them more likely to consent to participating in the assessments, Assessors worked to build a good connection with the students upon first meeting them by taking time to chat on

various topics of interest. This initial rapport also helped to build students' confidence to respond to the assessments.

- **Teams followed survey sampling methodology.** Selection of learners and teachers to participate in survey activities was done based on survey sampling instructions. It was a random selection and all participants were registered on a sheet of paper for easy follow up.
- **Assessors prepared an appropriate sitting arrangement for the assessments.** During assessments, assessors and students sat opposite each other on different seats, which encouraged good eye contact and offered ample space and good positioning for reading and writing during the assessments. Assessors also had space to place their materials and keep the tablets from view so as to not distract students during assessments.
- **Assessors followed instructions outlined for each task.** At all times, assessors read the task instructions as provided on the tablet. The instructions were given in area local language most familiar to many learners. Learners not conversant with area local language or English were discontinued and replaced. .
- **Assessors practiced examples provided for the different tasks.** On all tasks with examples, assessors demonstrated the examples as they appeared in the instructions and did not take shortcuts.
- **Assessors marked the last item attempted on timed tasks.** During timed tasks Assessors were mindful to mark the last item a learner attempted.
- **Assessors encouraged learners to point at items read.** C1 and C2 assessors encouraged learners to point by themselves on reading assessment tasks. Learners were reading at their pace, assessors only supported them in pointing at next item in case they stayed quiet.
- **Interview environment adequately ensured privacy and learner concentration.** Learner assessments are conducted in school compounds. This round, assessors were keen to choose places with the least distractions to ensure learner concentration. Assessors tasked schools to provide free rooms where available and used spaces far from or behind classrooms. NORC local staff also relocated a few assessors who were found conducting learner assessments in inappropriate places.
- **DQAs supervised and provided continuous mentoring of assessors to master EGRA administration while in the field.** Each DQA supervised a different team of assessors each day. DQAs organized daily debrief meeting to discuss feedback on individual/group performances. In these meetings, they also discussed feedback from RTI main office about data quality assurance. They mentored assessors on weaker areas identified. Assessors were also free to ask questions on problematic areas encountered that day.

### Constructive Feedback

- **Inconsistencies noted in administration and marking some tasks.**
  - **Inconsistencies in tapping the start button for timed tasks.** For timed tasks, such as letter sounds and reading comprehension, some assessor began the timer before providing the prompts to learners, and some began the timer after providing the prompts. With Lumasaba group, we observed an assessor who restarted the timer after providing the prompts on Letter sound knowledge task. This difference in assessment administration can affect results.
  - **Luganda letter sound knowledge task:** The set of letters used in the assessment started with letter “O”, which has same name and sound. As a result it was difficult for

assessors to apply early prompts in case learners read letter names instead of letter sounds. Some assessors prompted while others did not. This created inconsistencies in administering the assessment, which may affect the results. This should be clarified in future trainings.

- **Segmenting subtask.** Many C3 assessors had difficulty reading to the learner the whole word without segmenting it in syllables. This affected the assessments and adequate marking. It was hard for assessors to judge a learner saying what was read and those breaking the word into syllables. This should be emphasized in future trainings.
- **Oral reading and listening passage and comprehension.** Some assessors did not comply with the 10 seconds rule when asking comprehension questions. They were always in a hurry to ask the next question when a learner was quiet, not giving them a time to think through their responses. The 10 seconds rule should be emphasized in future trainings.
- **English vocabulary task.** Observed inconsistencies administering subtask on identification of body parts. Some learners were prompted to point to the body part whereas others were not. This creates differences in administering the task, which may affect results. It is important to clarify this issue, or add an instruction alerting assessors to remind the learners to point during the assessment, for consistency.
- **Learner context interview.** Some assessors judged learners responses based on the school surroundings and recorded what they thought was applicable. This was common when it came to questions about a learner's home environment e.g. whether there is a computer, car, etc. Assessors should be encouraged to record responses as provided and for strange responses, it should be noted down in comments section at the end, for the attention of data editors.
- **C3 assessors pointed for learners throughout on reading assessments.** C3 assessors were not encouraging learners to point for themselves during reading tasks. Assessors pointed for learners throughout. This may affect results because learners tend to read at pace determined by the assessors. Since C1 and C3 (Luganda, Ateso and Lumasaba groups) were able to aid learners point by themselves when reading, it is possible to train Lukhonzho language group to do the same.
- **Assessors were competent using the tablet technology.** All Assessors were able to use the tablets: they easily scrolled through different tasks, began and ended tasks, saved interviews, and created new interview forms for different students. However, they encountered freezing tablets and skipping in some subtasks. It was not clear why this happened to some tablets.

## **ANNEX 3: RESULTS DISSEMINATION WORKSHOP (JUNE 17, 2015) - PROCEEDINGS AND PARTICIPANT LIST**

## **Dissemination Workshop - Performance and Impact Evaluation of the USAID/Uganda School Health and Reading Program (SHRP)**

**17 June 2015**

**Protea Hotel, Kampala, Uganda**

### **OPENING REMARKS**

The Commissioner of Special Needs Education from the Ministry of Education Science Technology and Sports (MOESTS), Mr. Martin Omagor, opened the workshop by acknowledging the presence of the various participants, in particular recognizing USAID as the donor and the RTI as the lead implementing partner. He noted the Performance and Impact Evaluation (P&IE) contract by stating the important role that research, monitoring, and evaluation plays in helping ensure program outputs and impact.

### **REMARKS FROM USAID**

USAID's Deputy Mission Director, Mark Meassick, noted the low primary school reading statistics and high secondary school dropout rates as the drivers behind the Government of Uganda's use of mother tongue instruction in early grade reading. He noted that the Government's early grade reading approach is now being scaled up into 85 districts under SHRP, the Global Program for Education, and USAID's recently awarded Literacy Achievement and Retention Activity (LARA). Mr. Meassick mentioned the importance USAID attached to collaborating, learning, and adapting (CLA), and the role of the workshop in advancing that agenda, specifically providing a venue for collaboratively sharing key evaluation findings, and promoting better understanding and use of recommendations. He recognized NORC's (National Opinion Research Center at the University of Chicago) work on the impact and performance evaluations as well as adaptations RTI has made and continues to make as a result, e.g., scripted lesson plans for the School Family Initiative. He concluded by noting the importance of monitoring and evaluation in meeting SHRP's program objectives.

### **Presentation of the SHRP Mid-Term Performance Evaluation**

Betsy Bassan, performance evaluation team leader, Panagora Group, presented the results of the performance evaluation (see attached PowerPoint)

### **Discussion:**

During the ensuing discussion, a number of comments and questions arose, as well as clarifications regarding additional adaptations and program initiatives by RTI since the evaluation. This includes commencement of community mobilization activities including use of radio, leadership development activities, and use of technology (SMS texts to teachers and schools). A question was raised on the sustainability of the Field Assistant position and how that can be continued beyond external funding. A question was raised on the integration of health and HIV/AIDS in age appropriate ways in the early grade reading primers; while there is some, there is an opportunity for more, particularly in the P4 materials. The importance of maintaining the control schools for an additional year was discussed in response to a question regarding the timing for including control schools and districts into the program. The importance of local language board was mentioned. A participant mentioned the need for posters in schools to be produced in local languages so that parents can understand them.

## **Presentation of the SHRP Impact Evaluation - YEAR 2**

Alicia Menendez (NORC), Chief of Party and Impact Evaluation Team Leader (see attached PowerPoint)

### **Discussion**

Participants were excited about the figures and statistics and had many questions.

The USAID COR for P&IE, Joseph Mwangi, underscored the importance of impact evaluations with control groups in identifying whether an intervention is having a real impact, different than what is occurring within the control group not receiving the intervention.

A participant asked whether other similar interventions in the control districts could have caused the scenario we see. A SHRP representative said this has been examined and is not the case. Alicia said that some improvement in controls schools is expected as the students in these areas are attending school and learning, albeit not with the SHRP methodology. She noted that achievements in the both the treatment and control schools are quite low, and much less than desirable.

A participant noted the importance of understanding results variation by language, e.g., performance in R/R and Luganda languages is higher without SHRP interventions. Alicia noted the complexity of comparing across languages, which includes possible socio-economic differences. A SHRP representative noted that some languages use a particular letter very often, possibly making it easier for that group to perform better in letter sounds.

A participant noted the importance of community and parental participation, including in making sure students have food in schools, and that now that community mobilization activities are underway in SHRP, this may have a positive effect.

A participant asked whether the control schools include schools trained by Mango Tree, possibly accounting for their performance improvements. Alicia clarified that schools treated by Mango Tree were excluded from the impact evaluation.

A participant requested more interpretation of each graph. Alicia explained the exact interpretation of each graph in detail.

A person in the audience suggested that there was contamination especially through radio programs because they are listened to by those on control and treatment districts/schools. Alicia explain that the goal is not to evaluate the effect of radio programs (and that during the period of analysis they were not in place anyway). A question was asked regarding minimum standards for early grade reading. Alicia replied saying there are not standards for Uganda that can be used as reference, however the levels of achievement even in the best performing groups are well below acceptable levels of P2 reading.

The SHRP AOR, Mariella Ruiz-Rodriquez, noted the enthusiasm among schools for SHRP interventions, and the importance of USAID having proof through a rigorous impact evaluation that the

SHRP interventions works. Mr. Mwangi mentioned the importance of third party monitoring, issues raised by the performance evaluation (e.g., sustainability), the need for action on the points raised. He also said we must understand the numbers behind the impact evaluation: does learning in mother tongue make a difference in learning to read if we compare with the controls? Is the cost justified if the control schools can catch up in other ways or if there is not much difference between reading performance in the control and treatment schools. He concluded by saying that the experiment in Uganda will inform USAID investments in reading.

### **Closing Remarks, Assistant Commissioner for Basic Education**

The Assistant Commissioner concluding by expressing thanks for the results shared by the performance and impact evaluations. He reiterated the phrase from the performance evaluation, “what gets measured gets done.” He thanked USAID for funding SHRP and the rollout of the new programs (GPE and LARA) that will be operational in many additional districts. He noted that people are already excited about the SHRP and noted the spill-over effects to control schools. He closed the workshop by saying “*The child’s need is now*” and “*The struggles continues*” and thanking the organisers.

## List of Participants

| <b>Name of participant</b> | <b>Institution</b> | <b>Title</b>                  | <b>Email contact</b>   |
|----------------------------|--------------------|-------------------------------|--|
| Mark Meassiek              | USAID              | DMD                           |  |
| Joseph Mwangi              | USAID              | Strategic Information Advisor | <a href="mailto:jmmwangi@usaid.gov">jmmwangi@usaid.gov</a>                     |
| Rhobbinah Ssempebwa        | USAID              | PMS                           | <a href="mailto:rssempebwa@usaid.gov">rssempebwa@usaid.gov</a>                 |
| Jackie Batanda             | USAID              | PLS                           | <a href="mailto:jbatanda@usaid.gov">jbatanda@usaid.gov</a>                     |
| Mariella Ruiz Rodriquez    | USAID              | Director EYCD                 |  |
| Douglas Blanton            | USAID              | Development Officer           | <a href="mailto:dblanton@usaid.gov">dblanton@usaid.gov</a>                     |
| Opaman Amos                | UNEB/NAPE          | Senior Education Officer      | <a href="mailto:opamos@yahoo.com">opamos@yahoo.com</a>                         |
| Magera Deborah             | NCDC               | Curriculum Specialist         | <a href="mailto:mageradeborah@yahoo.com">mageradeborah@yahoo.com</a>           |
| Omagor Loican              | MoEST/SNE          | Commissioner/SHRP Coord.      | <a href="mailto:mlomagor2006@yahoo.com">mlomagor2006@yahoo.com</a>             |
| Dr CT Mukasa Lusambu       | MoEST/BE           | Assistant Commissioner        |  |
| Mutazindwa HA              | DES                | Director                      | <a href="mailto:hzyfa2008@yahoo.com">hzyfa2008@yahoo.com</a>                   |
| Derek Nkata                | RTI/SHRP           | Deputy Chief of Party         | <a href="mailto:dnkata@shrp.rti.org">dnkata@shrp.rti.org</a>                   |
| Consilous Rwanyonga        | RTI/SHRP           | Literacy Field Coordinator    | <a href="mailto:crwanyonga@shrp.rti.org">crwanyonga@shrp.rti.org</a>           |
| Scholastica Tiguryera      | RTI/SHRP           | Prof. Dev. Specialist         | <a href="mailto:stigurjera@shrp.rti.org">stigurjera@shrp.rti.org</a>           |
| Saeeda Prew                | RTI/SHRP           | Chief of Party                | <a href="mailto:sprew@rti.org">sprew@rti.org</a>                               |
| Julius Mulindwa            | RTI/SHRP           | DFA                           | <a href="mailto:jmulindwa@shrp.rti.org">jmulindwa@shrp.rti.org</a>             |
| Robinah Kyeyune            | RTI/SHRP           | Literacy Advisor              | <a href="mailto:rkyeyune@shrp.rti.org">rkyeyune@shrp.rti.org</a>               |
| Tracy Brunette             | RTI/SHRP           | M&E Director                  | <a href="mailto:tbrunette@shrp.rti.org">tbrunette@shrp.rti.org</a>             |
| Sarah Kyobe                | WEI/SHRP           | Health Advisor                | <a href="mailto:skyobe@shrp.rti.org">skyobe@shrp.rti.org</a>                   |
| Norah Kushaba              | WEI/SHRP           | Finance Manager               | <a href="mailto:kgnorah@yahoo.com">kgnorah@yahoo.com</a>                       |
| Elizabeth Ndyanabangi      | WEI/SHRP           | FPO                           | <a href="mailto:endyanabangi@shrp.rti.org">endyanabangi@shrp.rti.org</a>       |
| Lillian Viko               | VSO                | Program Manager               | <a href="mailto:lillian.viko@vsoint.org">lillian.viko@vsoint.org</a>           |
| Susan Mubbala              | SIL-LEAD           | Program Manager               | <a href="mailto:susan_mubbala@sil-lead.org">susan_mubbala@sil-lead.org</a>     |
| Augustine Wandera          | QED                | M&E Director                  | <a href="mailto:awandera@qedgrouppltd.com">awandera@qedgrouppltd.com</a>       |
| Geri Burkholder            | LARA               | COP                           | <a href="mailto:gburkholder@rti.org">gburkholder@rti.org</a>                   |
| Asiimwe Wilson             | CSR/RTI            | Director                      | <a href="mailto:direcsr@gmail.com">direcsr@gmail.com</a>                       |
| Tom Evans                  | ADC                | Country Manager               | <a href="mailto:tevans@africadevcorps.org">tevans@africadevcorps.org</a>       |
| Musinguzi Christopher      | Kabale district    | CCT                           | <a href="mailto:totamusinguzi@gmail.com">totamusinguzi@gmail.com</a>           |
| Javan R Lwamafa            | Kabale district    | Retired Principal             | <a href="mailto:ajavan.lwamafa@gmail.com">ajavan.lwamafa@gmail.com</a>         |
| Odwilo Simon               | Lira district      | Principal Loro PTC            | <a href="mailto:simonodwilo@yahoo.com">simonodwilo@yahoo.com</a>               |
| Vastina Beyendera          | Kabale district    | District Inspector            | <a href="mailto:vastinabeyendera@gmail.com">vastinabeyendera@gmail.com</a>     |
| Okalebo Peter              | Lira district      | CCT                           | <a href="mailto:peterokalebo@gmail.com">peterokalebo@gmail.com</a>             |
| Mutahunga Elia             | Kabale district    | Municipal Inspector           | <a href="mailto:mutahungaelia@gmail.com">mutahungaelia@gmail.com</a>           |
| Bwonyo Bosco               | Lira district      | District Inspector            | <a href="mailto:acupbwonyo@gmail.com">acupbwonyo@gmail.com</a>                 |
| Jayne Frances Offungi      | Lira district      | Municipal Education Officer   | <a href="mailto:jayneffunge@gmail.com">jayneffunge@gmail.com</a>               |
| Atto James                 | Lira district      | Deputy Principal              | <a href="mailto:attoowinyakullu337@gmail.com">attoowinyakullu337@gmail.com</a> |
| Ndagga Daniel              | Entebbe/Wakiso     | Municipal Inspector           | <a href="mailto:kikoblasio@yahoo.com">kikoblasio@yahoo.com</a>                 |
| Nshenereirwe Jacqueline    | Entebbe/Wakiso     | CCT                           | <a href="mailto:jacqueynshem@yahoo.com">jacqueynshem@yahoo.com</a>             |

## **ANNEX 4: CHALLENGES TO THE IMPACT EVALUATION, AS PRESENTED IN SEMI-ANNUAL REPORTS PRESENTED IN JUNE 2013, OCTOBER 2013, APRIL 2014, OCTOBER 2014**

1. Result 1: Delays in the implementation of Result 1 continued through October 2013. Although all the trainings have taken place, including refresher TOT and teacher training on Cluster 1 PI materials (teacher guides and primers), these instructional materials were still being distributed to Cluster 1 schools as late as September/October 2013. Our understanding is that materials have not reached all schools at the time of writing this report. In addition, the original plans that included three different treatment arms were modified and treatment was uniform across all schools. Baseline data collection for Cluster 1 was completed successfully in February, and follow-on data collection for Cluster 1 is being fielded among a sub-sample of primary schools. These delays and modifications to the implementation do not pose serious risks to the evaluation at this juncture. We plan to evaluate the impact of the program as it was implemented.

While the implementation changes/delays are not a risk to the evaluation design, an important fact to keep in mind, however, is that we do not expect to see the impacts of the full Result 1 intervention (teacher training and instructional materials) during this first impact analysis, using Oct/Nov 2013 data. However, the Oct/Nov 2013 data will provide us with an opportunity to measure the impact of multiple rounds of teacher training.

2. Result 1: The most recent version of the SHRP PMP indicates that no data will be collected from Cluster 2 in 2016. Going forward with this decision would imply that the impact evaluation for Cluster 2 would only be possible for P1 and P2 but not for P3. Given that Cluster 1 did not receive the full intervention in 2013, Cluster 2 will be the only group that will have a chance to receive three years of full treatment from the beginning of their primary education. The Evaluation Expert already mentioned this omission as a concern to USAID and to the IP as well.
3. Result 1: Data for the second EGRA wave are being collected as we write this report. Initial information from the field indicates low response rates (i.e. low numbers of students are being found) in the schools in the Central Region compared to baseline. We are currently working with the IP to try to address this problem and minimize the risks of having a small sample.
4. Result 2: There are several issues related to sample that have surfaced during the ongoing KAP data collection, which are likely to pose threats to the evaluation of Result 2 activities.

We noted in our first Semi-Annual Report that, it was not possible to include boarding or partial boarding schools -very common among post-primary establishments- in the evaluation sample, given delays in obtaining parental consent for the KAP Survey during the school year. We decided, however, to use the second round of the KAP survey (KAP2) to collect additional baseline data from Cluster 1 boarding and partial boarding post-primary schools by distributing parental consent forms to students before the school break. The idea was to ensure that the

baseline survey consisted of a representative sample of post-primary schools, thereby allowing us to generalize the results of the impact evaluation to all such schools in the districts.

We recently learned of several problems that the IP is encountering with the supplemental boarding school component of the second round of KAP surveys. These problems could potentially have serious implications for sample size and the representativeness of the post-primary school sample:

- ▶ The IP faced resistance to data collection activities from some schools, where principals cited concerns that the survey would take away from exam preparation time (national exams in post-primary schools begin in the 2nd week of October) and some head teachers did not distribute consent forms to students at all. These schools could not be interviewed.
- ▶ Some schools closed before the end of the term and consent forms were not distributed on time. These schools could not be interviewed.
- ▶ Other programs related to HIV/AIDS have interacted with some of the schools and, therefore, head teachers decided not to participate in KAP. This is particularly the case of private secondary schools. These schools could not be interviewed.
- ▶ The sample frame that the IP provided NORC for selection of the school sample for the KAP2 contained errors; it included schools that already participated in the first round of KAP. In cases where these schools were randomly selected for the KAP2 sample, they had to be removed from the sample and, where possible, replaced.

NORC has requested from the IP a list of all schools in the KAP2 sample with disposition comments for each of the schools. After evaluating the situation we will have a clearer impression of the effect that these problems can have on the evaluation. At a minimum, we expect a reduction in sample size.

5. Result 2: As mentioned above, SHRP decided not to include post primary establishments in new treatment districts (Cluster 2 and after). Therefore, we will only be able to assess the impact of the Result 2 intervention on post-primary educational facilities for Cluster 1 schools.
6. Result 2: Based on the most recent PMP, we note that the Result 2 intervention will no longer be conducted in Cluster 3 districts and schools. As a result, NORC will focus its evaluation of Result 2 on Cluster 1 and Cluster 2 schools.
7. Result 1: Given program implementation delays in Year 1, the academic term was delayed for one week in the 11 districts of Cluster 1 where the IP is working in order to build in time to prepare and have teacher guides ready for the second training of teachers. Additional classes to compensate for the one week delay are not currently planned. An equivalent delay did not occur in the control district schools; therefore, the academic year in those schools will be one week longer. We do not anticipate a visible effect, but it is worth mentioning how the reality of the program may affect the evaluation.
8. Result 2: After NORC selected the samples for the impact evaluation of the School Health activity, the focus of the intervention underwent some changes in order to align with PEPFAR priorities. We were informed that the intervention would target large schools (with over 150 students) in high HIV prevalence districts; this brought into question the external validity of the impact evaluation and the ability to include non-intervention districts with similar characteristics to treatment districts in the design. However, these new criteria do not seem to have affected the actual selection of districts and we will proceed with the original evaluation design. However the number of treatment schools

increased. The IP went ahead with the selection of schools for treatment and control before NORC could approve the selection. As a consequence no replacements for control schools were selected. This can result in a smaller sample than needed. The Evaluation Expert discussed this issue with the IP and USAID.

#### April 2014:

1. **Result 1:** During the Cluster 2 EGRA training and pilot test, the P&IE team observers noted some issues related to the implementation of three specific EGRA subtasks – Letter Sound Knowledge and Word Segmenting, and Oral Passage Reading - and that could have negative implications for the impact evaluation. Annex 3 describes the issues in great detail and also lays out the implications for the impact evaluation. In short, SHRP was using very stringent requirements for accepting letter sounds as correct; for example, while the EGRA toolkit states that “For consonants that can represent more than one sound (i.e., c, g), either answer is acceptable. For vowels, either the short or long sound is accepted (/i/ as in pin or as in pine),” in the SHRP implementation of EGRA only one sound per vowel was being accepted as correct. As well, local pronunciations of words – e.g. “muzzah” for mother – were being marked as incorrect. This raises the concern that learners who actually know correct letter sounds are assessed as not knowing them, since trainers were instructed during training to mark as wrong any very slight deviation from the “ideal” sound of a letter.

This approach can bias the assessment in favor of treatment schools, where students are being taught one correct letter sound or a specific pronunciation of a word, relative to control schools, where a broader set of letter sounds and pronunciations are being taught. We can take as an example the letter B: the sound of letter B is /b/ or /buh/. Both sounds are correct and accepted as building skills towards early reading ability. However, the current application of EGRA in Uganda only accepts a perfect clipped sound /b/ as correct. Marking /buh/ as wrong is likely to punish learners in control schools more than it punishes learners in treatment schools, because teachers in treatment schools are trained to teach /b/ as the only correct sound while teachers in control schools are likely to use either /b/ or /buh/ given that both sounds are considered correct. This approach of “teaching to the test” will bias impact findings in favor of treatment schools. NORC is exploring options for measuring this bias in order to adjust impact measures; towards this end, we briefly discussed some alternatives with USAID, such as measuring the bias by conducting experiments to test more and less restrictive versions of EGRA administration.

2. **Result 1:** Possible contamination of controls. Because the SHRP team is not planning to expand SHRP implementation to additional districts for Cluster 1, they are planning to implement Result 1 activities in control CCTs in the 11 original districts starting in 2014 in order to meet target numbers of trained teachers. However, they plan to exclude the control schools within the control CCTs which were selected for the EGRA data collection and intervene only in the schools from control CCTs which have not been included in the EGRA data collection. Hence, according to the SHRP M&E Team Lead, no teachers in any grade (P1 through P4) in the EGRA control schools will be trained; nor will instructional materials be distributed to these schools. CCTs associated with

---

<sup>1</sup> Similar problems exist with many other consonants such as D, T, P, K, G, etc.

<sup>2</sup> RTI International, EGRA Toolkit, March 2009  
<https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=149>

these control clusters will be strictly instructed not to provide any assistance to these control schools.

Strict exclusion of control schools from treatment is critical for the integrity of the impact evaluation design. While SHRP staff has assured us that no control schools will receive any semblance of the Result 1 interventions, we are nonetheless concerned by the possibility of contamination through CCTs or spillover of materials. Any contamination of the control schools will lead to underestimation of the effects of the SHRP Result 1 interventions. We have made this concern clear to both the IP and USAID, and requested that SHRP put in place adequate safeguards to ensure that the control schools in our sample will not be contaminated.

3. **Result 1: Non-systematic replacement of sample schools.** During Cluster 2 baseline data collection in Mbale district, the SHRP team opted to exclude control schools that use or were presumed to use Luganda and English instead of Lumasaaba as the medium of instruction. The appropriate procedure to replace these schools (following the replacement rule provided) was not followed. Two of these non-Lumasaaba instruction schools were replaced by schools in which the medium of instruction is Lumasaaba; these replacements were picked from the list of preselected schools designated as replacements. The rest of the non-Lumasaaba instruction schools in the district sample were neither assessed nor replaced. We indicated to the IP and USAID that this approach was neither appropriate to keeping the integrity of a random sample nor conducive to comparing SHRP schools to the average public school in Uganda. First, replacing sample schools with hand-picked replacements creates problems with the sample balance. Second, the aim of the evaluation is to assess reading ability of learners in English and local language. While it is not possible to test them in the local language (Lumasaaba, in this case) in schools that do not teach in Lumasaaba, it would still have been possible to test student's performance in English. As such, NORC's Evaluation Expert urged SHRP staff to conduct the EGRA in English in these schools as soon as we learned of the situation. However, the SHRP team did not comply with this request in a timely manner. Therefore, NORC decided that the impact analysis will need to exclude Mbale district altogether.
4. **Result 1: Manafwa district is encountering a serious crisis created by teacher transfers in the region.** We learned during field observations that most of the teachers trained by SHRP in January 2014 in this region have been transferred to other schools: four of the treatment schools visited by our local staff did not have a trained PI teacher, because s/he had been transferred. It will be critical to have information about the whereabouts of teachers trained by SHRP, since transfers of trained teachers away from treatment schools will have a severe effect on the impact evaluation. If these teachers end up at control schools, the impacts will be even more skewed. We will work with the IP and through our performance evaluation to try and capture the movement of trained teachers between schools.

#### **October 2014**

- I. **Result 1: Possible contamination in the control group.** In October, we learned from RTI that Mango Tree Project was working in Otuke, a control district for the SHRP evaluation, and providing their literacy intervention to two control schools in the SHRP sample. This occurred despite careful coordination between Mango Tree and SHRP. The schools in question were replaced for others, however the replacement schools have no baseline and their usefulness is limited. In addition, it is possible that some contamination has already occurred as we do not have any type of control over Mango Tree activities in the district.

2. **Result 1: Sample size adjustments between rounds.** The IP has changed sample sizes of each cohort between rounds. For the first cohort of students (Cluster 1), the February 2013 baseline included 280 schools to allow for analysis of 3 treatment arms, controls in treatment districts, and controls in comparison districts. However, the Cluster 1 Round 2 data collection conducted in October 2013, RTI collected data only in a subsample of treatment schools (168 of the 280), since a decision was made to only focus on one (and not 3) treatment. For Cluster 1, Round 3 in October 2014, however, RTI reverted back to data collection from 280 schools to account for that fact that the 3 treatment arms were implemented in the second year. This use of unbalanced panels does not preclude us from conducting a rigorous evaluation; however, it makes the process less transparent and prevents us from having measurements year by year without loss of information and precision. A similar change happened for Cluster 2. In this case, RTI requested NORC to calculate a sample size large enough to be able to analyze results at the district level. At baseline, in February 2014, data was collected from enough number of schools to calculate impact at district level; however, in for the first follow up in October 2014, the IP decided that district level analysis was not of interest and reduced the data collection to a subsample of the original schools. Although we will not be able to say anything about impact at district level, if properly implemented, this change should not prevent us from analyzing results at language level. In general, NORC recommends following the original samples over time to produce a more streamlined and transparent process and more comparable results across year.

#### **April 2015**

1. **Result 1: The Cluster 3 sample that NORC received from RTI in February** included only 1 control CCT per language. The reason for this was that all other non-treated clusters in the Cluster 3 districts were slated to receive treatment in 2016 so that SHRP could meet its student targets. SHRP staff explained to NORC that, for various reasons (security situation, ongoing MoES programs, budget constraints, etc.), expanding to new districts with Cluster 3 languages was not an option.

NORC expressed concern about this sampling change. Namely, that having only 1 control CCT in each district poses a risk to the impact evaluation because if for any reason that control CCT is different (for instance if it is a wealthier area, or if local authorities are much worse or better than treatment CCTs, etc.), the treatment and control will not be sufficiently similar to support an RCT. As such, we requested that, at a minimum, RTI add at least one more control CCT per district (even though more than one CCT would have been preferable). We also requested that all schools in the added control CCT remain untouched for the period of the study.

Ultimately, after lengthy email discussions, it was agreed that RTI would add 1 more CCT in each Cluster 3 district. However, due to the limited number of schools in these districts and the need to reach a specified target of students, SHRP would intervene in all but the control schools in these CCTs in 2016.

All parties agreed to this solution, while acknowledging the risks of contamination in control CCTs where some schools receive the SHRP interventions. Contamination of control schools could potentially lead to underestimation of the impact of SHRP interventions.

2. **Result 2: In February, NORC learned from WorldEd that, at NARC's request, they were dropping the KAP2 schools from the mid-term data collection and beyond** because adequate approvals from NARC had not been garnered for the new districts and schools in KAP2. As a result, future data collections (midline and endline) would only include KAP1 schools.

This decision has implications for the impact evaluation. The KAP2 data collection included schools from 4 new districts, which will no longer be part of the impact evaluation. Moreover, it also included boarding schools from the original KAPI sample that could not be included because of missing consent forms during KAPI data collection. Without the boarding schools in the KAP2 school sample, we will not be able to assess the impact of Result 2 activities on boarding schools.

In the absence of the KAP2 sample, the impact evaluation will only be able to assess the impact of the Result 2 interventions in secondary day schools and not secondary boarding schools. Indeed, although KAPI data collection included boarding schools, the boarding school sample visited during KAPI was not representative of that population. The boarding school students interviewed were not a random sample of students since only those with consent forms or those who were over 18 and could give consent themselves were included. The sample of schools was also not representative. As shown in the table below, partly boarding schools make up 39% of the population of schools and full boarding schools make up 15% of the population of schools; however in our KAPI data, they make up 20% and 3.7% of the sample of secondary schools, respectively. Given that neither the secondary school nor the student samples are representative, we will not analyze post-primary boarding schools, and will restrict our analysis of SHRP impacts on secondary schools to the day schools only.

|                 | <b>Population</b> | <b>KAP I DATA</b> |
|-----------------|-------------------|-------------------|
| Day Schools     | 46%               | 75.80%            |
| Partly boarding | 39%               | 20.40%            |
| Full boarding   | 15%               | 3.70%             |

3.