USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM IMPACT AND PERFORMANCE EVALUATION

Data Quality Assessment Report

PRESENTED BY:
NORC at the
University of Chicago
4350 East-West Highway
8th Floor
Bethesda, MD 20814

JUNE, 2013

NORC at the UNIVERSITY of CHICAGO

Table of Contents

Introduction 1

I. Evaluation Design 3

Sampling 3

Result 1 3

Result 2 4

Data Collection Plan 4

II. Data Collection Instruments 5

Result 1: Literacy 5

EGRA tool 5

Learner context questionnaire 6

Teacher/Head teacher surveys, classroom observation tool and school inventory tool 6

Result 2: HIV/AIDS 6

KAP (Knowledge, Attitudes and Practices) tool 6

III. Training 8

Enumerator and Supervisor Training Manuals 8

Result 1 8

Result 2 9

Enumerator and Supervisor Training Observation 9

Result 1 9

Result 2 10

Post-Training Pilot Observation 11

Result 1 11

Result 2 11

IV. Data Collection (Result 1 only) 12

Data Collection Roll-Out 12

Field Quality Control Procedures 13

V. Data Entry 14

Tangerine Programming for EGRA 14

Data Entry Template for KAP tool 15

VI. Datasets 15

Result 1 15

Result 2 15

Annex – Data Quality Assessment Plan 16

Introduction

One of the responsibilities of the Performance and Impact Evaluation (P&IE) Contractor for the School Health and Reading Program involves conducting Data Quality Assessments (DQA) for data being collected by the Implementing Partner (IP) that is directly relevant to the evaluation of the projects. As outlined in NORC's contract: "Throughout the five-year Literacy and Health Education Program, NORC shall collaborate with Literacy and the Ministry of Education and Sports (MOES) to ensure that the quality of data collected is adequate to enable rigorous external performance and impact evaluation. During the Evaluation Design process, NORC shall work collaboratively with program stakeholders to establish plans for the design of data collection tools and systems. These tools and systems, to be implemented as part of the Literacy and Health Education Program[1], must be acceptable to the P&IE contractor for performance and impact evaluation purposes. The P&IE contractor shall conduct and document Data Quality Assessments (DQAs) for baseline, mid-term, and final data collection, processing and analysis. Throughout the five-year program, the P&IE contractor shall provide technical expertise and work with program stakeholders to promote the highest possible quality of data."

Methodology

For the USAID/Uganda SHRP P&IE, NORC is taking a systematic approach to Data Quality Assessment by ensuring that all systems, protocols, tools and data collection activities are developed to ensure the highest possible data quality. Data Quality Assessment is, therefore, a process that includes both formal and informal on-going review of the sample design and procedures, data collection instruments, field procedures and practices, quality control protocols, and reporting mechanisms that are prerequisites for rigorous external performance and impact evaluation. As such, NORC and the Implementing Partner (IP) are in regular communication regarding all the steps involved in the evaluation.

This Data Quality Assessment report describes NORC's observations and comments on the tools, materials, training and data collection activities developed by the Implementing Partner for the evaluation of the SHRP intervention. It covers findings from our *ex-ante* data quality review activities for the first round of Result 1 (EGRA and associated tools) and Result 2 (KAP) data collection. Although EGRA data collection was completed in March 2013, NORC only obtained access to the complete EGRA dataset in June 2013; as such, we have not had the opportunity to conduct a quality review of that data. Hence, our Data Quality Assessment Report does not include *ex-post* a data quality assessments for the EGRA data. Finally, at the time this report was written, the data collection for Result 2 had just begun. The report therefore does not cover the data quality review activity associated with the roll-out of data collection for Result 2.

[1] Now known as the School Health and Readong Program (SHRP).

The organization of this report follows the components of NORC's DQA Plan, which appears in Annex I of this report; this DQA plan/checklist was also presented in NORC's January 2013 Performance and Impact Evaluation Design Report and was shared with the IP to inform them of NORC's DQA process. This DQA report presents key achievements, areas of concerns and recommendations made by NORC for each of the following aspects of the evaluation:

- Section I: Evaluation Design – Sampling and Data Collection Plan
- Section II: Data Collection Instruments – EGRA and associated tools, and KAP instrument
- Section III: Training – Training Manuals, Training and Post-Training Pilot Observations
- Section IV: Data Collection Period – Data Collection Roll-Out and Field Quality Control Procedures
- Section V: Data Entry Systems – Tangerine for EGRA
- Section VI: Datasets

I. Evaluation Design

This section describes activities and discussions held with the IP regarding the evaluation design. These discussions began during NORC's initial trip to Uganda in October 2012 and continued subsequently through phone conversations and emails. NORC and the IP worked closely together to finalize the sampling design and procedures and the data collection plan in order to ensure that they meet the needs of the rigorous impact evaluation proposed for the SHRP intervention in NORC's Design Report dated January 31, 2013.

Sampling

Result 1

In November 2012, through a series of conference calls and email communications, NORC's Statistician and Evaluation Expert worked on the estimation of the sample sizes in order to make sure that the data are sufficient for the desired level of precision and power. These power calculations and sample size estimates, which were shared with USAID in November 2012, took into consideration the Implementing Partner's (IP) decisions to move forward with three treatment arms, and desire to measure impacts for 4 language subgroups.

In addition, NORC requested two control groups – non-treated schools within treatment districts and schools in comparison districts – to ensure that the sample design would be adequate for assessing the impact of the intervention at both the school and district levels. As such, the sampling took into account the fact that there would be 5 groups and 4 languages, or a total of 20 cells of schools and students. We estimated the sample required to detect a double-difference measure of impact of magnitude D = 0.20 with a power of 90%. Based on these calculations for each of 20 cells, we estimated that it was necessary to have 14 schools, each with 30 P1 students, for a total of 420 students, who will be followed over subsequent years. With 20 cells (3 arms and 2 controls, and 4 language subgroups per group), the total sample size required amounts to 8,400 student in 280 schools; i.e. 8,400 P1 students at baseline in 2013 to be followed in November 2013, November 2014, and November 2015). Of the 280 schools, 168 (5,040 students) would constitute the treatment group, and 112 schools (3,360) would be controls.

Based on these estimations, the IP selected CCTs and assigned them to the different arms and randomly selected 168 treatment schools for the evaluation sample, from 410 randomly selected intervention schools. Control schools within the treatment districts were selected from the schools that were not selected for the intervention by NORC. NORC's evaluation expert also worked with the SHRP M&E leader to select comparison districts and sample schools within those districts (unfortunately the data available to perform the matching for selecting the comparison districts was extremely scarce).

NORC deemed the sample design and size presented in NORC's January 2013 Design Report sufficient for the conduct of the impact evaluation design, which was also presented in the same report. We requested that the IP inform NORC of any deviations from this sampling plan prior to implementing them, since such changes would have implications for the evaluation design approved by USAID.

Result 2

NORC staff also worked with the IP Result 2 team to estimate the sample for the School Health Program, which is comprised of three cells: treatment, controls within the treatment district, and controls in comparison districts. A school sample of 234 schools was deemed adequate to detect effects (double-difference impacts) of magnitude D = 0.20 with high power (90%). Out of 150 intervention schools, 78 schools were selected for assessment. A similar number of schools were selected for the in-district control and out-district control groups. Initially, all 50 intervention post-primary schools were included in the evaluation, as well as a similar number of post-primary schools for the in-district and out-district control groups. A total of 30 students per school were selected for data collection. Therefore, for the Health program, we anticipated that a total of 7,020 primary school students and 4,500 post primary school students would be included for data collection.

The Result 2 intervention underwent some redesign following the involvement of PEPFAR; as a result, a decision was made to focus the school health intervention on large schools (more than 150 students in P4-P7). As a result, NORC's Evaluation Expert reviewed and adjusted the original sample selection for the intervention. Furthermore, due to delays in implementation, it was not possible to distribute consent forms to students in time to obtain parental consent for data collection among minors in boarding schools, who do not live at home; therefore, NORC and the IP came to an agreement that boarding schools and partial boarding would be dropped from the sample, thereby limiting the data collection to day schools only[2]. The original sample of 50 post-secondary schools per arm was thus reduced to about half that number (the final number will only be known after the end of data collection given that the school's boarding status is only known at the time of visit by the enumeration team).This limitation has implications for the external validity of the evaluation design, an issue that NORC has discussed with USAID.

Data Collection Plan

RTI is responsible for all data collection related to the impact evaluation. As such, In December 2012 and January 2103, NORC's evaluation expert worked with the Result 1 and Result 2 Teams and M&E team to ensure that the timing and coverage of the data collection are adequate for the impact evaluations. Final data collection plans that the IP and NORC developed meet many, but not all, of the original objectives (as outlined in NORC's proposal) of the impact evaluation, and fits within the budget and logistical constraints of SHRP. The timing and coverage (grades/cohorts and numbers of students per grade/cohort) of the EGRA and KAP surveys over the five-year life of SHRP were presented in detail in NORC's Evaluation Design report, submitted to USAID/Ghana in January 2013. Excerpts from the Design Report are also presented in Annex I of the Semi-Annual Report.

The baseline data collection for the EGRA baseline was carried out at the beginning of the academic year (February 2013), as planned, and before the intervention took place. Data collection for Result 2 started on June 12, shortly after the beginning of the second term, and continues as we write this report. We note here that part of the Result 2 intervention, the PIASCY training for teachers, was conducted prior to start of data collection, in May 2013. This

[2] The IP is collecting some data from day learners and boarding learners who are aged 18 and over in partial boarding schools at request of MOES. NORC will explore the usefulness of using these data but anticipates those students are not representative of learners in partial boarding schools.

is not ideal and could potentially contaminate the base line data; however we do not anticipate a major impact that could compromise the IE.

II. Data Collection Instruments

NORC also worked closely with the SHRP M&E team to ensure that all instruments follow closely from the evaluation hypotheses and indicators necessary for the impact evaluation. Towards this end, NORC's subject matter experts and evaluation experts reviewed all instruments and provided extensive feedback to ensure that in addition to outcome indicators related to reading skills and HIV prevention knowledge, the data collection effort includes information on covariates (student's socioeconomic characteristics, parent education, home literacy environment, etc.) that need to be controlled for in the evaluation model.

Result 1: Literacy

EGRA tool

Starting in December 2012, subject matter experts from NORC and the IP worked on several iterations of the EGRA tool to finalize the subtasks to be included in the assessment tool. In addition to the subtasks that were already included in the original EGRA draft received by NORC, some of the recommendations that NORC made were to:

- **Add a receptive vocabulary sub-task:** Since vocabulary is considered a key component of early literacy development by the National Reading Panel (one of the main conceptual bases of EGRA) and is especially important for children learning in bilingual and low-income environments, NORC suggested adding a vocabulary sub-task. This has been done with EGRA tools used in other countries, such as the Mozambique test used by World Bank which included a receptive vocabulary sub-task. *The IP and NORC agreed that gathering this information would provide insight into how vocabulary contributes to literacy acquisition and would also allow for analysis of the relationship between literacy in the mother tongue and the second language, as schools increase their instruction in English, and therefore decided to add a receptive vocabulary sub-task in the English EGRA tool.*
- **Add an expressive vocabulary sub-task, or an assessment of the words that children can produce:** NORC and the IP discussed the possibility of adding such a test; one idea was to ask students to produce synonyms of certain words they would encounter in the reading passage. *However, expressive vocabulary is more sophisticated than receptive vocabulary, so it is likely that in the early grades, scores would be closer to zero than on a receptive assessment. NORC and RTI thus agreed not to include the expressive vocabulary sub-task in the current version of EGRA, but will discuss the possible inclusion of this sub-task in future rounds of data collection.*
- **Include a writing sub-task:** Although writing is a key outcome of literacy instruction, most assessments do not include a writing sub-task, likely because the National Reading Panel did not address writing in its seminal report, and because it is difficult to assess. Nonetheless, some EGRAs have included writing assessments in the form of a short dictation, which specifically analyze performance of spelling, orientation of text, spacing, capitalization, and punctuation. *However, the major challenge for such a sub-task is in analyzing the text produced. As such, similarly to the expressive vocabulary sub-task, NORC and the IP agreed*

not to include this into the current EGRA tool since most P1 students would likely score zero and since it is quite difficult to implement. The IP also suggested that a qualitative spelling task that would allow us to assess a child's ability to apply their emerging letter knowledge with phonological awareness might be better. SHRP may pilot such a spelling task in the future.

Learner context questionnaire

NORC reviewed an initial draft of the learner context questionnaire, which consists of a series of questions on student background and home literacy environment and is administered to students directly following the EGRA assessment. NORC recommended collecting additional data on both home and school language and literacy practices, specifically including adaptation of some questions that have been used on the Progress in International Reading Literacy Study (PIRLS). *The IP shared this interest in collecting valid data on student's reading environment and worked with their Ugandan experts to refine the learner context tool in order to capture this information.* These data, which will serve as covariates in the impact analysis, will help us conduct a more precise program analysis as well as examine some of the key barriers to literacy acquisition, both in terms of constraints at home and in school (as literacy is strongly influenced by factors outside of school).

Teacher/Head teacher surveys, classroom observation tool and school inventory tool

Finally, NORC reviewed the teacher/head teacher surveys, classroom observation tool and school inventory tool. Overall, these tools required minor adjustments and most of NORC's comments concerned the wording or response options for very specific questions.

The main concern that NORC raised with these tools was the need to ensure that they included unique identifiers that would allow their respective datasets to be merged with one another, i.e. unique school IDs and unique teacher IDs. *The IP responded to these concerns by making sure that schools and teachers have unique IDs that would allow for merging of all datasets.*

Result 2: HIV/AIDS

KAP (Knowledge, Attitudes and Practices) tool

The Result 2 instrument, which covers Knowledge, Attitudes and Practices regarding HIV/AIDS, was provided to NORC for review and comment in early January 2013. NORC's comments on the first two review versions included concerns about human subjects protections, consent, mode of administration, length (duration) of questionnaire, appropriateness of questions given respondent age, relevance of questionnaire items to the program key indicators, comprehensibility of terms, and format of questionnaire items.

NORC's recommendations were provided via email and phone and included, but were not limited to, the following:

- **Consent and IRB:** To meet basic human subject protections requirements, the instrument requires a consent and/or assent procedure for respondents and for respondent parents if the respondent is a minor, according to Uganda's legal age of majority. Uganda's National Guidelines for Research involving Humans as Research Participants, which NORC shared with the Implementing Partner (IP) in January, 2013, applies to " (1) all research involving

humans as research participants in Uganda, including research in social sciences and humanities."[3] NORC inquired about the IP's planned IRB submission schedule. Although the IP indicated a plan to request exemption, NORC advised consultation with the IP's US-based IRB to determine application steps, as NORC's experience with protection of human subjects indicated a full board review would be required by both local Uganda IRB and sponsoring entity IRB (in this case the IP), given the vulnerable population and sensitivity of the questionnaire items. The delays in developing the questionnaire and in seeking appropriate IRB approval delayed the data collection and forced a change in the sample for Result 2. More specifically, IRB requires parental consent for minors' participation in the survey. The parental consent form was not available before the end of the February-May school term, as initially planned, which would have allowed time for boarding students to bring consent forms home for signature. Because the parental consent form was not available until after students had returned to school for the June school term, boarding and partial boarding school students were not able to obtain parental consent, and these schools had to be dropped from the sample. Students attending day schools were able to bring parental consent forms home for signature after the start of the June school term, therefore day school students remained in the sample.

To the extent possible, NORC's Survey Specialist provided advice to the IP regarding IRB submission, reiterating the urgency of early submission to the local Uganda IRB and the IP's IRB, and the importance of obtaining IRB approval prior to any contact with subjects for pretesting or other data collection. The IP had intended to carry out a pretest in one or more schools during the week of 4 February, prior to Uganda IRB approval and prior to IRB application submission to IP's IRB, but at NORC's recommendation, abandoned this plan until appropriate approvals had been obtained.

- **Mode of administration:** Self-administration was selected for privacy and cost concerns. NORC expressed concerns about privacy for all respondents, and about formatting and question wording for ease of use.

- **Questionnaire content**
  - Length of questionnaire: NORC recommended carrying out internal timing exercises to better estimate survey duration, which appeared considerably longer than the estimated 30 minutes.
  - NORC expressed concern about the sensitive nature of some questions, including questions regarding sexual behavior, being administered to students in grades P4-P7. Some students could be as young as 8-9 years old. NORC recommended cognitive testing to be done before a pre-test with the real population.
  - NORC questioned whether, if the tools were not translated into local languages, students in the earlier grades (e.g. P4-P5) would understand enough English to respond.
  - NORC expressed concern about the use of technical and potentially unfamiliar terms and concepts related to sexual behavior.

[3] Uganda, 2007. National Guidelines for Research involving Humans as research Participants. Uganda National Council for Science and Technology. http://www.uncst.go.ug/dmdocuments/Guideline,%20Human%20Subjects%20Guidelines%20Marc.pdf

- NORC recommended standardizing questionnaire items into questions, rather than a mix of statements and questions.

- **Training and Data Collection procedures**
  - NORC recommended ensuring that the training and related materials give detailed instructions to group facilitators and include extensive role-playing to ensure standardized data collection protocols are followed.
  - To ensure high quality data, NORC requested that its DQA plan be shared with the Result 2 team to provide the team with a standard checklist of tasks and items NORC requires in carrying out data collection.
  - NORC recommended extensive pre-testing of the questionnaire and data collection procedures prior to training.

III. Training

This section discusses NORC's comments on the training manuals as well as observations of the enumerator trainings and of the post-training pilots for Result 1 and Result 2.

Enumerator and Supervisor Training Manuals

Providing enumerators and supervisors with a comprehensive training manual is essential in ensuring that procedures and administration of the interviews are standardized across all field staff. At a minimum, NORC wanted to ensure that the manuals covered interviewer techniques, confidentiality and consent, organization of fieldwork and sample requirements, and detailed description of the data collection instruments.

Result 1

Prior to the start of enumerator and supervisor training in February 2013, NORC had requested copies of the training manuals. However, only a few days before training was scheduled to begin, the IP informed NORC that there was no plan to create an enumerator manual. NORC considered this a significant weakness in training methodology and discussed our concerns with USAID. Upon USAID inquiring about this issue, the IP informed NORC and USAID that they would be providing a manual in the form of "the paper instruments themselves and a kit of relevant handouts" (Email from Saeeda Prew to Joseph Mwangi, 2/9/2013)

NORC did, however, receive a copy of the supervisor manual a few days before training. However, at that time, the manual did not appear sufficient for the training. NORC staff raised many questions and flagged numerous inconsistencies throughout the draft, which were then shared with the IP.

The manuals were only finalized after the end of enumerator/supervisor training. The revised version of the field work manual showed major improvements upon previous versions received by NORC and appeared adequate for the continuing support of the field teams.

In addition to the field work manual, RTI created one-pagers and job aids that provided excellent information in a concise way on sampling procedures, troubleshooting protocols for Tangerine and administration of EGRA.

Result 2

Similarly to Result 1, NORC requested copies of enumerator and supervisor training manuals in advance of the training, which was scheduled to take place in late May and early June 2013. NORC received a first draft training manual on May 21. NORC provided comments to the first and subsequent drafts of the training and field manuals via email and phone and included, but were not limited to, the following:

- Need for more specific information on logistics on advance contact with schools, document management for surveys and tracking forms, security procedures.
- Need for more specific and detailed instructions and practice for sampling procedures.
- Recommendation for clear protocols for 'at risk' or distressed respondents.
- Suggestion for more detailed training manual including module descriptions, chronology of training activities and a daily agenda.

The final versions of the field and training manuals were provided to NORC on June 1 and June 3, respectively, after the supervisor training. The final versions of both manuals showed significant improvement over early drafts and appeared sufficient for training and fielding purposes. NORC recommended additional in- person training and testing of supervisors and enumerators on carrying out sampling procedures, which can be a particularly difficult task to operationalize in the field.

Enumerator and Supervisor Training Observation

Result 1

NORC's Senior Literacy Evaluator and Resident Evaluation Manager observed the EGRA enumerator and supervisor training in Kampala from February 11th to February 16th 2013. Overall the training went well with few minor issues. Below we detail our main observations.

Key Achievements:

- **Well-organized Training:** The training was attended by about 100 enumerators. Individual instructors taught separate sections of the training and, used effective, methodical and systematic training strategies that incorporated extensive modeling for each of the subtasks of EGRA. Instructors appropriately answered enumerators' questions and gave enumerators multiple opportunities to practice and debrief after training sessions. The only subtest for which no modeling was done was the vocabulary subtest, and NORC discussed this with RTI trainers.
- **Systematic testing of enumerators:** Three inter-rater reliability tests were done throughout the training whereby one of the trainers modeled being a student while the enumerators filled out the assessment tool. These tests were done on paper and on Tangerine and the scores from these tests helped the IP determine who to keep for the final field teams, and dismiss others that were not up to standard midway through the training. The highest scoring enumerators were selected to be supervisors.

- **Good training on Tangerine:** After the pilot, each enumerator was assigned a tablet and asked to practice with it in order to identify any challenges they encountered. The IP went through all questions raised in order to make sure that enumerators had a good understanding of the electronic tool. Enumerators were also given time to train on Tangerine using role-playing exercises whereby one enumerator acted as an enumerator, a second one as a pupil and a third one as an observer.

Areas of concern:

- **Discrepancies between paper EGRA and Tangerine**: NORC observed that the paper EGRA questionnaire had some discrepancies with the online Tangerine tool and minor formatting/wording issues, mainly:
  - Teacher name item: NORC suggested using a teacher ID instead of using the teacher's name as this would likely lead to errors in spelling and complicate the data cleaning tasks.
  - Stopwatch: the stopwatch provided to enumerators counted up while the Tangerine timer counted down. Since enumerators are asked to record the time remaining on the stopwatch at the end of a task, this forced them to conduct calculations while in the field. Errors were noticed during training when enumerators would write times above 1 minute which is not possible since students are given 1 minute to read letters/words. NORC suggested either altering the stopwatch so that it would count down from 1 minute, or altering instructions so that enumerators report time used rather than time remaining.
  - Early stop rule: The test did not differentiate between a child who attempts nothing from a child who attempts but gets the first row of a subtask wrong. In both situations, the child gets the same score. Although the reasons for non-response might be different from the reasons for all incorrect answers.
- **Workload for supervisors:** It was also noted that the workload for supervisors seemed quite heavy. Supervisors are tasked with contacting schools, arranging all interviews including selecting the sample, conducting the teacher interview and school inventory, and uploading the data at the end of the day.

NORC shared these observations and recommendations with the IP immediately after the training.

Result 2

NORC's Evaluation Expert and Resident Evaluation Manager observed the interviewer training for Result 2 during the week of June 4- 8, 2013. NORC's Senior Survey Director was also scheduled to observe training but had to cancel her trip right before the start of the training due to a death in the family.

The Result 2 interviewer training went well. The training included 65 participants. It allowed sufficient time for each presentation, which was followed by a 30-40 minute reflection session that offered enumerators ample time to ask questions. The participants also appeared engaged throughout the training, reacting to presentations and sharing their own experiences. They contributed greatly to improving the survey tools as the training unfolded.

The training included group exercises, which were well supervised by the trainers. These group sessions proved crucial as they generated important issues that had not been foreseen by the Principal Investigators and enabled the IP to improve their survey tools.

Post-Training Pilot Observation

The in-class trainings for Result 1 and Result 2 were both followed by post-training pilots during which enumerators were given the opportunity to practice first-hand the skills they learned in class and for the data collection teams to resolve any potential field challenges before the start of the actual data collection. As such, it is expected that some issues arise during a pilot.

Result 1

NORC's Senior Literacy Evaluator and Resident Evaluation Manager attended the post-training pilot, following the in-class EGRA training. Below are observations as noted by NORC staff, all of which were communicated to the IP staff:

- **Duration:** Some of the assessments extended well beyond the 30 minute guideline although NORC understands that administration time typically declines within the first few days of data collection starting.
- **Prompting:** NORC observed several instances when children were waiting for enumerators to prompt them through tasks. In some cases, the enumerator reminded the student that he had to continue on his own, but in others, we heard the enumerators begin to prompt, for example, for each letter in the letter sounds sub-task.
- **Repeating Vocabulary:** On the vocabulary section, NORC observed that some enumerators were repeating the vocabulary words several times, prompting the student to point. Since the process of the vocabulary section is different from the others, NORC suggested modeling this for enumerators again (and giving them a bit more practice).
- **Pupil Stimuli:** NORC observed a few enumerators using the EGRA tool, rather than the pupil stimuli, to ask the children to read, making it difficult for students to name the letter sounds.
- **Assessment during Breaks:** Some assessments took place during the student break time; in addition to missing out on play, some of the children likely were missing their snacks. NORC suggested coordinating with the school to make sure those children have time to eat, if assessing during break time is unavoidable.
- **Instructions:** Some enumerators read instructions in both English and then Luganda, which is time consuming. NORC recommended trainers clarify to enumerators when to shift from one language to the other (rather than reading all instructions).

Result 2

NORC's Evaluation Expert, Resident Evaluation Manager, and Senior HIV/AIDS Evaluator attended the post-training pilot, following the in-class Result 2 training. Below are observations as noted by NORC staff, all of which were communicated to the IP staff:

- **Sample:** Enumerators followed the rules to create the random sample accordingly to the instructions received during training. This was one of the most difficult tasks they had to carry out and no teams reported problems.

- **Instructions:** Enumerators took their time to explain clearly to the learners the selection process, the right to refuse participation, and the rules of privacy. Students seem to feel comfortable through the process.
- **Survey**: Enumerators divided the learners in groups by age and conducted the survey. Students completed the questions with no apparent problems. An initial inspection of the data confirmed that they were able to follow instructions and complete the survey as expected
- **Duration**: For some groups the survey took more than one hour. It seemed too long for young learners and, therefore, some questions that seemed redundant were eliminated during the working meeting that took place the day following the pilot exercise.

IV. Data Collection (Result 1 only)

This section discusses NORC's comments based on observations of data collection as well as the adequacy of the procedures that were put into place to ensure quality control. Because Result 2 data collection has just started at the time of this report, we are only including observations for Result 1.

Data Collection Roll-Out

NORC's Resident Evaluation Manager conducted the following field visits to observe the EGRA data collection:

Date	Language	District
5th March 2013	Runyankore/Rukiga	Bushenyi
6th March 2013	Runyankore/Rukiga	Kabale
11th March 2013	Luganda	Wakiso
15th March 2013	Ateso	Ngora

Overall, data collection went as planned. Sampling did not pose a major problem and enumerators administered the EGRA without major difficulties. As expected, administration time of EGRA decreased significantly within the first few days of data collection, from 35-40 minutes at the beginning to 13-15 minutes per student.

Key Achievements:

- **Collaboration amongst team members:** NORC observed a high degree of collaboration among the team members. The enumerators, school level supervisors, and some Data Quality Supervisors (DQS) worked together to ensure they accomplished the tasks for the day. Supervisors helped enumerators carry out EGRA assessments before learners left school, especially during times when teams experienced delays in the morning. In fact, no team failed to complete P1 interviews before learners left for home even though they only had 2-3 hours to conduct interviews for 30 learners. In addition, the data collectors supported the supervisors for pupil sample selection.

- **Excellent time management:** The teams set off very early in the morning to ensure they reached schools on time. Teams were always at the work stations before 8.30am, even though some teams were accommodated in nearby districts and driving through Ugandan roads is quite hectic.
- **Entry into school community:** Although some schools had not been informed about the EGRA activities ahead of time (see below for more details), the teams were able to do the required introductions that convinced the school administrators to allow them to work. They worked efficiently and calmly with school administrators to mobilize students and teachers quickly. They also managed field challenges well even without the help of their DQS in some instances.
- **Management of tablets:** the DQS did a good job of keeping custody of tablets and recharging them. It was rare to find tablets running out of battery before the interviews were done. The enumerators received fully recharged tablets from the DQS every morning.

Areas of concern:

- **Advance contact:** Some school administrators were not aware of the planned field work. For instance, this happened in Otuke district, where head teachers were not informed of EGRA activities ahead of time. In some instances, head teachers and school deputies were absent when field teams arrived. However, field teams overcame this problem and were able to conduct the necessary tests and interviews.
- **Scheduling:** EGRA teams should ensure that data collection activities are not scheduled on community market days, because teachers report that pupils stay at home to look after younger siblings, or accompany their parents to the market. These markets, organised at sub-county level, occur in all upcountry areas apart from Kampala. The District Education Officers are aware of these days in their localities, and can help schedule data collection activities on non-market days when student attendance tends to be higher.
- **Sampling:** In some schools, children as young as 3 or 4 years old might be grouped together with P1 pupils even though they are nursery learners. Field teams encountered some challenges in the identifying these pupils and, in some cases, the determination of which students were nursery learners was made by teachers. Since it is difficult to determine the ages of the students, it is possible that some students who are eligible for EGRA were in fact excluded from the study. Unfortunately, there is not much that can be done to solve this issue since students in Uganda often do not know their own age and many of them suffer from stunting which makes them look younger and smaller than their actual age.
- **Supervisor workload:** As was mentioned in the training section, the supervisors were overloaded with responsibilities. They were also doing teacher and head teacher interviews, which were not piloted during the training; hence, they were relying on their supervisor manual to carry out these interviews. This lack of familiarity with the interview content could potentially lead to quality issues.

Field Quality Control Procedures

In this section, we review the procedures that were put into place by the IP for quality control.

- **Organization of field teams:** Field teams are well organized and roles and responsibilities of each team member were well defined prior to start of data collection. Each team was comprised of four enumerators and one supervisor. There were also four Data Quality

Supervisors (DQS) for each of the language groups covered by the data collection. The team supervisors would report to the DQS who in turn would report to the Task Manager from the Center for Social Research.

- **Supervision of teams:** The communication between Data Quality Supervisors and the IP was good. The DQS were contacted on a daily basis for feedback such that it was possible for the IP to learn about challenges encountered by different teams and advise accordingly and in real time. The DQS would also link up with the different teams in the evenings for debriefing, discussion and troubleshooting of main challenges of the day. On the other hand, as was mentioned previously, the team supervisors' workload was quite heavy, making it difficult for them to supervise the enumerators since they were also conducting EGRA assessments.

V. Data Entry

Tangerine Programming for EGRA

The NORC team had the opportunity to test Tangerine, the electronic tool used for the EGRA data collection. NORC found that it was well programmed, skips were respected and soft and hard validation checks have been programmed and tested. For instance, validation ranges for age of student or number of students per classroom have been included.

Furthermore, Tangerine allows for generation of a unique student ID number, and was designed in such a way that the EGRA assessment was immediately followed by the learner context interview, thereby avoiding any issues with merging the two datasets later on. Instructions to the enumerator and questions to the students were formatted different on screen, allowing for easy administration of the tool.

Tangerine is an easy to use tool and enumerators did not encounter major issues with it. The integrated timer on the tablet is easier to use than the stopwatch needed with the paper instrument and the electronic device allows for easy correction when students make mistakes but self-correct immediately.

Below are the main issues noted by NORC during initial Tangerine testing:

- **Segmenting Task/Vocabulary Marking Change:** In Tangerine, both the segmenting and vocabulary tasks asked enumerators to mark the correct answer, rather than incorrect. This could lead to confusion on two counts. First, they were trained to mark them incorrect on paper. Second, it requires them to shift gears from sub-test to sub-test since they mark incorrect answers in other sub-tests of the Tangerine EGRA. NORC proposed three possible solutions: 1) change the sub-tasks on the paper version to match Tangerine and also provide additional training on this task; 2) change the sub-tests so that the marking is consistent with the others; and/or 3) group these two sub-tasks together either at beginning or end of assessment so that enumerator only has to shift in his/her thinking of how to mark once.

- **Learner context questionnaire only in LL:** It appeared that the learner context questionnaire questions were only offered in local language. Earlier in the training, NORC

had inquired about why it was necessary to provide the instructions for each sub-test in English, rather than just providing them in the local language. We were told that some students might actually have better mastery of oral English than the local language (for example, in the case that their parents actually speak a third language at home). If this is likely, the learner context questions should also be offered in English, so that we can be sure that learners fully understand the questions.

- **Auto-stop when all is wrong in first line/self-correct:** Normally, if a child reads all the letter-sounds incorrectly in the first line, the auto-stop happens. However, there was no way to override auto-stop in the case that a child self-corrected on the final letter in the first line.
- **Data when a child decides after starting that s/he doesn't want to complete test:** When a child opts out of the test after administration begins, there should be a mechanism to document this.
- **Oral passage auto-stop difference from paper to tablet:** The auto-stop on the paper version occurs after 15 words, but on the tablet version auto-stop occurs after 13 words.
- **Vocabulary: replacement of word "desk" with "clothes":** It appeared that the word "desk" from the paper vocabulary test was replaced with the word "clothes" because some students don't have a desk. If that is the case, it should be changed on the paper version as well.
- **Segmenting task: allowed to enter both correct phoneme and "none correct":** The segmenting task allowed the enumerator to select that a child had identified one phoneme correct (such as /i/ in if) while simultaneously selecting "none correct".

Data Entry Template for KAP tool

NORC has not received the KAP data entry template for review prior to writing this report. Our review will be included in the next Data Quality Assessment Report.

VI. Datasets

Result 1

The EGRA data collection was completed in March 2013. However, NORC only obtained access to the complete EGRA dataset in June 2013; as such, we have not had the opportunity to conduct a quality review of that data. Hence, our Data Quality Assessment Report does not include ex-post data quality assessments for the EGRA data.

Result 2

Result 2 data collection is ongoing as we write this report. NORC will request the dataset from the IP once it is available.

Annex I – Data Quality Assessment Plan

The following data collection instruments, documents and activities will be reviewed as part of NORC's DQA. Note that this list may not be exhaustive. This list applies to both Result 1 (literacy) and Result 2 (health).

Instrument, Document or Activity	DQA checklist
Evaluation Design	
Sampling: sample size, selection (randomization process) of intervention schools, matching of comparison districts	▪ Sample size is sufficient for desired level of precision and power ▪ Sample design is adequate for assessing impact of intervention at school and district level (randomization of intervention schools is carried out correctly, selection of matched comparison districts is done using adequate statistical matching methods and with best matching data available)
Data collection plan: timing of data collection, selection of intervention/control in-district/control out-district schools for data collection	▪ Timing of data collection is appropriate for impact evaluation (baseline is prior to intervention, follow-ups are at regular intervals, endline is post-intervention; data collections happen at either beginning or end of school year) ▪ Data collection team has allocated adequate human and material resources to carry out collection within specified time period ▪ Timing of data collection and data delivery allow for annual impact evaluation and impact evaluation of 4-year SHRP within project deadlines ▪ IRB permissions have been obtained
Data Collection Instruments	
Result 1 (Literacy) EGRA tool Learner environment questionnaire Teacher/Head teacher surveys Classroom observation tool School survey CCT Monitoring Tool	▪ All tools capture information needed to calculate key indicators for performance and impact evaluation ▪ Questionnaires are ordered logically and structured to facilitate comprehension by respondents and use by data collectors ▪ Questionnaires are piloted and revised accordingly (adapted to Ugandan context ▪ Questionnaires include proper geo-referencing information and allow for easy merging of data: ✓ Questionnaires include case id, class id school id that
Result 2 (Health) KAP survey	

Instrument, Document or Activity	DQA checklist
	are standard across different instruments ✓ Questionnaires are designed for easy merging of longitudinal data ▪ Questionnaires allow capture of interviewer id, supervisor id, data enterer id (if applicable) ▪ Observation/review of pre-test results for tool development where applicable ▪ If possible, check that translations have been done correctly (may not be possible due to lack of staff with knowledge of local languages)
Training and Data Collection Period	
Enumerator and Supervisor Training Manuals	▪ Cover at a minimum: project description, basic interviewer techniques, confidentiality and consent, organization of fieldwork and sample requirements, tracking of sample, detailed description of data collection tools
Field Quality Control Procedures	▪ Organization of field teams provides adequate supervision/management ▪ Validation (back-check) procedures are included ▪ Field procedures includes proper tracking of sample and response rates: documentation of in-field sampling procedures (e.g. random selection of students within each classroom), proper use of disposition codes ▪ Interviewer feedback process is documented and used ▪ Proper use of unique ID codes for schools, students, etc, to allow for triangulation of data ▪ Mechanisms for reporting to Central Office/Level of supervision from Central Office is adequate ▪ Schedule for validation, tracking and interviewer feedback reports is clear
Enumerator/Supervisor training for Result 1 Enumerator/Supervisor training for Result 2	▪ Observations of trainings by NORC expert(s) ▪ Trainings are well-organized and trainers are well-prepared ▪ Role-playing and other practice exercises are included ▪ Interviewers demonstrate mastery of concepts and procedures through formal, documented assessment

Instrument, Document or Activity	DQA checklist
Post-training pilot	▪ Observation of post-training pilot by NORC expert ▪ Enumerators are well-prepared for field period ▪ Participation of NORC in post-pilot debriefings to gather lessons learned ▪ Debriefing lessons are implemented and communicated to field team prior to field period
Data Collection Field Report	▪ Data collection process and issues encountered during field period are documented (organization and structure of field teams, dates of field report, final response rates, reasons for non-response, challenges encountered and solutions)
Data Entry	
Tangerine data entry template for all applicable tools	▪ Paper instruments are reviewed for completeness prior to data entry ▪ Procedures for handling missing data are clearly specified and standardized across instruments and rounds of data collection ▪ Data entry templates match paper instruments ▪ Skips are respected ▪ For paper instruments, data is entered using double data entry method ▪ Soft and hard validation checks are programmed and tested prior to training for electronic instruments, and prior to data entry for paper instruments ▪ Upload are made available for review by data quality reviewer on a real-time basis (as they are uploaded)
Data entry templates for other surveys using paper questionnaires	
Datasets	

Instrument, Document or Activity	DQA checklist
EGRA data Learning environment data Teacher/Head teacher survey data Classroom observation data School survey data CCT monitoring data	▪ Datasets are well constructed: variable names, variable labels, value labels are included and correctly specified ▪ Datasets can be easily merged if needed (using of unique codes for merging across different datasets) ▪ Reserve codes are correctly used (including specification of legal skips and missing values) ▪ Proportion of missing values is within acceptable range ▪ Level of precision is adequate ▪ Data is internally consistent (See Annex 1 for more information on Guidelines for Data Cleaning and Assessment) ▪ Test datasets and interim datasets (pilot dataset, first 100 cases) are produced and delivered for DQA with adequate time for incorporating corrections prior to main data entry (for paper instruments)
Note: For all instruments, DQA covers pilot datasets, interim datasets (real-time uploads onto Tangerine server), full datasets	
Description of achieved sample sizes, calculation of response rates with breakdown of disposition codes	▪ Realized response rates are adequate to maintain level of precision and power needed for the impact evaluation ▪ Reasons for non-response are well-documented using standard codes ▪ Calculations of sample weights, when needed, are done correctly
Data documentation	▪ Proper metadata (see Annex 2) is included with the datasets (codebook at a minimum) ▪ Documentation of any cleaning steps taken before delivery of final datasets ▪ Delivery of both raw and cleaned datasets, de-identified, if required