



**USAID**  
FROM THE AMERICAN PEOPLE

# USAID/LIBERIA

## READ LIBERIA! IMPACT EVALUATION

GC-10F-0033M/AID-OAA-M-13-00010

## EVALUATION DESIGN REPORT (DRAFT)

**March 2017 (Revised April 2017)**

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. This publication was produced before the details of the program were available to the evaluator. 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.

# **USAID/LIBERIA READ LIBERIA! IMPACT EVALUATION**

## **EVALUATION DESIGN REPORT (DRAFT)**

March 17, 2017 (First review March 21, 2017. Second Review April 14, 2017)

PN 7617.016.01

Contract No. AID-OAA-M-13-00010

# CONTENTS

	1
Executive Summary	1
A. Introduction	3
A.1 PROJECT BACKGROUND	3
A.2 Evaluation objectives	4
B. EVALUATION METHODOLOGY	5
Level of Randomization	5
C. Sampling	6
C.1 School and Student Selection for the Sample	6
School Selection	6
Student Selection	7
C.2 Power Calculations	7
D. DATA COLLECTION APPROACH	9
D.1 Outcomes indicators and Instrument Development	9
Early Grade Reading Assessment	9
Additional instruments	10
D.2 Selection of firm and recruitment of enumerators	10
D.3 Programming into Tablets and Training	11
Programming Instruments	11
Training	12
D.4 Data Quality Control	13
D. 5 Operational Plan for Data Collection	14
D.6 Ethics Clearance	15
E. Analysis Plan	16
F. Limitations	18
Evaluation Design Timing	18
Non-compliance and contamination	18
G. Workplan	19

# LIST OF TABLES

Table 1: Sample Size by Survey Type ..... 1

Table 2: Data Collection Activities for the READ Liberia! Evaluation: Sample size, baseline and endline dates ..... 8

Table 3: Skills and corresponding sub-tasks included in baseline EGRA tool..... 9

Table 4: Workplan .....20

# LIST OF FIGURES

Figure 1: Theory of Change of the READ Liberia! Activity ..... 4

# LIST OF EQUATIONS

Equation 1: Difference-in-Difference Estimation..... 16

# LIST OF ACRONYMS

EDR	Evaluation Design Report
EGRA	Early Grade Reading Assessment
ETP	Education and Training Policy
ICC	Intra-Cluster Correlation
IE	Impact Evaluation
IR	Intermediate Result
IRR	Inter-Rater Reliability
MDES	Minimum Detectable Effect Size
MOE	Ministry of Education
NORC	NORC at the University of Chicago
RCT	Randomized Controlled Trial
STS	School-to-School International
USAID	United States Agency for International Development

# EXECUTIVE SUMMARY

Under the Reading and Access contract, NORC at the University of Chicago is conducting an external and independent impact evaluation of the Read Liberia! Activity. Read Liberia! is a multi-faceted 5-year activity designed to improve early grade reading skills of Liberian students in Grades 1 and 2 in six targeted counties (Lofa, Bong, Grand Bassa, Nimba, Margibi, and Montserrado); the activity also includes a pilot set of interventions to foster the development of literacy skills for pre-primary students in public kindergarten schools. NORC’s impact evaluation focuses addressing the question: *What is the impact of the Read Liberia! activity as a whole on the reading fluency and comprehension of second grade students<sup>1</sup>?*

To this end, for the Read Liberia! impact evaluation, NORC proposes to use a randomized control trial (RCT) design, where schools are randomly assigned to treatment and control groups. Given the focus of the evaluation, we will collect baseline data – using early grade reading assessments and associated school-based surveys - from a random sample of second grade students before the end of the current academic year 2016-17 (in May 2017) and endline data from another random sample of Grade 2 students in the same school 2-3 years later (repeated cross sections of Grade 2 learners in a panel of schools). The main outcome indicators for this evaluation are oral passage fluency and the reading comprehension, for which data will be collected through the EGRA. These two sub-tasks will be the key focus of our analysis. However, the EGRA instrument also collects data other subtasks measuring pre-literacy skills and listening comprehension, which we will also analyze for the impact evaluation.

The evaluation sample size will consist of 90 schools, divided equally among treatment and control groups, with 16 students (eight boys and eight girls) per school for a total sample of 1,440 students for each round of data collection. We will also collect data from one Grade 2 teacher and one head teacher in each school. The table below presents the instruments that will be used for the data collection along with corresponding sample sizes.

**Table 1: Sample Size by Survey Type**

Survey Type	Sample Size
Early Grade Reading Assessment	16 Grade 2 students in 90 schools. Total: 1,440 students
Student Survey	16 Grade 2 students in 90 schools. Total: 1,440 students
Teacher Survey	1 Grade 2 teacher in 90 schools. Total 90 teachers
Head Teacher Survey	1 head teacher in 90 schools. Total: 90 head teachers

---

<sup>1</sup> At the time of writing this report, the Read Liberia! activity is still in the procurement phase and an implementing firm has not been selected. Therefore, the details of the actual activity and its components are unknown. As such, the design of the IE is based on the information available in the Read Liberia! RFP and agreements reached during conversations with USAID/Liberia and USAID/E3/ED.

USAID/Liberia, in collaboration with the Ministry of Education, has defined a set of criteria that will be used to identify schools eligible to receive the READ Liberia! activities. These criteria include all schools that: 1) are public; 2) are located in the six selected counties (Montserrado, Margibi and Bong, Grand Bassa, Nimba and Lofa); and 3) have grade one and grade two classes. The evaluation sample will follow the same criteria for school selection. In order to ensure that we have adequate numbers of students in each sample school to power the evaluation, we will also limit the sample to schools that have at least 20 learners in Grade 2. Students will be selected randomly in each school.

NORC will work closely with subcontractors School-to-School International and the Khana Group (TKG) to conduct all data collection activities. Our field team, recruited by TKG, will consist of six field teams, each comprised of one supervisor and three enumerators, all drawn from a pool of candidates who have school-based (and preferably EGRA) survey experience and are familiar with tablet-based data collection. To end up with 24 field staff, we will train 30 candidates over an eight-day period. The first six days of training will consist of classroom instruction, role playing exercises, exercises to familiarize enumerators with tablets, inter-rater reliability exercises, and several rounds of group and paired practice with each subtask. During the last two days all trainees will participate in a pilot exercise in out of sample local schools to gain practical experience and practice in conducting the assessments with actual students, followed by a feedback session with all trainees to further improve practices and correct errors that were observed during the pilot exercise.

Strict and systematic data quality control measures will be instituted during the training and field period. NORC, STS and TKG will provide quality assurance during the training, while supervisors and Quality Control Officers will ensure quality assurance during the field period.

The evaluation faces several risks. Primary among them, is the fact that the design of the evaluation preceded the award of the Read Liberia! contract, which potentially poses a threat to the evaluation in the event that the eventual implementation of the activity diverges from the assumptions on which the design is based. Other risks are common to most evaluations and include the possibility of contamination across treatment and control groups; the evaluation team will attempt to detect and account for such contaminations during endline data collection. A final risk that USAID's Goal 1 team has raised concerns about test security related to the EGRA instrument being used for this evaluation, since it is the same assessment tools used in 2011 and 2015. However, there is no evidence of test leakage and, given time and budget constraints, NORC and USAID decided the creation and testing of a new tools was not warranted.

## A. INTRODUCTION

Under the Reading and Access contract NORC at the University of Chicago has been charged with the following activities:

1. Designing an external and independent impact evaluation (IE) of the Read Liberia! activity focusing on Grade 2 students, and
2. Collecting baseline data from Grade 2 students in Liberia by the end of the academic year 2016-17

At the time of writing this report, the Read Liberia! activity is still in the procurement phase and an implementing firm has not been selected. Therefore, the design of the IE is based on agreements reached during conversations with USAID/Liberia, USAID/E3/Ed and information included in the Request for Proposals (RFP) and corresponding amendments.

### A.1 PROJECT BACKGROUND

Liberian students' reading performance is below the desired levels. A recent program, the Liberia Teacher Training Program (LTTP), measured oral reading fluency in treated schools at only 14 words per minute. This was accompanied with poor performance in other literacy abilities, such as reading comprehension. To address this issue, USAID is implementing READ Liberia! (RL), a 5-year activity designed to improve early grade reading skills of Liberian students in grades one and two and pilot improvements in developing emergent literacy skills for Liberian students in public kindergarten schools in six targeted counties (Lofa, Bong, Grand Bassa, Nimba, Margibi, and Montserrado).

At the time of writing this report, this activity is still in procurement; therefore, the implementing firm has not been selected and details about implementation are not available yet. The RFP that bidders are responding to, however, presents a Results Framework with intermediate results (IR) linked to higher level objectives. This Results Framework links increased government commitment to and support of evidence-based early grade reading instruction (IR1), improved early grade reading instruction (IR2), improved systems for delivery (IR3), and increased parent, community and public-private partnerships support for early reading (IR4) to students' ability read with fluency and comprehension by the end of second grade.

The exact sub-activities to be implemented in order to achieve each of those four IRs are yet unknown; however USAID/Liberia posits that the IRs can be achieved *if* (1) Kindergarten students are taught the oral vocabulary lexicon needed for emergent literacy in English; (2) Official time allocations for reading instruction in the early grades are increased and enforced; (3) Data on early grade reading is collected and used to drive system-wide decision-making; (4) Reading interventions are refined and simplified with the key components needed for children to learn to read words and to understand what they read are preserved and improved if necessary; (5) Texts appropriate for early grade reading instruction are improved and additional leveled texts created; (6) Teachers receive intensive training and systematic coaching

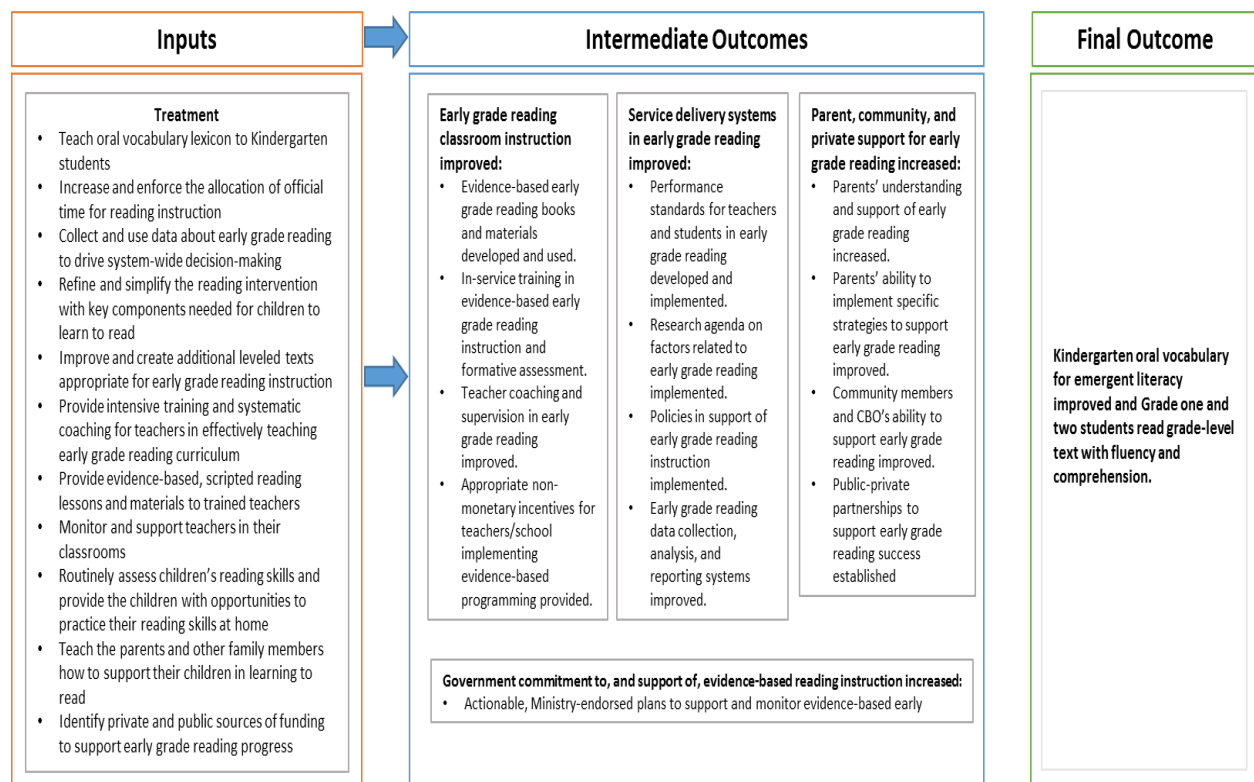


in effectively teaching the early grade reading curriculum; (7) Trained teachers receive evidence based, scripted reading lessons and materials to support reading improvement in the early grades; (8) Teachers are monitored and supported in their classrooms; (9) Children’s reading skills in the early grades are routinely assessed and children provided opportunities to practice their reading skills at home; (10) More parents and other family members learn how to support their children in learning to read; and (11) Private and public sources of funding to support early grade reading progress are identified.

Thus, while we assume the implementer will develop strategies and activities to provide these inputs, we do not know the details of the interventions that will be proposed, agreed to, and eventually, implemented. The full Theory of Change is illustrated below in Figure I and more details about the sub-activities expected to be included in Read Liberia! can be found in the RFP which is included in Annex I.

As such, the ensuing evaluation design is based on information available through the RFP and agreements reached with the USAID Liberia Mission on how key activities will be rolled out and implemented.

**Figure I: Theory of Change of the READ Liberia! Activity**



## A.2 EVALUATION OBJECTIVES

The main purpose of the impact evaluation will be to assess the causal impact of the RL activity on reading performance of second grade students.

Despite the different sub-activities that this intervention includes, the focus of this evaluation is to measure the effect of Read Liberia! as a whole. NORC has been instructed by USAID to plan for an impact evaluation focusing only on Grade 2 students<sup>2</sup>. The main goal of the IE will be to answer the following question:

***What is the impact of the Read Liberia! activity on the reading fluency and comprehension of second grade students?***

***To this end, norc will collect data related to grade 2 students before the end of the academic year 2016-17 (May 2017). these data will serve as the baseline for the evaluation.***

## **B. EVALUATION METHODOLOGY**

The NORC team decided that an impact evaluation is feasible and also necessary to properly answer the agreed upon evaluation question. An impact evaluation serves to assess the causal effect of a specific intervention on a set of outcomes. It allows us to attribute changes in an outcome to a specific intervention or set of interventions by answering the counterfactual question “What would have happened to activity participants in the absence of the intervention?” Ideally, this is done by observing the same program participants both with and without the intervention at the same point in time. Of course, this is not possible; at any given time, a participant either receives the intervention or not. Therefore, we can never directly observe the counterfactual and instead need to create a comparison group to serve as the counterfactual. Identifying a credible comparison group is a critical aspect of an impact evaluation and there are several approaches to do so.

The evaluation team will use an experimental or randomized control trial (RCT) approach to answer the evaluation question, with schools randomly assigned to treatment and control groups. The fact that the activity has not started yet and that the total number of schools is greater than the number of schools to be treated in the six counties, allows for an RCT. As we mentioned above, we want treatment and control groups to be as similar as possible; the primary advantage of a randomized design is that it minimizes the potential for selection bias—that is, the possibility that observed differences in outcomes between the treatment and control groups are due not to the impact of the project, but to other systematic differences between the two groups that the evaluation is not able to account for. Another advantage is that the randomizing approach also tends to be fairly simple to convey to stakeholders and policy makers and fair when not all schools can receive treatment at once.

### **Level of Randomization**

The randomization into treatment and control groups will occur at the school level. Randomization at the school level is possible given that all the sub-activities included in IR2, IR3

---

<sup>2</sup> The RL program will also target kindergarten students in 60 schools in in order to improve oral vocabulary for emergent literacy and grade one students. However, the present impact evaluation will not include them on this evaluation design.

and IR4 can be implemented at the school level, while the IRI sub-activities are common to the country as a whole.

The NORC evaluation team considered other options such as randomization at district or smaller area school-clusters level. Randomization at the district level is possible but requires a very large number of schools given that there are only 44 districts in the 6 counties where RL will be rolled out. Since this option has large cost implications, we discarded. Creating smaller geographic school clusters requires knowing the precise location of each school and a detailed mapping exercise. This process is not possible given the time and budget constraints.

## **C. SAMPLING**

Our methodological approach to answer the evaluation question requires collecting data before the start of the activity implementation –to make sure treatment and control schools are, on average, statistically identical- and after a few years of activity implementation.

There are two alternatives to collecting data to estimate the impact of RL on students at the end of Grade 2: (1) repeated cross sections at baseline and endline of Grade 2 learners in a panel of schools, or (2) panel data following students or their cohort from the time they start Grade 1 and until the end of Grade 2. Both approaches are commonly used in impact evaluations and each have advantages and disadvantages. Given that USAID/Liberia has indicated a strong preference for assessing learners this year (at the end of the 2016-2017 academic year), rather than waiting till the start of next academic year, and taking into account the budget implications of collecting panel data (additional costs of locating the same students, for example), we opted to use a repeated cross sections sample, collecting data from a random sample of Grade 2 students in May 2017 - as requested by USAID/Liberia - and again from a random sample of Grade 2 students in the same school in two to three years<sup>3</sup>.

### **C.1 SCHOOL AND STUDENT SELECTION FOR THE SAMPLE**

#### **School Selection**

USAID/Liberia, in collaboration with the Ministry of Education, has defined a set of criteria that will be used to identify schools that are eligible to receive the READ Liberia! activities. These criteria include all schools that: 1) are public; 2) are located in the six selected counties (Montserrado, Margibi and Bong, Grand Bassa, Nimba and Lofa); and 3) have grade one and grade two classes. The evaluation sample will follow the same criteria.

For the evaluation, NORC requested an additional criterion, that the schools have at least 20 learners in Grade 2. This request responds to the need to maximize the probability of collecting data from a minimum of 16 students (eight girls and eight boys) per school, which

---

<sup>3</sup> The exact length of the evaluation and the year in which endline data will be collected will be decided at a later date, and taking into account the activity details. There are trade-offs to consider: going soon minimizes the risk of teachers switching between treatment and control schools while postponing the endline gives the activity more time to be fully implemented and adjusted.

gives us the sample size necessary to have enough power to detect impact, given the number of schools in the sample (see sample size calculation in Section C2). The NORC evaluation team has specified 20 students rather than 16 to account for absenteeism which is quite common.

The EMIS list of schools provided by the MOE includes designations for urban and rural schools. Using this data, the sample is assigned proportionally to the number of urban and rural schools within each county. As a result, the sample is representative of the population of public schools which have grade 1 and grade 2 classes for each county in the sample.

The NORC evaluation team finished preparing the sample according to the above criteria on April 1.

### **Student Selection**

Sixteen second grade students - 8 girls and 8 boys - will be selected randomly in each school. All learners will be included unless they refuse to participate or severe disabilities preclude them from participating.

It is important to note that the school selection criteria define what kind of schools will be part of the baseline and, therefore, has implications for the RL activity implementation. For a successful evaluation, the baseline needs to be a representative sample of the schools that participate in the intervention; thus the criteria for school participation in the RL program have to match the criteria for baseline school selection. Once the baseline sample is drawn, the criteria for school selection cannot be modified. The implementer will need to work with schools that match the selection criteria used for the baseline or the ability to undertake an impact evaluation will be compromised as the sample will no longer represent the universe of schools with which the IP is working.

## **C.2 POWER CALCULATIONS**

This section provides the details of the calculations used to arrive at the recommended sample sizes for the IE. As with any quantitative analysis based on survey data, the required sample size is determined by a mathematical calculation that depends on a number of factors. These include features of the study design, properties of the data and outcome variables, and the desired precision of the analysis. For this exercise, we make the following assumptions:

- $\alpha$  is the significance level of the test, or probability of Type I error. We use the standard value of 0.05.
- $\beta$  is the power of test, where  $(1-\beta)$  is the probability of Type II error. We use the standard value of 0.8.
- $\rho$  is the intraclass correlation coefficient, or ICC. The ICC in the present case is a measure of how much variability lies between schools and how much lies within schools. Based on King, et al (2015) work with Liberian EGRA data, we assume an ICC of 0.3.<sup>4</sup>

---

<sup>4</sup> This figure corresponds to the one in Annex H for grade 2 in the Endline report of LTTP (King, Simon, Medina Korda, Lee Nordstrum and

- $r^2$  is the proportion of the variation in the outcome due to the covariates anticipated in the regression analysis. In our case, these covariates will include a range of household and individual characteristics, as well as school fixed effects. For this parameter, we assume an approximate value of 0.2.
- MDES is the minimum detectable effect size. The MDES is the smallest impact of the activity on the outcome variable that the evaluation will be able to detect. The selected MDES is 0.34 of a standard deviation, a medium size effect. This corresponds to the standardized effect for Oral Reading Fluency found in the midline evaluation of LTTP (The NORC team estimated this standardized effect using own calculations and data from the Midline report of LTTP.<sup>5,6</sup>) During a phone call (February 2017) USAID/Washington and USAID/Liberia considered it a reasonable effect to be expected given previous experiences and the intensity of the intervention.

Based on these parameters, the required sample size is 45 schools in each study group (treatment and control) with 16 students in each school, for a total sample of 1,440 students. The teacher and head teacher sample consist of one Grade 2 teacher and one head teacher per school and has the objective of completing the picture of the learners' education context and helping understand the mechanisms behind the impact effects. Table 1 presents a summary of data collection activities for this evaluation.

**Table 2: Data Collection Activities for the READ Liberia! Evaluation: Sample size, baseline and endline dates**

Survey Type	Sample Size	Baseline Date	Endline Date
Early Grade Reading Assessment	16 Grade 2 students in 90 schools. Total: 1,440 students	May, 2017	May/June 2019 or 2020
Student Survey	16 Grade 2 students in 90 schools. Total: 1,440 students	May, 2017	May/June 2019 or 2020
Teacher Survey	1 Grade 2 teacher in 90 schools. Total 90 teachers	May, 2017	May/June 2019 or 2020
Head Teacher Survey	1 head teacher in 90 schools. Total: 90 head teachers	May, 2017	May/June 2019 or 2020

Susan Edwards (2015). Liberia teacher training program: Endline assessment of the impact of early grade reading and mathematics interventions)  
 5 DeStefano, Joseph, Timothy Slade, and Medina Korda (2013). Liberia Teacher Training Program (LTTP): Midterm Assessment of the Impact of Early Grade Reading and Math Interventions.

6 We divide the DID for ORF and grade 2 in table 8 of the midline report, by the standard deviation that we calculated using the data in table 2 of the midline report.

## D. DATA COLLECTION APPROACH

### D.1 OUTCOMES INDICATORS AND INSTRUMENT DEVELOPMENT

#### Early Grade Reading Assessment

The NORC evaluation team will measure reading fluency and comprehension using the Early Grade Reading Assessment (EGRA). EGRA is a widely used tool designed to measure various aspects of reading ability which are, in general, considered foundational literacy skills (National Reading Panel, 2000; Snow, Burns & Griffin, 1998). These skills include at least five main areas: alphabets (letter knowledge and phonemic awareness), phonics, fluency, vocabulary, and comprehension. EGRA has been previously used to measure reading performance and pre-literacy skills in early primary grades in Liberia.

We will use an existing version of EGRA, created for Liberia students in 2011, to collect baseline data. The assessment has not been modified given the short time available before data collection. The skills tested in the EGRA tool and the corresponding sub-task are indicated in Table 2 below.

**Table 3: Skills and corresponding sub-tasks included in baseline EGRA tool**

Skill	Sub-task & Measurement
Orientation to print	Awareness of text direction, where to start reading, how read down a page. 3 questions.
Letter naming fluency	Number of letters correctly identified out of 100 in 60 seconds
Phonemic awareness	Identify words that start with a different sound. Number of correct identified sounds out of 10
Phonics/Alphabetic Principles	Number of nonsense words correctly decoded out of 50 in 60 seconds
Fluency	Familiar word fluency, number of correct words read out of 50 in 60 seconds
	Oral passage reading Number of words in a reading passage of 60 words read fluently (with accuracy) in 60 seconds
Reading Comprehension	Oral recall. Number of questions (out of 5) about a reading passage (read by student) answered correctly
Listening Comprehension	Oral recall. Number of questions (out of three) about a passage read aloud (by facilitator) answered correctly

The main outcome indicators for this evaluation are oral passage fluency and the reading comprehension; hence, these two sub-tasks will be the key focus of our analysis. However, the EGRA instrument also includes other subtasks measuring pre-literacy skills and listening comprehension, which we will be also analyze for the impact evaluation.

## **Additional instruments**

In addition to the EGRA tool, the baseline data collection will also include: (1) a short student background questionnaire which will be administered immediately following the EGRA to every student sampled, (2) a teacher questionnaire which will be administered to Grade 2 teachers in each school, and (3) a questionnaire administered to the school principal.

NORC will use validated questionnaires that have been used previously in Liberia along with the 2011 EGRA tool. We will slightly modify questions when necessary such that they are relevant to the objectives of the activity as defined in March 2017.

The principal survey includes questions about their education and experience, coaching and monitoring of teachers, interaction with education officers, parents, school resources, etc. The teacher questionnaire also contains questions about education and experience, interactions with the principal or others regarding coaching and monitoring, instructional practices related to reading, etc. Finally, the student context interview focus on the child home language, reading practices at home, some instructional practices in the classroom, and a few household possessions.

All instruments can be found in in Appendix 6-8 at the end of this report.

## **D.2 SELECTION OF FIRM AND RECRUITMENT OF ENUMERATORS**

NORC conducted a limited competitive bid process to select the data collection firm for this project. We contacted and interviewed a total of five firms that work in Liberia, and asked these firms to provide capabilities statements. Of those firms, only two had any experience conducting education related research in schools. Upon request, the two firms, the Agency for Economic Development and Empowerment (AEDE) and The Khana Group (TKG), submitted full technical proposals and budgets for this assignment, based on the Terms of Reference that NORC provided. Based on these technical proposals, the NORC team selected The Khana Group for the assignment. The decision was based on their prior experience with EGRA and EGMA assessments, references, their experience with tablet based interviewing, their specific experience using Tangerine software to conduct research, and their overall technical approach to the project. TKG was significantly more qualified and prepared for an assignment of this nature than AEDE in each of these areas.

TKG is a social impact management consulting and research firm that works in sub-Saharan Africa. With offices in Liberia, Ghana, Nigeria and the United States, the firm focuses on delivering customized solutions and quality services in research and policy analysis, monitoring and evaluation (M&E), capacity development, technology and innovation, and strategy across various sectors, including health, agriculture, education and economic growth. Specific to this project, TKG has led and supported EGRA assessments in Nigeria and Liberia. TKG also has experience using Tangerine software and recently supported an EGRA & EGMA assessment project using the Tangerine platform in Liberia.

TKG will be primarily responsible for recruitment of enumerators. Enumerator candidates need



to have survey experience in education area, preferably conducting EGRA, and be familiar with data collection using tablets. TKG has a pool of candidates that meets these criteria. Moreover, USAID/Liberia was able to provide NORC with a list of 30 enumerators who had conducted EGRA using Tangerine during a previous data collection. We have provided this list to TKG, who will invite those enumerators, if available, to participate in the current data collection.

Enumerators and supervisors will be recruited as one group with at least 25% more candidates invited than are needed for the field period. All participants will begin training with the understanding that (1) those who demonstrate the greatest level of proficiency during training and show good organizational skills will be invited to become supervisors during the field period, (2) those who demonstrate the next highest levels of proficiency during training will be invited to become enumerators during the field period, and (3) the remaining participants will not participate in the field work unless another member of the field staff is unable to continue with the project during the field period (these individuals will be considered reserve enumerators and will be ranked by ability).

At the end of training, participants will be divided into six teams, each consisting of one supervisor and three enumerators for a total of 24 field staff. We plan to invite 30 participants to training, which means we expect 6 reserve enumerators to be trained and available should anyone on a field team be unable to complete the field work.

### **D.3 PROGRAMMING INTO TABLETS AND TRAINING**

NORC's subcontractor on the Reading & Access contract, School-to-School International (STS), will support all EGRA-related data collection activities. The NORC team has worked with STS in the same capacity in several similar projects in the recent past. STS' primary responsibilities will include development of training materials, co-leading the enumerator and supervisor trainings in Liberia with the assistance and cooperation of TKG, adapting programming and troubleshooting the instruments on Tangerine, managing the data on their secure Tangerine server, overseeing field launch, and helping manage quality control and data cleaning measures throughout the project's field period.

School to School International (STS) is a non-profit organization based in the United States that seeks to transform education and school health in developing countries so that every child can succeed in school. For over a decade, STS has partnered with individuals, government ministries, non-governmental organizations, private foundations, and research institutions to improve the quality of children's learning through educational research and integrated support solutions for children in primary school.

#### **Programming Instruments**

STS has been able to obtain the 2011 EGRA instrument and companion tools programmed in Tangerine. STS will adjust the programming of the instruments as necessary (EGMA subtasks used during the previous administration will be deleted, for example) and test all tools in Tangerine. The EGRA, Student Context, Teacher, and Principal survey instruments can be found in Annexes 6, 7, and 8 at the end of this report.



## Training

NORC and STS will carry out the training of enumerators and supervisors in the administration of EGRA and other tools in collaboration with TKG. The NORC/STS team places a strong emphasis on participatory and demonstration activities during training to hone the skills of participants and enable them to put theory into practice. To this end, the training will include activities such as:

- Role-play exercises: facilitator demonstrations and trainee demonstrations
- Small-group and whole-group practice
- Inter-rater reliability (IRR) exercises
- Use of videos, whenever possible (including for IRR exercise and role-plays)

In addition to EGRA, we will train enumerators and supervisors in the administration of the teacher and principal questionnaires, and student context interview.

The training will start on April 24<sup>th</sup> and will last a total of eight days. The first six days of training will consist of classroom instruction, role playing exercises, exercises to familiarize enumerators with tablets, inter-rater reliability exercises, and several rounds of group and paired practice with each subtask. The first day of training will also include sessions on ethics, confidentiality, and important aspects of working with children.

After the classroom portion of training is complete, all trainees will participate in a pilot exercise in out of sample local schools to gain practical experience and practice in conducting the assessments with actual students. In addition to gain real experience and confidence, the main focus of this pilot exercise will be to evaluate each enumerator and allow the trainers to observe how well each of them is able to conduct the assessment in an environment that closely mirrors the situations they will encounter during field work. Groups of trainees will be assign a school where each one will be required to conduct a minimum of three student assessments each. In addition, student sample selection and teacher and principal interviews will be practiced.

After the day of in-school piloting, the final day of training will consist of a feedback session with all trainees to further improve practices and correct errors that were observed during the pilot exercise. Supervisors will also be selected on the final day and will be taken through an additional half day training session to ensure they all understand the duties and responsibilities expected of them as supervisors. These duties and responsibilities will be discussed throughout training so that both enumerators and supervisors share the same expectations for every team member's role and responsibility to the team.

The full training agenda, along with other details about the enumerator and supervisor training, can be found in the training manual in Annex 2 at the end of this report. Annexes 3 and 4 contain training materials for Tangerine and Table use and EGRA respectively.

## D.4 DATA QUALITY CONTROL

For the training, NORC and STS will create a data quality control checklist that will be used by trainers and supervisors to assess the level of skills of enumerators. At least 25% more trainees than necessary will be invited to training so that only the most qualified are selected for the actual field period. This selection will be based on assessments during the classroom training using the IRR tests and during the pilot test, as well as their score on the training exit quiz that NORC and STS will develop and implement. Trainers from NORC, STS and TKG will all evaluate each participant's level of understanding and competency and will jointly decide on assignments of supervisors, enumerators, and reserve enumerators in a manner that creates the highest quality teams available from the total group of training participants.

During the field work, supervisors will travel to each school with their team of enumerators. Supervisors will oversee the entire operation in the school and interviews throughout the day, ensuring that EGRA and the other survey instruments are administered correctly. Supervisors will also contribute to data collection activities on an as-needed basis to ensure all interviews are conducted before the end of the school day. In addition, supervisors will also monitor the data upload every day and immediately communicate any problems to TKG's Data Manager, who will further communicate with STS staff as needed.

Data quality assurance in schools will take two forms. First, both supervisors and quality control officers will directly observe interviews from start to finish. Second, interviews will be spot-checked where a portion of an interview is observed by a supervisor or quality control officer. Supervisors will directly observe at least 10% of all completed interviews conducted by their team throughout the field period. Quality control officers will directly observe an additional 10% of all interviews conducted throughout the field period. As a result, a minimum of 20% of all interviews will be quality controlled by direct observation of the full interview. Supervisors will also conduct spot checks on a minimum of 15% of the interviews conducted by their team each day. Quality control officers will also conduct additional spot-checks during their visits to schools, as time permits. For all spot-checks and observations, supervisors and quality control officers will use the data quality control checklist that NORC and STS will develop. In total, a minimum of 20% of all interviews will be directly observed and 15% of all interviews will be spot checked, either by the team supervisor or the quality control officer visiting the various schools. This will ensure data quality is evaluated consistently among all teams and in an even and systematic manner throughout the entire field staff. Additional information about quality control can be found in the training manual in Annex 2.

As mentioned above, TKG has also hired three Quality Control Officers to further assist with monitoring and support throughout the field period. These Quality Control Officers, who will participate in the full training, will be individuals who TKG has worked with in the past in a similar capacity and have extensive experience in monitoring enumerators during similar research projects. Each quality control officer will be assigned two teams to monitor throughout the field period. They will visit one (or both, depending on geographic constraints) of the teams they are assigned each day to further ensure all procedures for data collection are being adhered to by all team members, including the supervisors for each team.

Finally, TKG has provided a communications and technology plan that will ensure smooth communication and usage of technology, especially with regard to phones, the internet, and electricity during the field period. Supervisor will be available via phones provided by TKG and will be easily accessible whenever the teams are within reachable range. Because of the limitations with internet access in rural areas in Liberia, TKG has mapped the best internet service providers per county and will provide each team with modems for the appropriate counties. This will help limit upload issues with tablets. TKG has also invested in 300W Car Power Inverters with two 110V AC outlets and 2 USB ports to provide 700 Watt of peak power. These devices will provide power to charge laptops, smart phones and tablets using the vehicle's charging port. TKG invested in these devices and they have eliminated problems related to batteries draining on tablets or other data collection devices. In addition, each team will be equipped with a power bank. In very rare occasions teams can also use local charging boots to charge the devices as a last resort.

## **D. 5 OPERATIONAL PLAN FOR DATA COLLECTION**

The sample has been selected by the NORC team and approval was obtained from the MOES to contact CEOs, EDOs, and schools to alert and engage them in the upcoming data collection activities (see authorization in Annex 5).

Eight staff members from TKG will be deployed to the field prior to launch of the survey to contact all districts and all schools and ensure accuracy of the sample list. The three main objectives of these visits are to: (1) gain advance entry into the schools and explain the purpose of our data collection; (2) verify the eligibility of the schools for the RL intervention and evaluation sample; and (3) verify the exact location of the schools using GPS enabled tablets.

For the actual field period, a total of six teams, each comprised of three enumerators and one supervisor will administer all instruments over a period of three weeks. Each team will be assigned schools to visit throughout the field period in a manner that is as geographically compact as possible in an effort to minimize travel between schools during the field period. Each team member (all three enumerators and one supervisor) will be assigned a specific tablet and each team will have one extra tablet in the event of a technical failure while the team is in the field. Each tablet will be loaded with Tangerine and will be fully tested prior to sending it to field. Each team member will also have a laminated set of stimuli (show cards) which will be used during the administration of the EGRA. In the event that a tablet and the backup tablet both fail while a team is in the field, each team will also bring a set of paper questionnaires and timers so they can administer the EGRA to students without the assistance of a tablet. In the unlikely event this has to be done, data for those interviews will be entered at a later time and TKG will arrange for replacement tablets to be sent to the team.

Each team will be supplied with a vehicle throughout the field period so they will not have to rely on public transportation to access the schools. Teams will travel together to each school and will remain together until all work is completed in that school. We anticipate that each team will be able to complete all 18 interviews (16 students, one teacher, and one principal) in one school in a day. With 90 schools in the sample, each team will be required to conduct interviews in 15 schools. This means 15 working days will be budgeted for the primary data

collection, with an additional three days added in the work plan to account for cleanup of any missing schools in the event a school was not in session on the day of the initial visit, to account for the replacement of schools for unforeseen reasons, or to allow for a potential revisit to a school if a quality control need is identified. As such, a total of 18 working days are budgeted to complete data collection. Quality Control Officers will also work with the teams, as described above, to assure the quality standard of the data collected throughout the entire field period.

Following enumeration, quantitative data will be exported into STATA and processed through two rounds of cleaning, first by TKG, and then by STS staff in the United States, who will house the server which store all of the data as it streams in from the tablets. Data will be transferred between NORC, STS, TKG, and ultimately sent to USAID using a Secure File Transfer Protocol (SFTP) system. As data is reviewed, all changes to the raw data made during cleaning and analysis will be recorded in .do files, leaving the raw data intact and ensuring the process is completely replicable. Each respondent will be automatically assigned a unique identifying number by the Tangerine software so that respondent records in the dataset can be cleaned and to ensure duplicates are not included.

After the project is complete, NORC will maintain the integrity of the EGRA instrument by removing the instrument from any materials which will be released to the public. This will help ensure that the instrument is able to be accessed and used in any way which could influence or assist students who make take the assessment in the future. NORC will also ensure that all data are handled in accordance with our ethical guidelines for data security, which includes maintaining confidentiality of all respondents through all phases of reporting and data storage. As no personally identifying information is being collected, respondent confidentiality and anonymity should be easy to maintain and the final data set will not pose a risk of breach.

## **D.6 ETHICS CLEARANCE**

All human subjects related activities in this IE are subject to IRB ethics clearance and proper procedures to in place to obtain approval. NORC at the University of Chicago IRB forms have been submitted on March 15 and approval was obtained on March 29, 2017.

USAID/Liberia mission had informed NORC that no local IRB approval was required, however NORC subsequently learned that ethics clearance from the IRB at the University of Liberia is actually required. Materials for the University of Liberia's IRB were submitted to the head of the IRB committee electronically on March 31 and hard copies were delivered to the IRB by TKG on April 3. An IRB review meeting was held on April 13 in Liberia with the participation, in person, of TKG project manager and researcher, and the participation of NORC evaluation team PI and researchers over the phone. During the meeting a small number of adjustments, primarily related to the language in the consent statements and documentation, were requested by the IRB committee. While the consent statements originally submitted already covered obtaining assent from children prior to their being interviewed, NORC also created a written student participation consent statement to be read and signed by each principal in each school (see Annex 9). Since the principal is primarily responsible for the students throughout the school day, this consent will ensure the principal is fully aware of the activities the students will be taken through and the principal's written consent is being complete in lieu of seeking

additional parental consent, which is logistically prohibitive to obtain for each student on the day the teams arrive at each school. These adjustments have been made and updated materials were resubmitted electronically on April 16, 2017. No further hard copy submissions are required.

In addition to the IRB requirements and guidelines we will follow, TKG will ensure that interviewers are equipped with pedagogical interviewing techniques to ensure that children are comfortable to respond to the tasks. Every interviewer will undergo TKG's child safeguarding training as part of the full training exercises. These are TKG's standard policies in working with children and represent their understanding of the importance of putting children in the center of research projects such as this. TKG will also leverage the Children Participatory Assessment approach, which involves structured dialogue with children. TKG has demonstrated their understanding that children view, experience and communicate their needs and protection concerns in various ways. However, they also understand that children often lack access to and are excluded from adult decision-making processes and face additional barriers to their participation. As such, TKG has agreed to work with NORC throughout the process of training and field work to ensure they are meeting all its child's protection policies.

In addition to the children specific ethical and theoretical training field teams will received, they will also be trained on standard ethical procedures for interviewing and personal research in general. This training will include detailed explanations of confidentiality and anonymity and all enumerators, supervisors, quality control officers, TKG field support staff and data management staff will be required to sign a pledge of confidentiality. This ethics training will also include proper handling of electronic data before, during and after the interview, secure handling of all supplemental interviewing materials, and standards for proper and ethical treatment of respondents before, during and after the interview.

## E. ANALYSIS PLAN

As explained in Section B, an experimental approach will be used to create a treatment and a control group of schools. We will estimate the causal impact of the program on EGRA scores using regression models of the following form:

### Equation 1: Fixed Effects Model

$$Y_{ijt} = \beta_0 + \beta_1 \delta_{\text{time}} + \beta_2 \mathbf{X}_{ijt} + \gamma(\delta_{\text{time}} \times \delta_{\text{treat}}) + \varphi_j + \varepsilon_{ijt}$$

where:

$Y_{ijt}$  is the EGRA sub-task score for student  $i$  in school  $j$  at time  $t$ ,

$\delta_{\text{time}}$  is a dummy variable equal to one at endline,

$\delta_{\text{treat}}$  is a dummy variable equal to one for observations in the treatment group,

$\mathbf{X}_{ijt}$  is a vector of control variables,

$\varphi_j$  is a school-level fixed or random effect,

$\varepsilon_{ijt}$  is a random error term, and

$\beta_0, \beta_1$ , the vector  $\beta_2$  and  $\gamma$  are parameters to be estimated

The parameter of interest is  $\gamma$ , the coefficient on the interaction term between the time and treatment dummies. The value of  $\gamma$  will provide an estimate of the impact of Read Liberia! relative to the control group on EGRA scores.

The parameter of interest is  $\gamma$ , the coefficient on the interaction term between the time and treatment dummies. The value of  $\gamma$  will provide an estimate of the impact of RL relative to the control group on EGRA scores.

Although the evaluation question focuses on the reading fluency and reading comprehension scores, we plan to estimate the impact of RL on all the sub-tasks include in the EGRA tool.

Note also that the models will include a set of control variables  $X_{ijt}$  reflecting student characteristics that will be obtained from short surveys of students in the course of administering EGRA. Potential control variables include age, home language if different from English, whether child reads at home regularly, and some basic questions about household assets to estimate wealth status, among others. Incorporating these control variables into the analysis will improve the precision of the impact estimates and help controlling for any potential change in characteristics of learners between baseline and endline as well.

In addition to estimating the overall impact of Read Liberia! on reading outcomes, we will also carry out a sub-group analysis to examine whether the impacts were different for boys and girls. Depending on the findings derived from baseline data, we can explore some interactions between school characteristics and the RL activity. For example, if we find that student characteristics and performance are different between rural and urban schools, we could study whether the effects of the activity are different accordingly to the school location.

Furthermore, as mentioned before, the data collection effort will include two other survey instruments to gather information from the second grade teachers and school principals. At baseline, we will collect some information from teachers about, for example, the training and coaching they may have received. More information about these and other sub-activities impacting on schools and teachers could be collected at endline. These data will be useful to learn how sub-activities were implemented (according to teachers and principals) and if the treatment dosage received by teachers has a differential effect on the student reading performance<sup>7</sup>.

Additionally, once the RL activity is developed, it will be become clear what monitoring indicators may be available to the evaluation team. It is possible that those indicators could be also incorporated into the analysis to see how differences in the implementation of the activity

---

<sup>7</sup> We assume all teachers are in principle supposed to receive the same amount of training and coaching. If coaching, for example, were endogenous and less able teachers receive more and more able teachers less, then, the results would be misleading.

across schools –if any- may correlate with learners’ performance.

Finally, USAID/Washington expressed interest in collecting information on classroom practices in a small sample of schools to study the take up and fidelity of implementation of the RL activity by teachers and coaches. Once collected, these data could be shared with NORC in order to include a small descriptive analysis of the extent in which teachers and coaches’ behavior and practices resembles the Read Liberia! Methodology and related (not in a causal way) to learners’ performance.

These additional analyses will help to understand fidelity of implementation of the activity, and perhaps mechanisms behind findings. However, it is important to note that it will not be possible to estimate separately the causal impact of different sub-activities of RL. .

## **F. LIMITATIONS**

### **Evaluation Design Timing**

The evaluation design for RL was completed before the activity was procured and fully developed; therefore, only the information included in the RFP was available to the evaluation team. Although this is not ideal, the evaluation team used their previous experience evaluating other similar activities and agreements with USAID Liberia about project implementation to plan for the IE and the baseline data collection.

For the same reason, the evaluation team does not know what monitoring data may become available to support the evaluation. Again, based on the knowledge acquired during previous experiences evaluating this type of activities, the evaluation team will be collecting data that relate to the implementation of some of the sub-activities, directly from teachers and principals. Once the monitoring indicators are defined and agreed upon, the evaluation team will study how to incorporate them in the IE.

### **Non-compliance and contamination**

In cases where the treatment received does not correspond to the treatment assigned, due to non-compliance or other reasons, there is nothing that can be done to remedy the situation. The inference has to be based on the original randomly assigned groups. However, it is important to understand where and why the deviations have occurred, because that will affect how the findings should be interpreted.

There are several possible sources of contamination that could affect our evaluation design. First, it is possible that some learners do not stay in the same school from the start of grade 1 until we assess them at the end of grade 2. It is not likely that a lot of movement will be seen given that students will only be finishing grade 2. However, if switching occurs between treatment and control schools, there would be learners that do not maintain their original assignment status during the period of the evaluation and this would tend to underestimate the impact of the intervention.

The ideal way to assess the extent of this issue would be to have a system that tracks the learners. Some countries have such systems in place but this, as far as the NORC team knows, is not the case in Liberia. In order to mitigate this potential problem, a good alternative is to ask the learners assessed at endline whether they have been in the same school in the previous year and use the information to evaluate the extend of the problem and control for it in the analysis. <sup>8</sup>

Second, it is also possible that teachers and/or school principals switch between treatment and control schools. Again if this situation takes place it would lead to an underestimation of impact of the RL activity. The risk of teachers and principals switching schools grows with the time elapsed between baseline and endline. The approach we will use will be similar to that used with students switching schools, gathering information from teachers and principals.

Finally, we are assuming that all sub-activities included in RL will be implemented at the school (or its correspondent immediate community) level, with the exception of the sub-activities that target the Liberian government education system as a whole. This assumption means that only treatment schools receive the RL intervention, however there is always some risk of spillovers. This may happen if, for example, schools share RL materials or RL reading instruction approach. It is quite unlikely that this will occur and if it does, the possibility of an impact on the control schools is quite remote. The coaching sub-activity, that sometimes presents some challenges, is planned to be delivered to RL schools using mostly the school principals and US Peace Corps volunteers. CEOs and DEOs will also be trained in monitoring and coaching and although it is likely that the direct interaction with the teachers is quite limited, it would be best to direct their efforts to the treatment group until the endline data is collected. This should limit the risk of contamination of control schools

## **EGRA Instrument**

An EGRA tool used previously, in 2011 and 2015, will be used for the baseline data collection. This is normally the recommended approach because it allows comparability and also allowed to prepare the baseline data collection in a very short period of time. USAID/Washington raised some concerns about test security related to the EGRA instrument but it was decided that the risk is minimal and did not warrant the creation of new tools, given time and budget constraints.

## **G. WORKPLAN**

Below we present a detailed work plan schedule for activities for the baseline period of the impact evaluation. This workplan represents tentative dates for milestones based on the most current information NORC has available. These dates are subject to change as implementation details become available and pending receipt of necessary approvals for work to proceed from

---

<sup>8</sup> In some conversations with USAID it was mentioned the possibility of requesting the implementer to produce a roster of students in all the schools they will be treating every year. This should include enough data so we can track student transfers between control schools and treated schools (including all treated schools, not only the ones NORC will survey).



USAID and the University of Liberia's IRB review committee.

**Table 4: Workplan**

ACTIVITY	DATE	RESPONSIBLE PARTY	SUPPORT FROM
<b>Local firm selection</b>			
Identify shortlist of local firms for data collection/ Develop SOW/Interviews/budgets	Feb/March	NORC	
Submit final selection to USAID for consent	5-Apr	NORC	
USAID/Washington Finalize subcontract	14-Apr	USAID	
<b>Evaluation Design</b>			
Finalize EDR (USAID review due March 31)	14-Apr	NORC	
<b>Sampling</b>			
Send all necessary information to construct sample frame to NORC	28-Feb	USAID Liberia	USAID/W
Mission to submit selection criteria eligible schools to NORC	9-Mar	USAID Liberia	USAID/W
Create sample frame; select sample including back-up schools	24-Mar	NORC	USAID/Liberia
<b>Instrument development and testing</b>			
Obtain Tangerine programmed EGRA and supplemental assessments from RTI	5-Apr	STS	NORC
Review and make minor changes to EGRA and companion instruments	18-Apr	NORC	USAID/W
Program changes into Tangerine	20-Apr	NORC	STS
Finalize instruments	20-Apr	NORC	STS
<b>IRB and Other Approvals</b>			
Submit IRB application with draft instruments to NORC IRB	19-Mar	NORC	
Submit IRB application to Liberia authority	31-Mar	NORC	TKG
<b>Data collection</b>			
Organize enumerator training/training materials	Apr 10-14	NORC	STS
Develop QA procedures	Apr 1-10	NORC	
Train Enumerators for Full Data Collection	24-Apr – May 3	NORC	
Field Baseline	4-28 May	TKG	NORC
Data Quality checks during baseline data collection	4-29 May	NORC	STS/TKG
Weekly progress reports on data collection	4-29 May	NORC	TKG
<b>Data Prep, Analysis and Reporting</b>			
Baseline data set preparation	June	TKG/STS/NORC	
Analysis of baseline data	July	NORC	
Prepare Baseline Report	Aug	NORC	
Randomize schools into treatment and control, give list to USAID/Liberia	Aug	NORC	
Review of baseline report	July	NORC	
Final baseline report	Aug	USAID/Liberia	USAID/W
Prepare and submit public use datasets to USAID	Sep		
Dissemination	15-Sep TBD	NORC NORC	USAID/W USAID/Liberia/USAID/ W