



USAID
FROM THE AMERICAN PEOPLE

USAID/ETHIOPIA READING FOR ETHIOPIA'S ACHIEVEMENT DEVELOPED COMMUNITY OUTREACH (READ CO) PROGRAM

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

EVALUATION DESIGN REPORT

December 2015

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by NORC at the University of Chicago. 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/ETHIOPIA READING FOR ETHIOPIA'S ACHIEVEMENT DEVELOPED COMMUNITY OUTREACH (READ CO) PROGRAM

EVALUATION DESIGN REPORT

December 2015

PN 7617.007.01

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

CONTENTS

LIST OF ACRONYMS	1
A. PROJECT BACKGROUND	1
B. THEORY OF CHANGE	3
C. EVALUATION QUESTIONS AND OBJECTIVES	8
C.1 Final Outcomes	9
Research Question #1: Reading skills outcomes	9
C.2 Intermediate Outcomes	9
Research Question #2: Student motivation and behavior	13
Research Question #3: Parental engagement	14
Research Question #4: Program Take-Up	14
Research Question #5: School funds raised for reading-related activities	16
D. EVALUATION METHODOLOGY AND APPROACH	17
D.1 Research Question #1: Reading skills outcomes	17
D.2 Research QuesTions #2-5: Intermediate Outcomes	19
D.3 Power Calculations	19
For EGRA data	19
For the parent survey	21
D.4 Sample Selection and Experimental Design	22
School and Student Selection	22
Parent Selection	24
Teacher Selection and School Surveys	25
E. DATA COLLECTION APPROACH	26
E.1 Instrument Development	26
Early Grade Reading Assessment	26
Additional instruments	27
E.2 Selection of firm and recruitment of enumerators	27
E.3 Programming into Tablets and Training	27
E.4 Data Quality Control	28
E.5 Operational Plan for Data Collection	28
E.6 Workplan	28

LIST OF EXHIBITS

Figure 1. Evaluation Sampling Design	2
Figure 2. READ CO Theory of Change	7
Table 1: Evaluation questions.....	8
Table 2: Expected intermediate outcomes and indicators	10
Figure 3: MDES vs. Number of Schools, EGRA.....	20
Figure 4. MDES vs. parents per school, parent survey	21
Table 3. Final Sample Size.....	22
Figure 5. Sampling Process	25
Table 4: Early Literacy Skill, Subtask.....	26
Table 5: Workplan	29

LIST OF ACRONYMS

EDR	Evaluation Design Report
EGRA	Early Grade Reading Assessment
ETP	Education and Training Policy
GEAC	Girls Education Advisory Committee
ICC	Intra-Cluster Correlation
IE	Impact Evaluation
InEHD	Institute for Education, Health and Development
IQPEP	Improving the Quality of Primary Education Program
IR	Intermediate Result
IRR	Inter-Rater Reliability
KETB	Kebele Education and Training Board
KII	Key Informant Interview
MDES	Minimum Detectable Effect Size
MOE	Ministry of Education
NORC	NORC at the University of Chicago
PMP	Performance Monitoring Plan
PTSA	Parent Teacher Student Association
RCT	Randomized Controlled Trial
READ CO	Reading for Ethiopia's Achievement Developed Community Outreach
READ II	Reading for Ethiopia's Achievement Developed Institutional Improvement
READ M&E	Reading for Ethiopia's Achievement Developed Monitoring and Evaluation
READ TA	Reading for Ethiopia's Achievement Developed Technical Assistance
RSEB	Regional State Education Bureau
RTI	Research Triangle Institute
SC	Save the Children
SIP	School Improvement Plan
STS	School-to-School International
SUR	Seemingly Unrelated Regressions
T1	Treatment Arm 1
T2	Treatment Arm 2
TOT	Training of Trainers
USAID	United States Agency for International Development

A. PROJECT BACKGROUND

The USAID Reading for Ethiopia's Achievement Developed Community Outreach Project (READ CO) is a four-year project implemented by Save the Children (SC) and designed to support the Ministry of Education (MOE) and the Regional State Education Bureaus (RSEBs) to implement a comprehensive community and school-based early grade reading program that complements the READ Technical Assistance (READ TA) and READ Institutional Improvement (READ II) projects. READ CO strives to improve early grade reading achievement in 2400 schools by increasing students' access to print materials, offering more time for them to practice reading, and fostering increased parental, school, and community engagement in support of children's improved reading. Major activities of READ CO include building government and private-sector capacity to provide supplementary reading materials, school and community reading activities, and training for teachers and parent teacher student associations (PTSAs) on how to manage and promote these activities effectively.

A notable component of the READ projects is its focus on mother tongue-based instruction, in line with the Government of Ethiopia's Education and Training Policy (ETP) introduced in 1994 that mandates the use of mother tongue in primary education (Grades 1-8). English will be the medium of instruction only starting in secondary education but is taught as a subject beginning from grade 1. The READ projects (READ TA, READ CO, READ II) will create materials and enhance teacher's instruction in seven mother tongues covering 5 regions of Ethiopia: Amharic, Tigrinya, Afaan Oromo, Af-Somaali, Sidama Afoo, Wolayttatto, and Hadiyyisa.

NORC at the University of Chicago, through the USAID Reading and Access Evaluation Contract, has been charged with conducting the external impact evaluation (IE) of READ CO. The IE uses a Randomized Controlled Trial (RCT) design in which a subsample of the 2400 target schools are randomly assigned into one of two treatment groups or the control group. The first treatment will provide reading materials, training and coaching inputs for school-based reading support activities to school directors, teachers, and youth volunteers. The second treatment group will provide the inputs of the first treatment group, plus training and coaching support to Parent Teacher Student Associations (PTSAs), Girls Education Advisory Committees (GEAC) and Kebele Education and Training Boards (KETB) to promote parental and community engagement in both school- and community-based reading activities and increase the level and quality of parent-child engagement in reading activities at home.¹ Both treatment arms will include coaching and mentoring of the training participants in facilitation strategies, supporting students, monitoring children's reading activities and motivating learners; the first treatment focusing on school-based reading activities and the second treatment focusing on both school-based and community-based reading activities.

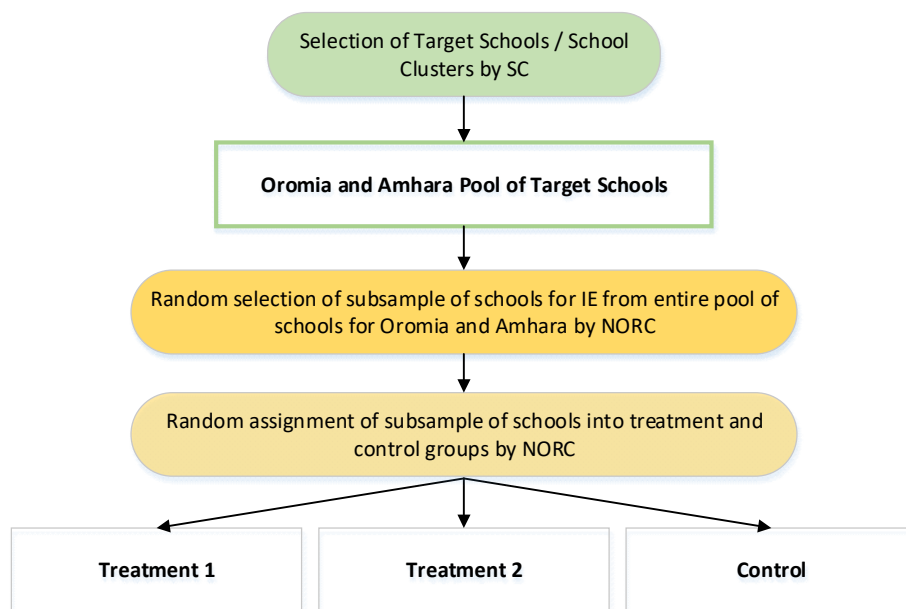
The main purpose of the IE will be to assess the causal impact of READ CO on reading outcomes of primary school children. The evaluation will measure reading outcomes using the Early Grade Reading Assessment (EGRA), a widely used tool to measure various aspects of

¹ READ CO Program Description – Revised Oct 12 2015.

reading proficiency. Additionally, the IE will also assess READ CO’s impact on other intermediate outcomes such as participation of children in reading activities at school and in the community/home, time spent reading independently and parental engagement in support of children’s reading development. To measure these outcomes, the evaluation will use four types of survey instruments to collect data at the level of the teacher, school, student, and student’s household respectively. The IE will be implemented in the regions of Amhara and Oromia, and focus on the Amharic and Afaan Oromo languages. It will assess the impact of READ CO over a period of approximately two years².

A graphical representation of the overall IE design is presented below. A more detailed figure appears later in this report.

Figure I. Evaluation Sampling Design



In section II, we describe our understanding of READ CO’s Theory of Change and list possible indicators to measure. Section III describes our main evaluation questions and objectives; Section IV details the evaluation methodology; Section V presents the sampling approach and finally Section VI describes the data collection approach including development of instruments, training of enumerators and data collection plans.

² The timing of the endline and a potential midline are still under discussion with USAID.

B. THEORY OF CHANGE

The evaluation questions and design are informed by the Theory of Change model presented in Figure 1 which is based on the READ CO PMP, Program Description and other documents received by NORC from SC. The Theory of Change presents the causal pathways by which the components of the READ CO intervention shown in the “Inputs” column lead to the goal of improved early grade literacy skills shown in the last column. To measure each of the outcomes, several indicators have been proposed.

The “Inputs” column describes the two treatment arms that READ CO will implement for the IE. Below we provide more information on these inputs based on documents received by READ CO as well as conversations with the READ CO Chief of Party (COP) for both treatment arms. The communities in Treatment Arm 1 (T1) will receive school-based inputs only. These school-based inputs consist of:

- Establishment of reading corners in classrooms (G1-4)
- Establishment of reading/writing clubs specifically targeting struggling learners (but open to all) facilitated by youth volunteers (G1-4, G5-8).
- Curriculum-aligned reading materials including decodable texts for G1-2, leveled readers for G1-8 and read-aloud texts for G2-4.
- Reading corner kits that include mats for sitting, letter card sets, slates, story books and read-alouds.
- 2-day training workshops for school directors, teachers and youth volunteers on facilitating school-based reading activities, supporting learning, promoting independent reading and taking books home.

The communities in Treatment Arm 2 (T2) will receive all school-based inputs from Treatment arm 1 in addition to community based inputs as follows:

- Establishment of community book banks (G1-8)
- Reading/writing camps (G1-8).
- Book bank kits that include curriculum-aligned reading materials story books, mats for sitting and a sturdy container.
- 2-day training workshops on facilitating community-based reading activities, supporting learning, promoting independent reading with the use of book banks and taking books home. These additional trainings will be administered to Parent Teacher Student Associations (PTSAs), Girls Education Advisory Committees (GEAC) and Kebele Education and Training Boards (KETB), school directors and youth volunteers and will also promote school improvement plan development to support reading and mobilization of existing resources.

The READ CO Performance Monitoring Plan (PMP) describes three Intermediate Results (IR). We address each of these IRs in turn below.

IR 1: ACCESS TO SUPPLEMENTARY READING MATERIALS IN VERNACULAR LANGUAGES INCREASED

One of the main goals of READ CO is to enhance the school and home literacy environment by developing and providing supplementary reading materials which are aligned with the new reading curriculum implemented in 2014 and developed in the mother tongue of the students. To make these materials available to students, READ CO will set up reading corners in classrooms (T1 schools) as well as book banks in the community (T1 and T2 schools)³. These supplementary materials will also be made available to students through school (T1) and community-based reading activities (T1 and T2) such as reading camps, reading buddies programs, read-a-thons, and reading competitions among others.

We would thus expect to observe (1) enhanced literacy environment both at school for T1 schools and in the community/at home for both T1 and T2 schools, which can be measured with the proportion of schools/communities that have a reading corner/book bank as well with the actual number and type of supplementary reading materials available in these reading corners and book banks; and (2) increased access to and use of these supplementary reading materials, which can be measured with the proportion of students who report taking books home from the school reading corner or the community book bank.

IR 2: OPPORTUNITIES FOR CHILDREN TO PRACTICE READING AND WRITING CREATED

We hypothesize that with more interesting school-based and community-based reading activities, students will have more opportunities to practice reading and writing and engage more in these activities as can be measured by student's participation rate in both school-based and community-based reading activities. As a result, students will spend more time reading and writing independently, thus increasing their "time on task".

Furthermore, they will be more engaged in the classroom and more motivated to learn to read. With this increased motivation, students may be less likely to miss school leading to higher attendance rate which could also lead to increased time on task and improved reading skills.

³ Both reading corners and community book banks will be established directly by READ CO through the provision of materials (reading corner and book bank kits) that will include a specific set of reading materials as well as containers, mats, etc. Reading corners are established in classrooms and are therefore school-based inputs (and part of Treatment 1) while the book banks are community-based (and therefore only part of Treatment 2). In addition to reading corners and community book banks, it is possible that reading rooms will be established at the school. Reading rooms are akin to libraries, targeting specifically G1-G4 students. However, unlike reading corners and community book banks, READ CO will not establish reading rooms directly. Instead, it will provide technical assistance to communities (to PTSAs in particular) to develop school improvement plans and budgets (through capitation grants) to establish these reading rooms. In other words, reading rooms will be established by the community with only technical support from READ CO and no direct provision of materials or financial contribution.

IR 3: PARENTAL AND COMMUNITY ENGAGEMENT TO SUPPORT LEARNING TO READ STRENGTHENED

Finally, given that READ CO's focus is on promoting literacy throughout the community, we expect increased parental and community engagement in the form of participation in and facilitation of school-based and community-based reading activities. READ CO intends to train youth volunteers in Treatment 1 and Treatment 2 schools to facilitate school-based reading/writing clubs targeting struggling learners specifically (although these camps will be open to all students) and which will be supported by school teachers. In addition in Treatment 2 schools, READ CO intends to train community members (PTSA, GEAC, KETB) as well as school directors and youth volunteers in the facilitation of community-based reading activities such as read-a-thons and reading camps⁴, and will encourage parents' participation in reading activities at home using strategies promoted by READ CO. We also expect parents and the community to participate in the development of School Improvement Plans and in fundraising efforts mobilizing existing resources and volunteers⁵. The funds raised would be used for acquiring additional learning materials (e.g. for school reading rooms). All of these activities and materials would in turn also enhance the literacy environment at school and in the communities.

ADDITIONAL INTERMEDIATE RESULTS

Finally, we also hypothesize that intermediate outcomes on teachers might range from higher rates of school attendance to more effective use of classroom time during reading instruction as well as better and increased facilitation of school-based reading activities such as reading/writing clubs, reading corners and reading buddy programs. Furthermore, trainings in facilitation skills, strategies for engaging and motivating learners and for monitoring children's reading activities would first improve teacher knowledge and practice, as a means to improve student literacy outcomes. Thus, we would expect that teachers will demonstrate higher quality literacy instructional practices⁶, as reported by students and as observed in the classroom.

ADDITIONAL NOTE

Although Treatment 1 is broadly designed as only containing school-based inputs while Treatment 2 includes community-based inputs in addition to Treatment 1 inputs, we understand that in practice there may be some form of parental engagement and community engagement in both treatment arms. For instance, youth volunteers will be trained in the

⁴ Reading and writing clubs are school-based activities facilitated by youth volunteers and supported by teacher; they will be meeting on a regular basis (e.g. every weekend) and while open to all, will specifically target struggling learners as priority (these struggling learners will be identified by the youth volunteers, parents, teachers). On the other hand, reading camps are community-based activities and less formal/structured in nature. They are voluntary activities that the parents/community will organize at a specific point in time (e.g. during school vacation). In both reading/writing clubs and reading camps, students will be able to access reading materials.

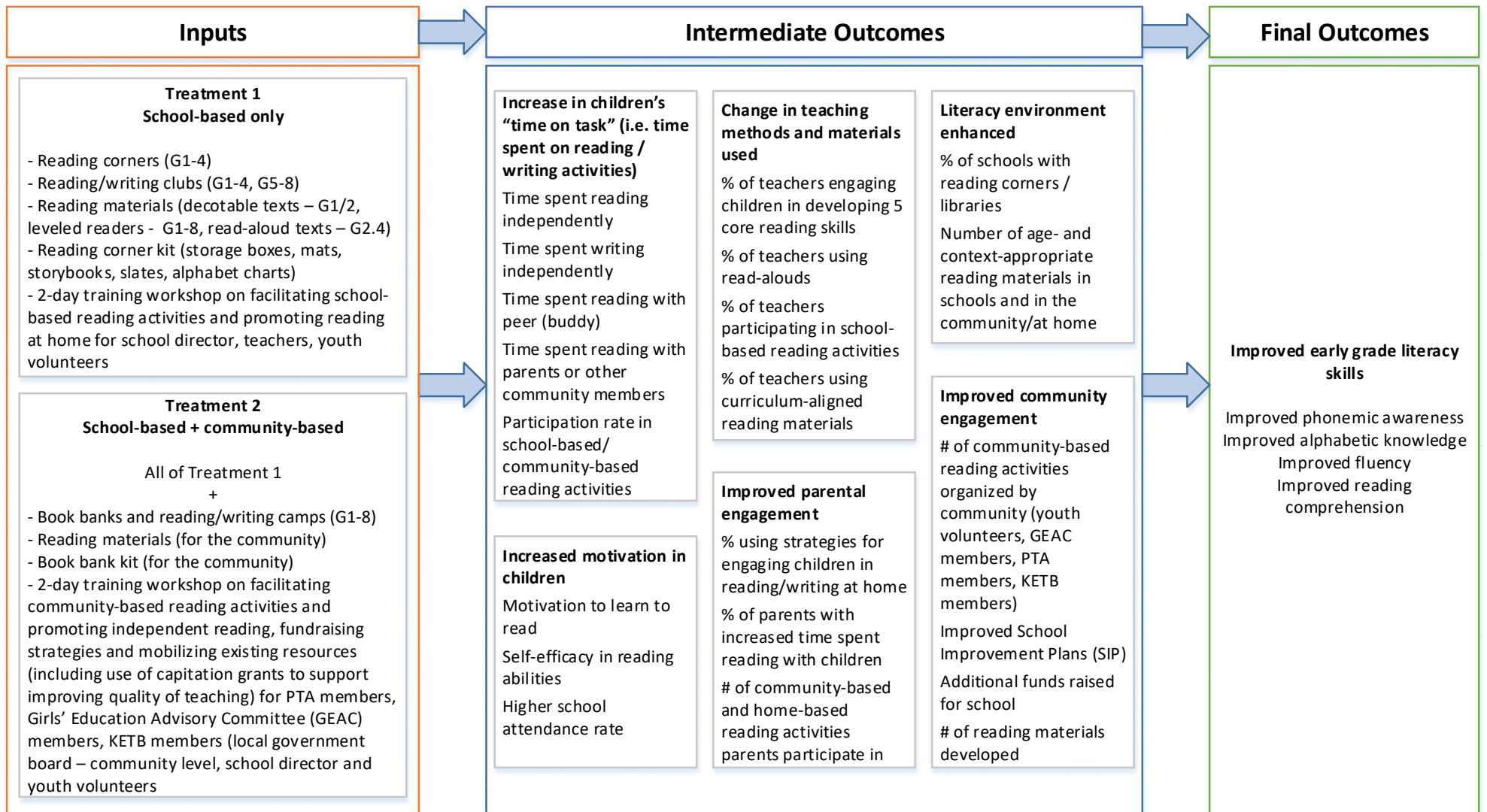
⁵ See Footnote 3.

⁶ These instructional practices may include: asking students to sound out words, introducing new vocabulary to students, asking students to retell a story they read, etc. However, at the time of writing this Evaluation Design Report, the exact content of the training curriculum has not yet been defined, therefore inclusion of a comprehensive list of specific strategies is not possible.

facilitation of school-based reading and writing clubs, and PTAs will be provided technical support in the development of School Improvement Plans which may lead to acquiring reading materials that will be used for establishing reading rooms in the schools. At the time of writing of this Evaluation Design Report, the READ CO implementation team was still finalizing its set of activities and specific training curriculum content.

As a result, the theory of change model presented in Figure 2 below does not distinguish separate causal paths for Treatment 1 and Treatment 2 at this point in time. Instead we provide a general theory of change framework that describes intermediate and final outcomes more globally.

Figure 2. READ CO Theory of Change



C. EVALUATION QUESTIONS AND OBJECTIVES

READ CO is primarily intended to improve reading skills for the students who participate in the program. The main focus of the evaluation is thus to measure the impact of READ CO on improving reading outcomes, as measured by the Early Grade Reading Assessment (EGRA) tool that will be administered to students. In addition, the evaluation will also consider several questions related to the intermediate outcomes presented in the Theory of Change, which will be measured using a variety of data sources.

The main evaluation questions are listed in Table I. Each evaluation question is discussed in depth below.

Table I: Evaluation questions

#	Outcomes	Questions
1	Reading skills	<ul style="list-style-type: none"> • Q1A: What was the impact of READ CO on learners' reading skills for Treatment Arm 1 as compared to Control? What was the impact of READ CO on learners' reading skills for Treatment Arm 2 as compared to Control? • Q1B: Did Treatment Arm 2 cause a greater improvement in reading skills for students as compared to Treatment Arm 1?
2	Student Motivation and Behavior	<ul style="list-style-type: none"> • Q2A: What was the impact of READ CO on increasing time/frequency that students spent reading independently, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control? • Q2B: Did Treatment Arm 2 cause a higher increase in the time/frequency that students spent reading independently as compared to Treatment Arm 1? • Q2C: What was the impact of READ CO on student engagement, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control? • Q2D: Did Treatment Arm 2 cause a larger improvement than Treatment Arm 1 in student engagement?
3	Parental engagement	<ul style="list-style-type: none"> • Q3A: What was the impact of READ CO on parental engagement, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control? • Q3B: Did Treatment Arm 2 cause a larger improvement than Treatment Arm 1 in parental engagement?

4	Program take-up (participation in reading activities and use of reading materials/trainings)	<ul style="list-style-type: none"> • Q4A, option 1: To what extent are teachers utilizing the READ CO curriculum materials, reading corner kits and recommended practices in their facilitation of school-based reading activities? • Q4A, option 2: To what extent have teachers changed their literacy instruction practices in general? • Q4B: What proportion of students in Treatment Arm 1 have participated in the school-based reading activities? What proportion of students in Treatment Arm 2 have participated in the community-based reading activities?
5	School funds raised for reading-related activities	<ul style="list-style-type: none"> • Q5A: Have schools in Treatment Arm #2 successfully obtained additional funding, and if so to what extent has it been used on reading-related activities?

C.1 FINAL OUTCOMES

Research Question #1: Reading skills outcomes

The reading skills outcomes that are the main focus of the evaluation are reflected in the following evaluation questions:

- Q1A. What was the impact of READ CO on learners’ reading skills, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control?
- Q1B. Did Treatment Arm 2 cause a greater improvement in reading skills for students as compared to Treatment Arm 1?

Our approach to addressing this question takes cognizance of educational research on literacy and measurement of outcomes. The consensus among the reading research community in the United States is that effective reading instruction attends to *at least* five main reading skill areas including alphabetics (letter knowledge and phonemic awareness), fluency, vocabulary, and comprehension and phonics (National Reading Panel, 2000; Snow, Burns & Griffin, 1998).⁷ Based on this research, the Early Grade Reading Assessment (EGRA), a brief oral reading assessment that tests these skills, will be used to measure program impacts on literacy (RTI International, 2007). Because our impact evaluation will be focused on Oromia and Amhara, we will administer the EGRA in Oromo and Amharic.

C.2 INTERMEDIATE OUTCOMES

In addition to assessing the impact of READ CO on literacy outcomes, the Evaluation will also consider a number of intermediate outcomes along the causal chain as shown in the Theory of Change. Investigating intermediate outcomes can provide valuable programmatic lessons by assessing the extent to which different mechanisms and associated program components have contributed to the observed impacts on the final outcomes. In cases where the impact of the

⁷ These five skills are not meant to be all inclusive; however, considerable empirical research has been conducted in these skill areas that has indicated they are important predictors of reading.

program on final outcomes is less than anticipated, investigating intermediate outcomes can help to shed light on what went wrong.

Table 2 illustrates the intermediate outcomes that the evaluation could consider, along with the level at which each outcome occurs and associated research questions based on Table 1, as well as proposed measurement questions⁸, and the type of instrument that will be used to collect the data.

Table 2: Expected intermediate outcomes and indicators

Level and Associated Research Question	Indicator	Measurement Question	Expected Outcomes	Instrument(s) used
1. Student: Research Question #2	1.1 Attendance	How many days did [student] come to school last week?	Decrease in reported absences	Student questionnaire Parent questionnaire
	1.2 Motivation to Read	Is [student] interested in reading?	Increase in reported motivation to read.	Student questionnaire
	1.3 Self-efficacy	Does [student] consider reading easy? Does [student] think practicing will help [him/her] become a better reader?	Increased self-efficacy	Student questionnaire
	1.4 Time spent reading	How often/ how much time does [student] spend reading by themselves?	Increased frequency/duration of time spent reading independently	Student questionnaire Teacher questionnaire Parent questionnaire
	1.5 Participation in reading activities	Does [student] participate in school-based or community-based reading activities? If yes, which ones? How often?	Increased participation in school-based and community-based reading activities Increase in time spent reading	Student questionnaire Parent questionnaire READ CO PMP Data (Indicator #14)

⁸ The final formulation of the questions in each of these instruments will differ slightly as questionnaires are revised through the process of pre-testing, enumerator training, field piloting and translation.

Level and Associated Research Question	Indicator	Measurement Question	Expected Outcomes	Instrument(s) used
	1.6 Use of materials	Does [student] bring home reading books from the classroom or from the school library/book bank to read at home? What was the name of the last book [student] took home?	Increase in taking books home / improvement in home literacy environment	Student questionnaire Parent questionnaire READ CO PMP Data (Indicator #7)
2. Teacher/School Research Question #4	2.1 Attendance	Is there any day [teacher] did not come to school last week?	Decrease in reported absences	Student questionnaire Teacher questionnaire
	2.2 Literacy Instruction	Does [teacher] use READ CO materials when he or she is teaching reading?	Increase in access to materials during reading lesson	Teacher questionnaire
		During the literacy hour or reading activities, does [teacher] use specific reading instruction strategies such as (but not limited to): <ul style="list-style-type: none"> - Having the class repeat sentences teacher said - Having students retell a story they read - Having students sound out unfamiliar words - Having students learn meaning of new words - Having students read aloud to teacher and other students - Have students read on their own during school time 	Increase in effective literacy instruction activities	Teacher questionnaire Classroom observation (nature and frequency TBD)

Level and Associated Research Question	Indicator	Measurement Question	Expected Outcomes	Instrument(s) used
		- Have students play games related to the lesson		
	2.3 Organization of and participation in school-based reading activities	Is there an increase in the type/number of reading activities at school? Do teachers and their students participate in these activities?	Increase in school-based reading activities / increase in opportunities for children to read independently	Teacher questionnaire School questionnaire READ CO PMP Data (Indicator #13)
	2.4 Availability of materials	Does the class have a reading corner? If yes, what type of materials are available? Does the school have a reading room and/or library? If yes, what type of materials are available?	Increase in availability of curriculum-aligned, age-appropriate, context-appropriate reading materials at school	Teacher questionnaire School questionnaire
3. Community / parents Research Question #3	3.1 Engagement in reading	Does anyone at home read with [student]? ⁹ Does anyone at home read with [student] using specific strategies promoted by READ CO? Does anyone in the community read with [student]?	Increase in reported read aloud at home	Student questionnaire Parent questionnaire

⁹ This indicator may be measured by frequency, type of reading, and motivation to read as well. For example, children may be asked with what frequency they engage in different types of reading activities outside of school (e.g. read aloud, listen to someone read aloud, talk about books, read for fun, etc.). They may also be asked to what extent they agree with questions such as: “I enjoy reading” or “Reading is boring”.

Level and Associated Research Question	Indicator	Measurement Question	Expected Outcomes	Instrument(s) used
Research Question #5		Does anyone at home read newspapers, religious texts or books?	Increase in reported observations of home reading	Student questionnaire Parent questionnaire
	3.2 Perception of importance of literacy	Do parents think reading with their children is important/useful?	Increased community and parental engagement in literacy development	Parent questionnaire READ CO PMP Data (Indicator #16 and #17)
	3.3 Community-based reading activities	Do community/parents participate in reading activities such as reading/writing camps, read-a-thons, book banks?		
	3.4 Availability of materials	Does the community have a library/community book bank? If yes, what type of materials are available?	Increase in availability of curriculum-aligned, age-appropriate, context-appropriate reading materials in the community	
	3.5 School Improvement Plans (SIPs)	Do community/parents participate in developing SIPs?		
	3.6 Fundraising and use of funds	How much money was raised for the school? How was this money used?	Improved home/community literacy environment	

On the basis of the outcomes shown in table 3, we identify the following evaluation questions related to intermediate outcomes that the evaluation could consider:

Research Question #2: Student motivation and behavior

An important intermediate outcome that the evaluation will assess is the extent to which participating in READ CO is associated with students spending more time reading independently:

- Q2A: What was the impact of READ CO on increasing time that students spent reading independently, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control?
- Q2B: Did Treatment Arm 2 cause a higher increase in the time that students spent reading independently as compared to Treatment Arm 1?
- Q2C: What was the impact of READ CO on student engagement, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control?

- Q2D: Did Treatment Arm 2 cause a larger improvement than Treatment Arm 1 in student engagement?

Data to address these evaluation questions will come primarily from the Parent Questionnaire. A sample of parents of students in the treatment and control groups will be asked a series of questions about the frequency and amount of time that their children spend reading and frequency of reading episodes as well as about use of reading materials. As a supplemental approach to the analysis, the evaluation will include similar questions in the student questionnaire, as these questions can be included with minimal cost implications. However, an important disadvantage here is that the reliability of the responses from students is expected to be relatively low, due to the age of the respondents and the high potential for social desirability bias when asking students about the amount of time they spend reading. Thus, the analysis for these questions will focus on the parent survey data, with the student survey data serving as a means of cross-checking the findings. For more details on associated indicators, see Table 2, Indicators 1.1 through 1.6.

Research Question #3: Parental engagement

The evaluation will also measure changes in parental engagement in supporting literacy development.

- Q3A: What was the impact of READ CO on parental engagement, for both Treatment Arm 1 and Treatment Arm 2 as compared to Control?
- Q3C: Did Treatment Arm 2 cause a larger improvement than Treatment Arm 1 in parental engagement?

A challenge in assessing this outcome is that engagement can be difficult to measure. Careful attention was devoted to developing an appropriate set of survey questions and piloting the questionnaire with technical input from the READ CO team in order to ensure that questions are aligned with the READ CO training content. Questions include perception of importance of literacy, frequency and amount of time parents or other household members read with their children at home, participation in community-based reading activities, and use of specific strategies and skills taught by READ CO to support children's reading. For more details on associated indicators, see Table 2, Indicators 3.1 through 3.3.

Research Question #4: Program Take-Up

Another set of outcomes relates to whether the services that READ CO is offering are being utilized as intended. This includes the extent to which READ CO teacher trainings and materials are translating into changes by teachers in the number and quality of school-based reading activities that they organize and facilitate, as well as the extent to which students are participating in the extra-curricular Treatment Arm 2 activities.

Assessing how teachers' behavior has changed as a result of READ CO materials could be done in one of two ways. First, the evaluation could identify a specific set of materials and practices associated with READ CO and consider the extent to which teachers in the two Treatment Arms have adopted them. For example, the evaluation will consider the amount of time spent in class on reading-related activities, and the extent to which teachers are following strategies promoted by the READ CO training curriculum during school-based reading activities:

- Q4A, option 1: To what extent are teachers utilizing the READ CO curriculum materials, reading corner kits and recommended practices in their facilitation of school-based reading activities?

Alternatively or in addition, the evaluation could identify and measure impact on a more general set of classroom practices that READ CO would be expected to influence, rather than the specific outputs associated with READ CO, as follows:

- Q4A, option 2: To what extent have teachers changed their literacy instruction practices in general?

While the focus of READ CO is on improving school-based reading activities rather than on improving literacy instruction itself (which is the focus of the READ TA project), the teachers will still receive mentoring and coaching on how to engage and motivate students in reading and writing as well as on how to monitor their student's reading activities. It could be that these practices are used during the school-based reading activities as well as more broadly speaking during their more formal literacy instruction. Further analysis and discussion with the READ CO team and USAID in the course of finalizing the design report and as the intervention is being rolled out will establish the most appropriate formulation of Q4A.

Changes in teacher practices could be assessed using a process of systematic classroom observation¹⁰ to gather data on classroom and other school-based reading activities, or alternatively by interviewing teachers at midline or endline to determine changes to their teaching practices¹¹. It is important to bear in mind that these outcomes are at the school level, and the number of schools included in the RCT is not sufficient to make statistical comparisons between the treatment and control groups. Thus, the analytical approach for these outcomes would be a pre-post analysis of teachers in the two treatment arms. For more details on associated indicators, see Table 2, Indicators 1.4, 2.2 and 2.3.

The evaluation will also consider the extent to which students are participating in the school-based and community-based reading activities, as follows:

- Q4B: What proportion of students in Treatment Arm 1 have participated in school-based reading activities? What proportion of students in Treatment Arm 2 have participated in community-based portions of the intervention, including book banks and reading and writing camps?

While Treatment Arm 1 includes only school-based interventions that will apply to all students, Treatment Arm 2 includes additional activities that each individual student may or may not participate in. In assessing the effectiveness of Treatment Arm 2 and the contribution of the different components to impacts, it is important to understand the extent to which students have participated in these activities. While questions related to participation in school and community-based reading activities will be included at baseline in the student and parent questionnaire, we expect that more useful information will be gathered at midline/endline once

¹⁰ The frequency and timing of these classroom observations are yet to be determined.

¹¹ While the baseline teacher questionnaire will include questions related to reading instruction, these questions will be refined and more targeted at midline/endline once the training curriculum is more specific and trainings have been conducted.

implementation of the intervention has started. For more details on associated indicators, see Table 2, Indicators 1.5, 2.3, 3.3.

Research Question #5: School funds raised for reading-related activities

Treatment Arm #2 includes workshops to assist schools in increasing the reading-related focus of their School Improvement Plans (SIPs), and to raise funds and mobilize existing resources to implement these activities. The following evaluation question assesses this component:

- Q5A: Have schools in Treatment Arm #2 successfully obtained additional funding, and if so to what extent has it been used on reading-related activities?

To answer this question, we will use the READ CO PMP data (Indicator # 16). In addition, data will be collected using Key Informant Interviews (KIIs) with school officials in each Treatment Arm 2 school. These KIIs will obtain information on overall levels of school funding before and after READ CO, as well as the levels of school spending on different types of activities. The KII data will be corroborated with official records if it is feasible to obtain and interpret them. Further investigation in the course of finalizing the Design Report will be needed to determine whether such records are available as well as the appropriate structure of the KII questions to categorize different types of spending and identify reading-related expenditure categories.

The analysis will be a descriptive summary of the data that will aid in the interpretation of the findings related to the other evaluation questions concerning Treatment Arm #2. If schools were able to increase funding for reading-related activities, that suggests that observed positive impacts of Treatment #2 may have occurred through this channel. Conversely, if Treatment Arm 2 schools did not increase funding for school-related activities, then any observed impacts of the additional Treatment 2 should be attributed to other components. For more details on associated indicators, see Table 2, Indicators 3.5 and 3.6.

D. EVALUATION METHODOLOGY AND APPROACH

D.1 RESEARCH QUESTION #1: READING SKILLS OUTCOMES

The methodology to address the reading skills outcomes under evaluation question I will use a randomized controlled trial (RCT) set-up, with schools randomly assigned to treatment and control groups. Such randomized methods are widely considered the “gold standard” in impact evaluation, for example the USAID Evaluation Policy states: “Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured.”¹² 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.

For the EGRA-related outcomes related to Evaluation Question I, data will be collected in two rounds: a baseline prior to program implementation, and an endline two to three years after the conclusion of the program¹³. Ideally, we will collect student-level panel data by administering EGRA to the same individual students at both baseline and endline¹⁴. The main advantage of a student-level panel is that it would allow for student-level fixed effects, which can improve the power and precision of the analysis if baseline outcomes are highly correlated with endline outcomes. However, information obtained in the early stages of data collection suggests that re-interviewing the same students at endline may not be feasible. Based on initial visits to the school, rates of absenteeism based on school rosters tended to range from 20%-35%. In addition, data from EMIS suggests that we could expect a drop-out rate from baseline to endline of approximately 20%. Thus, without repeated visits to each school we would only expect to be able to re-interview approximately half of the baseline students. While we will continue to explore this possibility, the alternative would be a repeated cross-section at the school level, in which one random sample of students would be drawn from each class at baseline and a second random sample at endline.

We will estimate the impact of the program on EGRA scores using regression models of the following form:

$$Y_{ijkt} = \beta_0 + \beta_1 \delta_{time} + \beta_2 X_{ijkt} + \gamma_1 (\delta_{time} \times \delta_{treat1}) + \gamma_2 (\delta_{time} \times \delta_{treat2}) + \alpha_i + \varphi_j + \varepsilon_{ijt}$$

Where:

Y_{ijkt} is the EGRA score component for student i in school j in region k at time t ,

δ_{time} is a dummy variable equal to one at endline,

¹² USAID Evaluation Policy p. 2

¹³ A midline round of data collection and analysis one year after the baseline is also being considered by USAID

¹⁴ Based on EMIS data, we would expect an attrition of about 20%.

δ_{treat1} and δ_{treat2} are dummy variables equal to one for observations in Treatment Arms 1 and 2 respectively,

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

α_i is student-level fixed or random effect,

φ_j is a school-level fixed or random effect,

ε_{ijt} is a random error term, and

β_0, β_1 , the vector β_2 and γ are parameters to be estimated

Our measures of program impact are γ_1 and γ_2 , the coefficients on the interaction term between the time and treatment dummies. Note also that the models will include a set of control variables \mathbf{X}_{ijkt} 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, whether the parents read to the child at home regularly, parents' education/literacy, sex of the household head, and some basic questions about household assets to estimate wealth status. Incorporating these control variables into the analysis will improve the precision of the impact estimates.

The values of γ_1 and γ_2 will provide estimates of the impact of Treatment 1 and Treatment 2 respectively relative to the control group on EGRA scores. In addition, a statistical hypothesis test for equality between γ_1 and γ_2 will determine whether there is a statistically significant difference between the impacts of Treatment 1 and 2. We will test the impact on each EGRA subtask. For the subtasks, the analysis will use a Seemingly Unrelated Regressions (SUR) framework to account for correlation between error terms in the individual regressions.

In addition to estimating the overall impact of READ CO on reading outcomes, we will also carry out a sub-group analysis to examine whether the impacts were different for sub-groups of interest such as girls, poor households, households with more educated parents, and other types of households. The data to identify the sub-groups will be based on parent surveys that will be administered to the parents of a sub-sample of students at each school, as well as a brief student survey that will be conducted in conjunction with the EGRA.

Finally, we note two additional methodological issues that the team will assess once the data is available. First, we expect that an individual student's scores on the different subtasks of the EGRA may be strongly correlated. If this is the case, the econometric methods should take this into account, for example by using the Seemingly Unrelated Regressions (SUR) framework, which allows the error terms in models of different outcomes to be correlated. Secondly, the sampling approach has been designed to carry out two independent estimations of the impact of READ CO on EGRA scores for Amhara and Oromia respectively. The reason for doing the two estimations separately is that combining EGRA scores across different languages is thought to be problematic. However, it may be possible to increase the power for both the analysis of the impact of READ CO on EGRA scores and the subgroup analysis by pooling the data and using a modeling approach that accounts for differences across languages. The evaluation team will investigate both of these issues once the midline or endline data is available.

D.2 RESEARCH QUESTIONS #2-5: INTERMEDIATE OUTCOMES

For the research questions #2-5, the analysis will use different methods depending on the particular question. For question #2 – Student motivation and behavior – and question #3 – Parental Engagement, the approach will follow rigorous impact evaluation methods in order to evaluate the causal impact of the program on these outcomes. Data will be collected using parent surveys at baseline and endline from samples of parents in Treatment Arm 1, Treatment Arm 2, and the control group, and the impact will be measured using a regression model similar to the one described in the previous section. For question #4 – Program Take-Up, the analysis of the outcomes related to uptake of materials and participation by students in extra-curricular activities will be a descriptive analysis based on data from the treatment areas only, intended to help explaining the findings related to reading skills outcomes, and the analysis of changes in classroom activities may be based on teacher interviews and/or classroom observations using descriptive analysis to assess changes over time. Finally, question #5 – School Funds Raised for Reading-Related Activities will also use a descriptive analysis based on Key Informant Interviews and school records.

D.3 POWER CALCULATIONS

This section provides the details of the calculations used to arrive at the recommended sample sizes for the EGRA and the household survey that will be used to evaluate intermediate outcomes. 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. In practice, calculating the required sample size requires choices and assumptions about a range of these parameters. We carry out the power calculations using the Optimal Design software package, which can perform these calculations under a variety of scenarios.

For EGRA data

As described above, we consider the impact evaluations of the program in Amhara and Oromia as two separate evaluations and calculate the required sample size for each. Calculating the minimum required sample size requires a number of assumptions about parameters as follows:

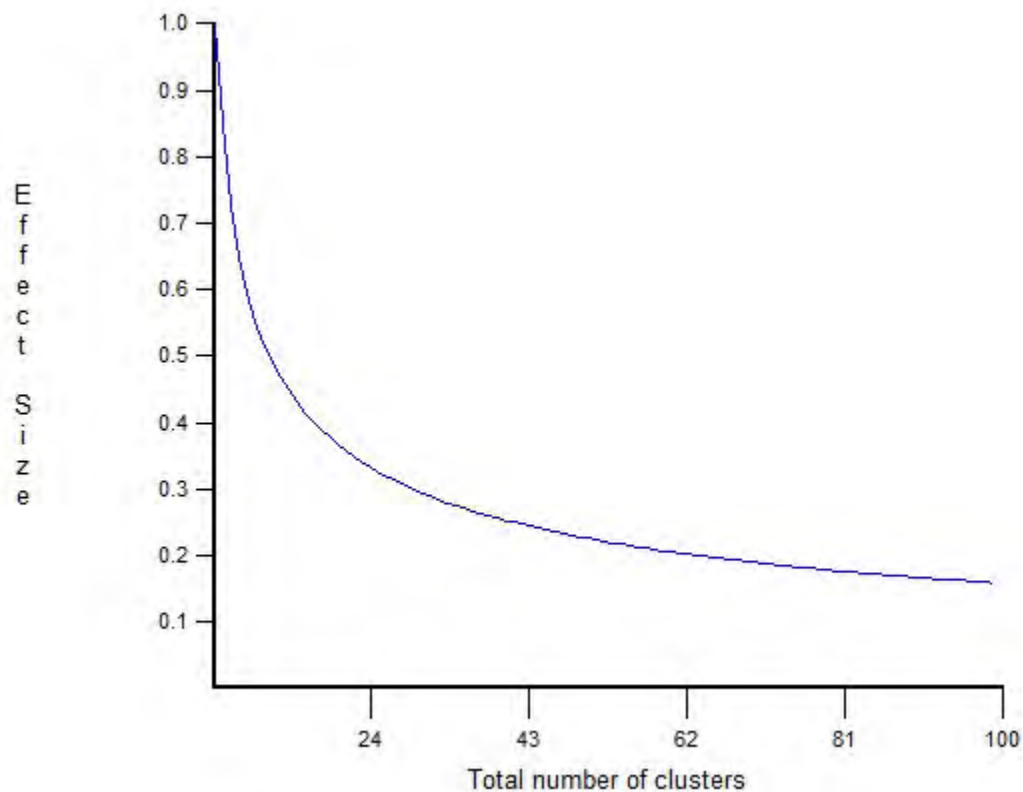
- α is the significance level of the test, or probability of Type I error. We use the standard value of 0.05.
- β is the power of test, where $(1-\beta)$ is the probability of Type II error. We use the standard value of 0.8.
- ρ is the intraclass correlation coefficient, or ICC. The ICC is a measure of the extent to which variation in outcomes is due to cluster-level factors as opposed to individual-level factors. In our case, the ICC reflects the extent to which differences in EGRA scores are due to differences between schools, rather than differences between individual students. Based on results from previous studies, we assume an ICC of 0.10.
- 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 the intertemporal correlation between

outcomes values at baseline and follow up. For this parameter, we assume an approximate value of 0.3.

- MDES is the minimum detectable effect size. The MDES is the smallest impact of the program on the outcome variable that the evaluation will be able to detect. The MDES is an important parameter in power calculations, as the choice of its value can have a substantial impact on the required sample size. Ideally, the value of the MDES is chosen based on previous studies that have considered the impact of similar interventions on the same outcome variables. While we are not aware of any cases of impact evaluations of very similar interventions to READ CO, our review of previous studies in sub-Saharan Africa that use EGRA scores as an outcome variable indicates effect sizes in the range of 0.17-0.35.

Figure 3 plots the MDES against the minimum required number of schools in the sample for each of the two treatment groups plus the control group. We assume a cluster size of 30 students per school¹⁵.

Figure 3: MDES vs. Number of Schools, EGRA



Taking into consideration effect sizes in previous studies as well as the budgetary implications, NORC and USAID agreed upon a MDES of 0.23, which requires a sample size of 25 schools in

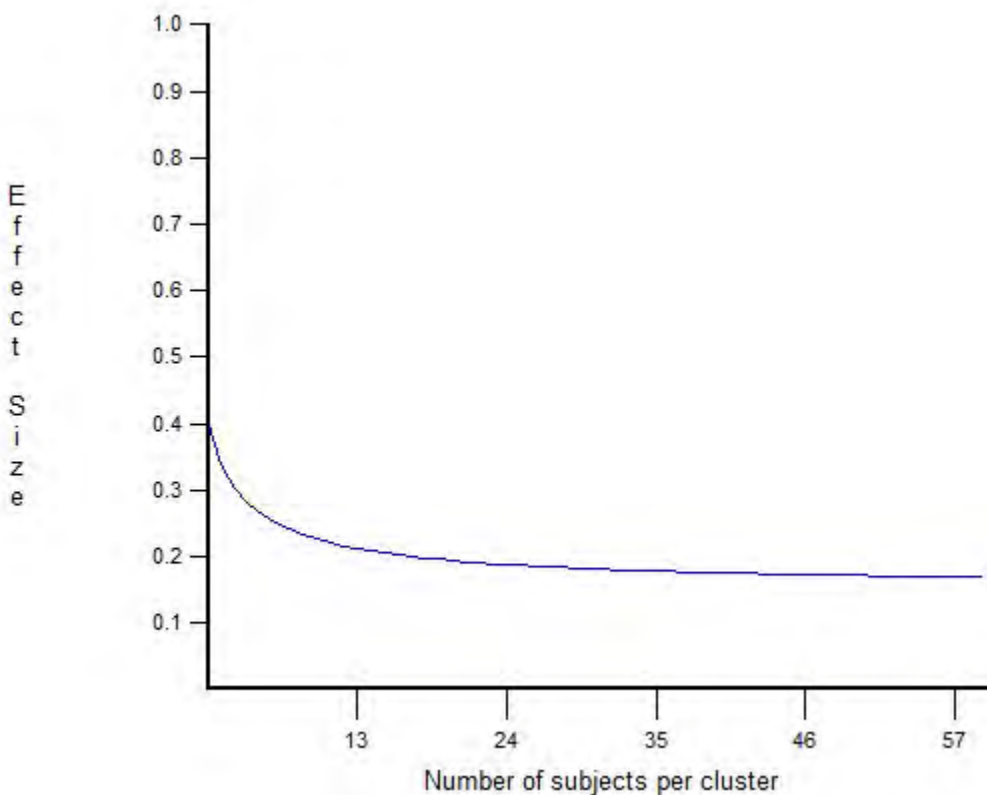
¹⁵ Note that it is possible to select a higher or lower number of students per school, which would then require adjusting the number of schools to compensate. We considered alternatives for the number of students, but based on the final calculations 30 students per school with 25 schools per treatment arm reflects the most cost effective approach to achieving a MDES of 0.23.

each treatment arm and 25 in the control group for each of the two regions. With 30 students per school, the total sample size is 4,500 for each round of the survey.

For the parent survey

As above, we set $\alpha = 0.05$ and $\beta = 0.8$. We assume that $\rho = 0.15$ instead of 0.10 to reflect the fact that we will pool the Amhara and Oromia data in our estimations of these outcomes. In addition, we assume a higher r^2 of 0.5 since we expect to have panel data for the household survey. Panel data allows for fixed effects, which substantially increases the proportion of the variation that the covariates can account for. We set the number of schools at 50 in each treatment arm based on the EGRA sample size calculations above. Figure 4 then plots the MDES against the number of students per school:

Figure 4. MDES vs. parents per school, parent survey



For the parent survey, NORC and USAID determined that the number of households per school will be 10, which corresponds to the same MDES of 0.23 as the EGRA analysis. An important consideration here was that 10 households can be surveyed at each school with minimal implications for the overall budget. As discussed below, the plan for the EGRA data collection includes one day devoted to administering the EGRA at each school. In the same day, the data collection team would also be able to survey 10 households without the need for additional vehicles to transport more enumerators to the site. Increasing the sample size further would require the data collection team to either spend additional days at each site, or assign an additional vehicle and more staff to each survey team to cover more households in a day. Either of these would have substantial implications for the budget. Moreover, we expect

that a MDES of 0.23 will be sufficient to measure our intermediate outcomes at a high level of precision, particularly since many of these are binary outcomes, which tend to be easier for the evaluation to detect. We thus opt for a sample size of 10 households per school to balance budgetary and analytical concerns.

The final sample sizes for both the EGRA and the household survey are summarized in Table 3 below:

Table 3. Final Sample Size

Region	Treatment Group	Schools	EGRA Students per school	EGRA Students	Parents per school	Parents
Amhara	Treatment Arm 1	25	30	750	10	250
	Treatment Arm 2	25	30	750	10	250
	Control	25	30	750	10	250
Oromia	Treatment Arm 1	25	30	750	10	250
	Treatment Arm 2	25	30	750	10	250
	Control	25	30	750	10	250
Total		150		4,500		1,500

D.4 SAMPLE SELECTION AND EXPERIMENTAL DESIGN

In this section, we describe the approach to sample selection and experimental design. The approach to selecting the sample for the analysis of reading skills outcomes is designed to balance the concerns of the evaluation with those of implementation.

School and Student Selection

As described above, the RCT design requires an approach to randomly allocating schools between Treatment Arm 1, Treatment Arm 2, and a control group. Within this framework, however, there is significant leeway to allow USAID, the implementing partner Save the Children (SC), and the Ethiopian Government’s Ministry of Education (MOE) to have input into the location and approach to implementation.

The impact evaluation will target two of Ethiopia’s eight regions. Following discussions with USAID and SC, it has been agreed that these will be Amhara and Oromia. Including these two regions in the study allows the evaluation to measure impacts across two separate language groups. While caution will be needed in drawing lessons from our results that apply to other language groups, including more than one language strengthens the case for external validity across language groups if the findings are similar for Amhara and Oromia.

Within these two regions, the process for selecting the sample for the EGRA and Student Survey was as follows:

1. The Woredas in which READ CO will be implemented in the first year were selected. READ CO and MOE led this process based on the concerns of project implementation.
2. Within each Woreda, READ CO and MOE then identified target schools for inclusion in READ CO. The target schools were selected using three criteria:

Criteria 1: Target schools do not currently have a library or reading corner. A proxy for whether or not that is the case is the presence/absence of another intervention (such as PACT, ANFEAE, PROPRIDE, TDA). In other words, target schools should not be supported by these other interventions.

Criteria 2: Target schools score level 1 or level 2 on the 4-point scale of the self-assessment project conducted by the MOE School Inspectorate. This self-assessment project rates school quality along dimensions such as condition of infrastructure, availability of classroom materials, etc.

Criteria 3: Only school clusters with least 3 schools meeting criteria 1 and 2 are eligible and can be considered Target School Clusters.

In accordance with implementation targets, READ CO selected a total of approximately 792 target schools in Amhara and 960 in Oromia, with an additional 120 schools in each region selected as potential replacements for schools that were later determined to be ineligible, or had been assigned to the control group. READ CO intends to systematically verify the eligibility of the initially selected target schools, but since this process would not be completed by the time of baseline data collection, it was necessary for NORC and SC to agree upon criteria for verifying school eligibility in advance so that NORC could replicate SC's approach in the sampled schools for the evaluation. Following discussions between NORC and SC, it was agreed that the process for verifying school eligibility would be to check if the schools had a functioning library with a dedicated, full-time librarian, and at least three schools in the cluster. Schools that had been initially selected as target schools would be replaced if they failed to meet either of these criteria.

3. From the list of target schools in each of the two regions, NORC randomly assigned 25 schools to Treatment 1, 25 schools to Treatment Arm 2, and 25 schools to the control group in each region. NORC carried out the random assignment so that a maximum of one school would be assigned to each group from each "school cluster," an administrative designation in Ethiopia typically comprising 3-5 schools. All schools in the same cluster will then receive the same treatment. The motivation for ensuring that all schools in the same cluster receive the same treatment reflects a concern raised by SCF that assigning schools within the same cluster to different treatment arms could lead to controversy and raise concerns over fairness among stakeholders. In addition, all schools in the same cluster minimizes the potential for "spillover" effects whereby teachers, students, or communities from neighboring schools could learn about READ CO and adopt some of the practices of the program. In order to reduce the risk of differences between the treatment and control groups, randomization used a stratification process based on the information that was available about each of the

target schools. Schools were placed into stratified “bins” on the basis of region, zone, and distance from the Woreda, with school clusters randomly selected from each bin in proportion to percentage of the population of schools in each bin.

NORC also drew samples of “replacement” schools in each bin and region to be used if any of the schools in the evaluation sample turn out to be ineligible following the verification process.

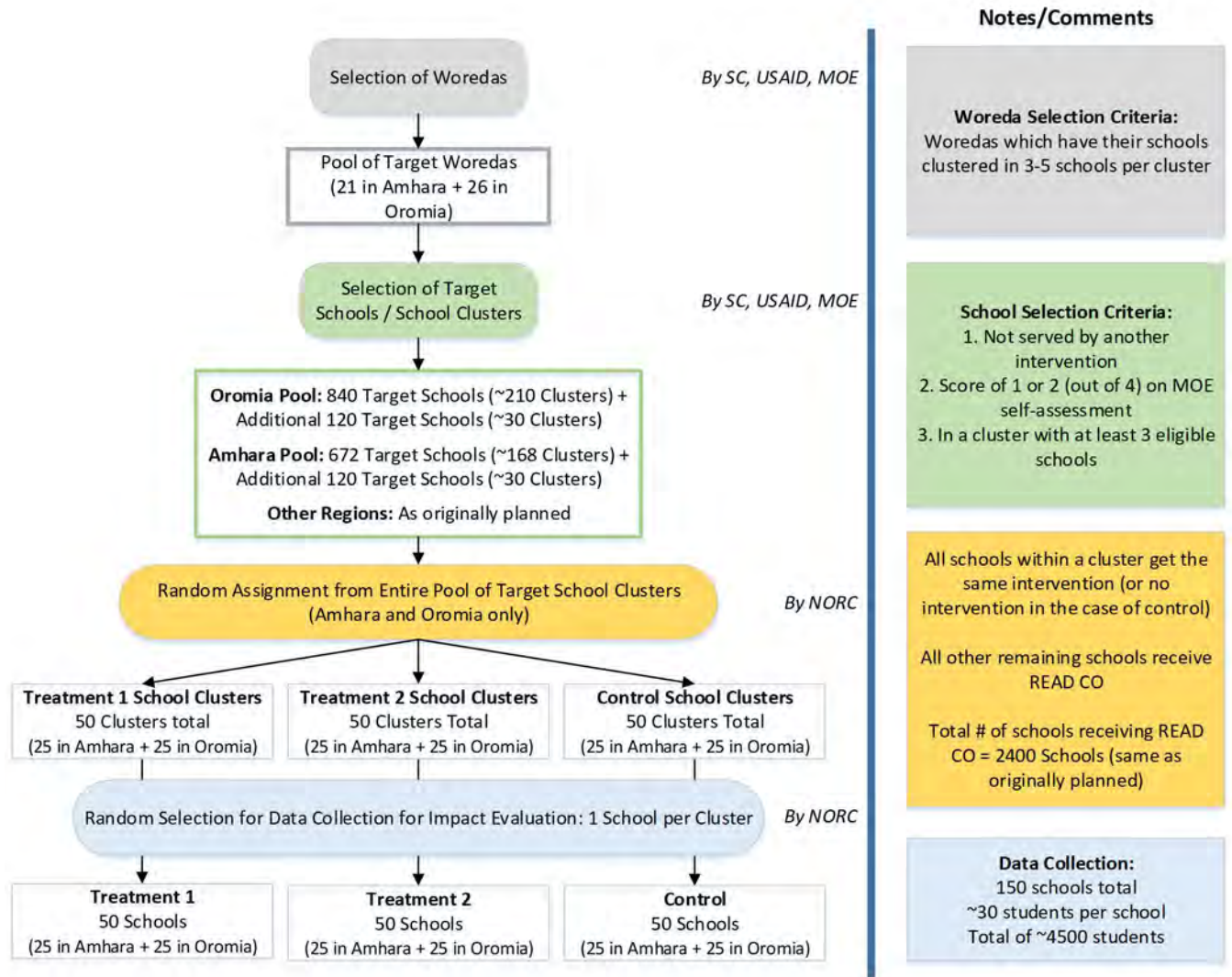
4. At baseline, the Evaluation Team will take a random sample of second grade (G2) students from each school that is to be included in the evaluation. At endline, the surveyed students will correspond to those in G2 at baseline, so that at endline two years later, we would expect these students to be in fourth grade (G4) and sample G4 students. The reason for selecting second grade in particular for the evaluation is to maintain a focus on early reading outcomes, while avoiding very low levels of reading knowledge that we would expect from first grade students. Statistical power calculations described in the previous section have established that the appropriate number of students per school is 30. Initial school visits indicated that some classes in the sample schools may have less than 30 students present at the time of the interview. As a result, the data collection team will complete the EGRA with 33 students in cases where at least that many students are present in order to compensate for any shortfall in the sample size that arises from smaller classes. In schools with less than 33 Grade 2 students present at the time of the data collection team visit, all students (i.e. a census) will be assessed.

The sampling process is summarized in Figure 5.

Parent Selection

The sample for the Parent Survey will be parents of a sub-sample of the students who are included in the EGRA and Student Survey. For each school in the Treatment Arm 1, Treatment Arm 2, and control group, a sub-sample of approximately 10 students will be randomly selected from the sample of 33 who are participating in the EGRA and student survey. Enumerators will then locate and interview one parent or caregiver of each of the 10 students, so that for each school, the EGRA, Student Survey, and Parent Survey can be completed in one day. The statistical power calculations in the previous section indicate that this subsample will be able to measure intermediate outcomes with a high degree of precision.

Figure 5. Sampling Process



Teacher Selection and School Surveys

Finally, in each school, the Grade 2 teacher will also be interviewed along with a school administrator (in most cases the principal) for the school-level survey. In schools that have more than one Grade 2 class, one of these Grade 2 classes will be randomly selected and its corresponding teacher will be interviewed (in these cases, the 33 students will be sampled from the class of the selected teacher). In schools that have multiple Grade 2 teachers that each teach specific subjects, the literacy/reading teacher will be interviewed.

E. DATA COLLECTION APPROACH

E.I INSTRUMENT DEVELOPMENT

Early Grade Reading Assessment

EGRA is comprised of multiple subtasks that focus on the five main reading skill areas outlined above. Within each of these five areas, there are multiple subtasks that can be selected for inclusion, based on local needs and the goals of the assessment system. Table 4, below, outlines the early literacy skill subtasks that are included in the EGRA tool used for this evaluation.

Table 4: Early Literacy Skill, Subtask

Early Literacy Skill	Subtask	Measurement
Alphabetic Knowledge	Fidel identification in Amharic and letter name knowledge in Afaan Oromo	Number of fidels and letter names correctly identified out of 100 in 60 seconds
Phonemic/phonological Awareness	Phonological awareness in Amharic and ending sound awareness in Afaan Oromo	Number of phonemes correctly identified out of the total of 10 words
Phonics/Alphabetic Principle	Nonword decoding	Number of nonwords correctly decoded out of 50 in 60 seconds
Fluency	Oral passage reading	Number of words in a reading passage of approximately 60 words read fluently (with accuracy) in 60 seconds
Reading Comprehension	Oral recall	Number of questions (out of five) about the reading passage (read by student) answered correctly
Listening Comprehension	Oral recall	Number of questions (out of five) about a passage read aloud (by facilitator) answered correctly

Benchmarks for the EGRA tool have been established in 10 countries, but not for Ethiopia, to our knowledge. That is, the EGRA tool provides a snapshot of early literacy skills but does not provide guidelines for interpreting which children can be considered “readers” or what level of performance should be expected on each sub-test. Ethiopia has implemented a national EGRA in 2010 and in 2013/2014 through the USAID-funded Improving the Quality of Primary Education Program (IQPEP). The American Institutes for Research (AIR) which holds the READ M&E contract conducted adaptation workshops to modify these tools in early November 2015 and is in the process of finalizing them at the time of writing of this report. Therefore, because these EGRA tools will not be finalized in time for the READ CO baseline data collection, NORC will use the IQPEP 2013/2014 tools. NORC and its partner STS have conducted a thorough review of the tools to ensure that instructions are correct, orthography is accurate and to re-randomize the order of fidels and words. Furthermore, IQPEP tools have already

been piloted and used and therefore the quality of items and subtasks is known. These tools were first developed by the Research Triangle Institute (RTI) in 2010 under EdData and the total Item Correlations and Cronbach alpha scores provided by RTI are within acceptable limits.

Additional instruments

In addition to the EGRA tools, the baseline data collection will also include: (1) a student background questionnaire which will be administered immediately following the EGRA to every student sampled, (2) a teacher questionnaire which will be administered to the Grade 2 teacher in each school, (3) a school-level questionnaire and (4) a parent questionnaire which will be administered to the parents of a subset of the students who have been sampled for EGRA.

In order to develop these associated tools, NORC worked from similar validated questionnaires that have been used in previous projects, in particular the Ethiopia IQPEP tools that were administered along with EGRA, and tools used by RTI for the USAID/Uganda School Health and Reading Program. NORC modified these tools based on the indicators targeted by the READ CO intervention, as described in Section II and III and worked in collaboration with USAID and the READ CO staff to tailor all questions such that they are carefully mapped to the objectives of the program. The instruments also contain questions related to student, household and school characteristics that can be used as covariates in the final analysis.

Following this first development phase, the tools were translated into Afaan Oromo and Amharic and pretested with a convenience sample of 15 students, 10 parents, 5 teachers and 1 school principal in each language in one Afaan Oromo-speaking school and one Amharic-speaking school located in the vicinity of Addis Ababa. Based on the feedback gathered during this pretest, NORC revised the instruments, and engaged USAID and READ CO in their finalization.

E.2 SELECTION OF FIRM AND RECRUITMENT OF ENUMERATORS

NORC has selected the Ethiopian data collection firm InEHD, Institute for Education, Health and Development, to carry out the EGRA and associated data collection activities through a competitive process.

InEHD has primary responsibility for the recruitment of enumerators. However, NORC and STS provided the firm with guidelines in terms of education and experience level. To the extent possible, we will hire enumerators who have experience administering EGRA, drawing on the pool of enumerators that have participated in the 2010 national EGRA exercise.

E.3 PROGRAMMING INTO TABLETS AND TRAINING

NORC's subcontractor, School-to-School International (STS), will program and test all tools into Tangerine. NORC/STS will carry out the training of enumerators and supervisors in the administration of EGRA in collaboration with InEHD. Given the anticipated size of the field teams, we will first conduct a Training of Trainers (TOT) for all supervisors; this TOT will last 5 days. Following the TOT, we will conduct the full enumerator training which will last 6 days, followed by a pilot test and debriefing.

NORC/STS places a strong emphasis on participatory and demonstration activities during training that 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
- Inter-rater reliability (IRR) exercises
- Use of videos, whenever possible (including for IRR exercise and role-plays)

In addition to EGRA, enumerator and supervisors will be trained in the administration of the parent questionnaire, teacher questionnaire and school questionnaire.

E.4 DATA QUALITY CONTROL

For the training, NORC and STS will create a data quality control checklist which will be used by trainers and supervisors to assess the level of skills of enumerators. A larger number of enumerators than necessary will be invited to training, such that only the best are selected for the actual field period. This selection will be based on an assessment during the classroom training using the IRR tests, as well as during the pilot test, as well as their score on the training exit quiz that NORC and STS will develop.

During the field, supervisors will remain in close contact with all their enumerators and conduct spot-checks and observations, ensuring that EGRA and other survey instruments are administered correctly. Furthermore, a field coordinator for each region will be in the field during the entire field period in order to conduct similar spot-checks and observations and ensure that all protocols are adhered to. For all the spot-checks and observations, supervisors and regional field managers will also use a data quality control checklist that NORC and STS will develop.

E.5 OPERATIONAL PLAN FOR DATA COLLECTION

As soon as the sample was selected, NORC worked with USAID and READ CO to obtain approval for contacting the schools in order to alert them of the impending data collection activities. InEHD organized a pre-data collection visit during which all schools were visited in order 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 READ CO intervention and data collection and (3) obtain student lists and contact information of parent teacher student association (PTSA) members and principals in order to facilitate contacting parents during the field period.

For the actual field period, a total of 14 teams comprised of five enumerators (of which one is an assistant supervisor) and one supervisor each will be administering all instruments over a period of 3 weeks. Each team will administer assessments and interviews at one school per day and will be assisted by community guides who will be responsible for contacting the parents and setting up appointments with them.

E.6 WORKPLAN

Below we present a detailed work plan schedule for activities for the baseline period of the READ CO impact evaluation.

Table 5: Workplan

Work Plan	
Activities	Month
<p><i>Develop and finalize impact evaluation design</i></p> <p>NORC’s Principal Investigator will travel to Ethiopia to meet with USAID and the implementer to. These meetings will also enable the PI to gain a stronger understanding of the intervention activities and roll-out plans which will inform the final selection of indicators to be measured by the evaluation. Based on information gathered and agreements reached during the trip, NORC will finalize the EDR.</p> <p>During this period, NORC will:</p> <ul style="list-style-type: none"> — Submit a draft Evaluation Design Report (EDR) to USAID — Engage in additional communications with implementer, as necessary, based on comments/feedback — Submit budget for impact evaluation — Finalize EDR 	October 2015
<p><i>Conduct sample selection, select data collection firm, develop data collection instruments and training materials, obtain IRB approval</i></p> <p>NORC will continue working closely with the IP to finalize the sample selection. In order to do this, the IP will send a complete sample frame to NORC and NORC will be responsible for assigning schools to the treatment and control groups. In parallel, NORC will develop and pretest the data collection instruments including the household questionnaire, the student questionnaire, the teacher questionnaire and the classroom observation. The EGRA instruments will be the same ones as those used for the National EGRA exercise (if possible). NORC will also select a data collection firm and finalize budget negotiations for the data collection. In collaboration with the data collection firm, NORC will develop all associated training materials and obtain IRB approval, as well as inform the Ministry of data collection activities. Under the supervision of NORC, the data collection firm will recruit enumerators and plan all logistics for training and data collection. Specific NORC activities during this period will include:</p> <ul style="list-style-type: none"> — Conduct the selection of schools for evaluation sample, randomizing schools into treatment 1, treatment 2 or control groups. — Develop, program, and pretest the following data collection instruments: teacher survey, classroom observation tool, household survey, student survey — Review national EGRA instruments, and modify if necessary — Develop training manuals, data collection protocols/plans, and quality control procedures for field work — Review all proposals from data collection firms, select one firm and finalize budget; submit paperwork to USAID contract office for subcontractor approval — Recruit enumerators, develop operational plan for data collection 	October – November 2015

Work Plan	
Activities	Month
<ul style="list-style-type: none"> — Submit IRB protocol and obtain IRB approval for data collection — Inform the Ministry of data collection plans — Send advance letters to schools to inform them of purpose of survey and approximate arrival date of enumerators. 	
<p>Conduct enumerator training and baseline data collection</p> <p>NORC's partner, School-to-School International (STS) will have primary responsibility for training enumerators in the administration of EGRA and household interviews. The training will include a 5-day Training of Trainers, followed by a 7-day training of enumerators which include a 2-day pilot test. Following the training, enumeration teams will travel to their respective zones and start data collection following NORC's quality control procedures. NORC will conduct regular reviews of data quality and will work with the data collection firm to clean and finalize the dataset. Specific activities will include:</p> <ul style="list-style-type: none"> — In collaboration with data collection firm, NORC and STS will carry out the baseline enumerator training and pilot testing of instruments — Conduct field observations during the first week of data collection — Conduct quality reviews of data — In collaboration with data collection firm, conduct data cleaning to obtain final clean datasets for EGRA, student survey, teacher survey, household survey and classroom observations. 	November – January 2016
<p>Conduct baseline analysis</p> <p>The analysis of baseline data will be designed, most importantly, to ensure the similarity between treatment and control groups. The analysis will present descriptive statistics for all indicators of interest (impact indicators, covariates) for treatment and control groups, describe any concerns with the data, and suggest adjustments to the impact evaluation design, if necessary. Prepare baseline report.</p>	Feb - March 2016
<p>Monitor Implementation of READ CO and Communication with IP</p> <p>NORC will meet regularly with the IP to ensure that program implementation is adhering to the impact evaluation design. Where implementation deviates from plan, NORC will discuss with USAID and the IP any necessary adjustments to implementation plan and/or evaluation design. NORC will engage in ongoing dialogue and discussion with USAID and the IP about evaluation implementation, and challenges. NORC will also maintain regular contact with the IP in order to obtain quarterly reports and PMP data which NORC will review on a regular basis.</p>	Throughout life of the project
Conduct midline data collection and analysis	To be decided
Conduct endline data collection and analysis	To be decided