USAID
FROM THE AMERICAN PEOPLE

PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM: RESULT 1 INTERVENTIONS

Impact Evaluation Report, Year 1

30 APRIL 2014

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

PERFORMANCE & IMPACT EVALUATION (P&IE) OF THE USAID/UGANDA SCHOOL HEALTH AND READING PROGRAM: RESULT 1 INTERVENTIONS

IMPACT EVALUATION REPORT, YEAR 1

30 April 2014

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

PRESENTED TO:

USAID/Uganda

Joseph Mwangi

PRESENTED BY:

NORC at the University of Chicago

4350 East-West Highway, 8th Floor

Bethesda, MD 20814

Telephone: (301) 634-9413

Fax: (301) 634-9301

DISCLAIMER

CONTENTS

INTRODUCTION

NORC at the University of Chicago, in collaboration with Panagora, is conducting the Impact and Performance Evaluation of USAID/Uganda's School Health and Reading Program (SHRP), implemented by RTI. RTI (referred to as the implementing Partner or IP in the rest of the document) is implementing SHRP as two separate activities: (1) Activities related to Result 1, Improved Early Grade Reading and Transition to English, and (2) activities related to Result 2, Improved HIV/AIDS Knowledge, Attitudes and Practice. This report focuses on activities related to Result 1 (early grade reading) and presents the findings of the impact evaluation of the first year of the project on Primary 1 students.[1]

The report begins with a description of SHRP's Improved Early Grade Reading and Transition to English (Result 1) activities, followed by a description of the evaluation design, an analysis of baseline balance, and findings from the impact analysis.

A. PROJECT DESCRIPTION

For its Improved Early Grade Reading and Transition to English (Result 1) intervention, SHRP focuses on the nexus of language, pedagogy, and instructional materials to significantly improve students' early grade reading and P3 literacy scores, as well as bring to scale a "Ugandan led 'reading policy'" (RTI International, 2012, p. 1). To this end, the program will build institutional capacity, support policy development and help institutionalize the training, support structures and policies necessary for sustainability. The intervention works at multiple levels. In early grade reading (Result 1), they consist of the following support the MoES:

- At the school level, the intervention is comprised of training teachers in early grade literacy instruction using students' mother-tongue in P1-P3 and with a transition to English in P4; distribution of textbooks and instructional materials and teacher guides in local languages and English; and support supervision provided to teachers.
- At the district level, training for district education officials is raising awareness and building district-level support for the new curriculum.
- At the national level, MOES systems and pedagogical and language frameworks aim to strengthen and support mother-tongue based EGR and transition to English. The intent is to support the strengthening of policies related to reading; increase advocacy for reading at multiple levels (e.g. student, teacher, school, district, and national); and generate and use data for programmatic decision making.

Together, these interventions are expected to improve the instruction and learning environment of students and eventually lead to improved literacy skills.

Teacher trainings under Result 1 are implemented through Coordinating Center Tutors (CCTs), who are school support workers in charge of monitoring education quality within their Coordinating Centers (CC). Under the district education structure, each CCT is responsible for a certain number of schools

[1] At the time of submission of this report, the Result 1 team had collected data for: Round 1 for Cluster 1 (Feb-Mar 2013), Round 2 for Cluster 1 (Oct 2013) and Round 1 for Cluster 2 (Feb 2014), and the Result 2 team had collected baseline KAP data, but no follow-up data. Therefore, this report only focuses on the impact of Result 1 using Cluster 1 Round 1 and Round 2 data. The first impact evaluation for Result 2 will only be carried out after the first follow-up data collection scheduled for 2015.

within a district (one district typically has multiple CCTs). The CCTs selected for the intervention receive training directly from SHRP and, in turn, deliver teacher training and program support in their schools, thus following a Training of Trainers (TOT) model.

Teacher trainings at the school level focus on pedagogy with an emphasis on using structured lesson plans and learner books. These lesson plans provide teachers with a practical step-by-step process for implementing the transitional bilingual approach mandated by the Ugandan EGR policy. In addition, SHRP is developing materials to support early grade reading. These materials are being adapted in order to take into account the different needs of learners at different stages of cognitive and academic development, and the linguistic characteristics of the different local languages, rather than be translated directly from one language to another. Furthermore, in order to develop these materials, SHRP works with MoES and Local Language Boards (LLBs) to standardize orthographies of the target languages. All materials follow the same general pedagogical framework to facilitate guidelines for textbook development and teacher training. Teachers receive these instructional materials in their local language of instruction and English. Teachers also receive ongoing monitoring support supervision from SHRP staff who observe classes and provide constructive feedback.

Additionally, at the district level, workshops and training for district education officials are serving as a forum for raising awareness and building district-level support for the new curriculum.

Finally, at the national level, SHRP is working with MoES and the Sector Policy Management Working Group to develop a Uganda-specific reading strategy, which will include policies in the areas of Local Language Board development, textbook development, printing as well as Special Needs. SHRP is also assisting the MoES in advocating for reading outside the classroom. Together with MoES, SHRP aims to raise awareness of local language development, reading instructions and special needs learners by using national communication campaigns through mass media and mobilizing local communities. The national level activities also include a component to strengthen the MoES' ability to monitor reading achievement for research and programmatic purposes.

Initially, the Implementing Partner (IP) had planned to develop variations of the Result 1 intervention and use experimental approaches to identify the most effective intervention to scale up. The three slightly different interventions, or "treatment arms" planned for the first year of the project included: (1) Basic Program: teacher trainings alone; (2) Basic Program + manpower support: teacher trainings with a certain number of CCT visits to schools per year; and (3) Basic Program + SMS support: teacher trainings with SMS support by CCTs. However, during the implementation phase, the IP decided that it was not feasible to implement these different treatment arms; as such, the intervention under evaluation in this report is the Basic Program, with no manpower support nor SMS support.

A.1 SHRP IN YEAR 1

In Year 1 of the intervention, SHRP worked in 4 local languages (Luganda, Runyankore/Rukiga, Ateso, and Leblano) in 11 districts[2], covering approximately 430 schools: approximately 20 CCTs and 1,659 teachers were trained[3]. By the end of Year 1, two teacher trainings had been conducted. However, there were several delays in the development and distribution of instructional materials and textbooks. These materials and textbooks were not distributed until August/September 2013, i.e. approximately 6

[2] Throughout the life of the project, SHRP will be working in a total of 12 local languages (4 languages in the first year, 4 additional languages in the second year, and 4 additional languages in the third year).

[3] USAID/Uganda School Health and Reading Program Performance Management Plan, Annex C, Version September 19 2013.

months after the start of the school year. The actual roll-out of the intervention is important to keep in mind as we interpret the results of the findings of the impact evaluation.

B. EVALUATION DESIGN

To assess the impact of the School Health and Reading Program, NORC is using a combination of an experimental (randomized controlled trial, or RCT) and quasi-experimental (matched comparisons) design. This mixed-method design allows us to estimate the combined effects of the district- and school-level interventions that comprise the School Health and Reading Program; it also allows us to isolate the effects of the school-level intervention.

An impact evaluation (IE) is conducted 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 program 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.

The ideal comparison group stems from the use of experimental methods in which eligible participants are randomly assigned to receive the intervention or not. Randomization ensures that, on average, characteristics of the treatment and control groups are statistically identical, with the only difference being their participation in the intervention. In this case, any measured difference in outcomes between the groups over time can be attributed to the program. When random assignment is not possible, quasi-experimental methods, such as statistical matching, are used to establish a comparison group.

The impact evaluation design for the SHRP Result 1 intervention uses both the random assignment of schools to treatment and control groups within SHRP intervention districts (experimental design) and the selection of matched comparison districts (quasi-experimental) in which SHRP is not operating. The experimental design allows us to isolate the effect of school-level interventions from district-level interventions, while the inclusion of non-intervention districts allows us to measure the impact of the district level interventions and the combined district+school level intervention package.

B.1 DISTRICT-LEVEL COMPARISON GROUP

In the first year of the program, the IP implemented the literacy intervention in 11 Districts located in 4 different language areas (Table 1). These 11 districts were chosen by the IP and MoES, and were not part of the evaluation design. Therefore randomization at the district level was not possible. However, we were able to select comparison districts that are similar in some key characteristics to the treatment districts. Although we had intended to pair a comparison district to each treatment district, budget and logistical restrictions expressed by IP resulted in the selection of only one district per language adding to a total of four comparison districts to compare against 11 treatment districts.

The four control districts were selected by matching non-intervention and intervention districts in a specific language area according to district characteristics such as NAPE 2011 results on P3 proficiency in oral reading, P3 proficiency in literacy in English, and P6 proficiency in literacy.4 Because we were matching only one comparison district to more than one intervention district, we computed a weighted

[4] Unfortunately, no information about HIV and AIDS knowledge, attitude and practices was available at the time of matching districts.

average of treatment districts' proficiency scores, where the weights are proportional to the number of schools participating in the program during the first year. Through this matching process, we selected four control districts - Buikwe, Ngora, Otuke, and Ibanda (Table 1).

Table 1. Treatment and Comparison Districts

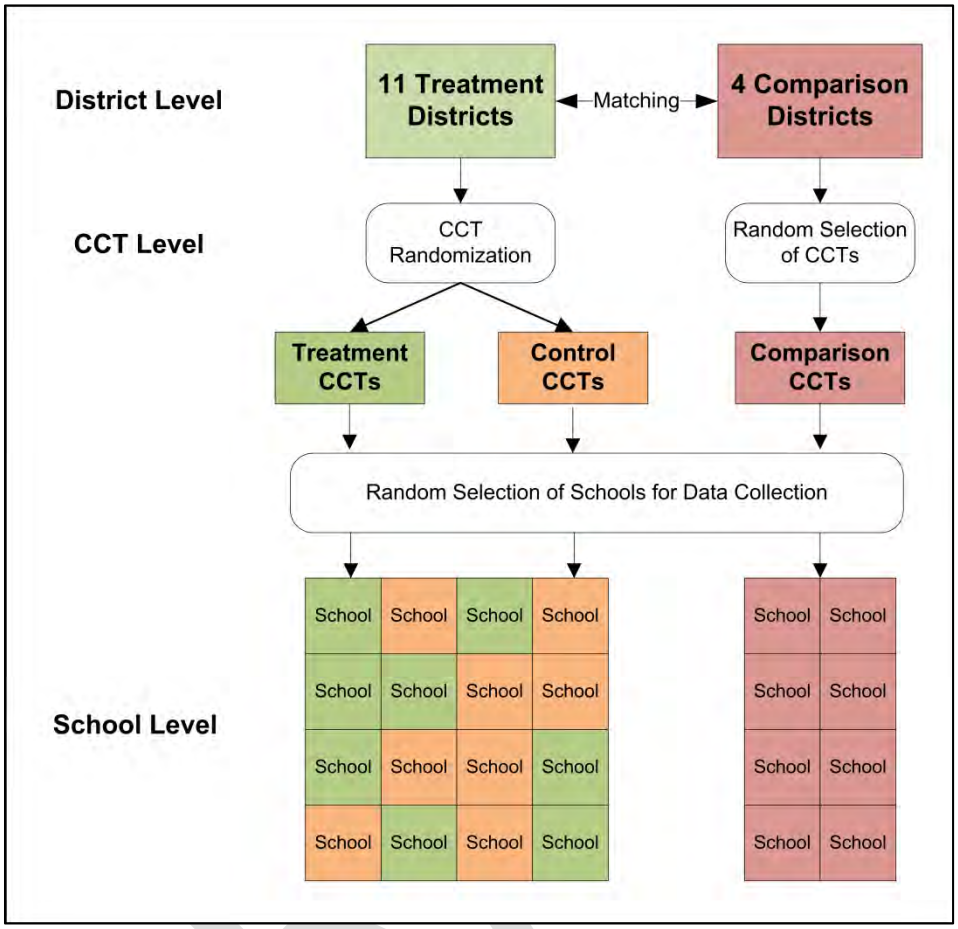
Region	Language Area	Treatment District	Comparison District
Central	Luganda	Gomba Wakiso	Buikwe
East	Ateso	Katakwi Kumi Serere	Ngora
North	Leblango	Apac Lira Kole	Otuke
South West	Runyankore/ Rukiga	Bushenyi Kiruhura Kabale	Ibanda

B.2 ASSIGNMENT OF SCHOOLS TO TREATMENT AND CONTROL GROUPS AND ESTIMATION STRATEGY

The CCTs are responsible for training teachers and providing follow-up support and assistance in implementing the Result 1 interventions. Towards this end, SHRP conducted training workshops (training of trainer workshops) for CCTs in different regions. Since each CCT is responsible for several schools, randomizing at the school level would imply that a CCT would have to treat schools under his or her jurisdiction differently if some were designated as treatment schools and others as controls. Therefore, randomization at the school level had a high risk of 'contamination' between treatment and control groups. To avoid this potential problem, we opted to randomize at the CCT level, assigning the entire cluster of schools under a CCT to either the treatment or the control group. In comparison districts we randomly selected CCTs whose school clusters will serve as out-of-intervention-district controls.

Figure 1 illustrates the evaluation design, i.e. random assignment into the treatment or control group within treatment districts and random selection of CCTs within comparison districts.

Figure 1. Assignment to Treatment, Control and Comparison Groups at the District and CCT levels



The difference in outcome indicators between treatment schools (green) and control schools (orange) will show the effect of the school-level intervention, given that both types receive the district-level intervention but only treatment schools receive the school-level program. The difference in outcomes between control schools and comparison schools (red) can identify the effect of the district-level intervention. While none of those schools benefits from the school level programs, the control schools are exposed to the district level treatment. The impact of the complete intervention package can be measured by estimating the difference in outcomes between treatment and comparison schools.

We estimate the impact of the intervention for each language within the first year of school on a number of literacy outcomes using a difference in differences model. To evaluate the impact of the intervention package on literacy score, *Y*, for students in grade P1 within a specific language group, we use the scores collected at the beginning of the school year 2013 (baseline) and from the end of the same school year (endline), and regress the individual test score *Y* of student *i* on the treatment status *T* of the school *s* in district *d,* a dummy variable *P* indicating period (i.e. 0 or 1 respectively for baseline and endline) and their interaction *T***P*:

$$Y_{isdp} = \beta_0 + \beta_1 T_{sd} + \beta_2 P + \beta_3 T_{sd} * P + \beta_4 D_d + \beta_5 X_i + \beta_6 Z_s + \varepsilon_{isd}$$

where X_i are individual characteristics of the student *i*, such as age and sex, Z_s are school characteristics such as language of instruction, D_d is a vector of dummy variables indicating the districts. As we mentioned T_{sd} is a dummy equal to 1 if school *s* in district *d* received the intervention and 0 otherwise,

P indicates the period –baseline or endline, therefore, our coefficient of interest is $\beta 3$, which shows the differential effect of the treatment at the endline.

B.3 IMPACT INDICATORS AND DATA COLLECTION INSTRUMENTS

Literacy is comprised of multiple skills, both receptive and productive. Successful readers must be able to identify letters and their corresponding sounds, segment and blend those sounds to form words and sentences, master appropriate vocabulary, and make meaning from text, among other skills. They must also be able to demonstrate their understanding and engagement with text through writing. To assess the effectiveness of the SHRP in reaching its goal to improve early grade reading and transition to English, specific key literacy skills were assessed.

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).[5] Based on this research, the Early Grade Reading Assessment, (EGRA) a brief oral reading assessment that tests these skills, is used to measure program impacts on literacy (RTI International, 2007).

EGRA is comprised of multiple sub-tests that focus on the five main reading skill areas outlined above. Within each of these five areas, there are multiple sub-tests that can be selected for inclusion, based on local needs and the goals of the assessment system.

No clear benchmarks for the EGRA tool have been established. 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. At the same time, EGRA has been used to assess early literacy skills in more than 50 countries around the world; thus, performance of students participating in SHRP can be compared with the range of performance of other children on EGRA in other low-income countries.

A notable component of SHRP is its transitional bilingual design. That is, literacy instruction begins in one of four mother tongue languages, with English introduced as a subject area nearly simultaneously (within 4-8 weeks after mother tongue instruction has begun). The language of instruction will then increasingly transition from mother tongue to English over the course of four years. Because of this transitional bilingual design, the impact evaluation necessarily requires a heteroglossic[6] approach to assessment. Therefore, early literacy skills are assessed in mother tongue and English.

This transitional bilingual design affected the selection of sub-tests included in the EGRA tool in each language. For example, because most grade 1 students cannot be expected to have prior knowledge in English language or literacy, the sub-tests that have been selected to assess literacy in English are aimed at capturing lower skill levels; in contrast, students are expected to already possess basic linguistic knowledge in their mother tongues and the EGRA sub-tests that have been selected aim to capture a distribution of literacy skills that include higher level abilities. Table 2 describes the sub-tests that are included in English and mother tongue EGRA.

[5] 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.

[6] A "heteroglossic" approach conceptualizes literacy learning in both mother tongue and English as interconnected, co-existing, and mutually reinforcing.

Table 2. Early Literacy Skills, EGRA Subtasks

Early Literacy Skill	Sub-test	Measurement	English	Mother -Tongue
Alphabetic Knowledge	Letter Sound Knowledge	Number of letter sounds correctly identified Number of letter sounds correctly identified out of 100 in 60 seconds	X	X
Phonemic Awareness	Segmenting	Number of phonemes correctly identified out of the total number found in 10 words (exact words and number of phonemes to be determined)	X	X
Phonics/Alphabetic Principles	Nonword decoding	Number of nonwords correctly decoded out of 50 in 60 seconds	X	X
Fluency	Oral passage reading	Number of words in a reading passage of approximately 68 words read fluently (with accuracy) in 60 seconds	X	X
Reading Comprehension	Oral recall	Number of questions (out of five) about a reading passage (read by student) answered correctly	X	X
Listening Comprehension	Oral recall	Number of questions (out of three) about an passage read aloud (by facilitator) answered correctly	--	X
Receptive Vocabulary	Oral identification of common objects	Number of common objects correctly identified	X	--

In addition to the EGRA, the following data collection instruments were developed and administered:

- Learner Context interview: to determine learner attendance to pre-school, socio-economic status and home literacy environment
- Teacher interview: to determine language of instruction in the school, teacher qualifications (including attendance to SHRP trainings) and the amount of support received from head teachers and CCTs
- Head Teacher interview: to obtain school enrollment information and determine participation in SHRP Training and amount of support provided to teachers
- School inventory: to determine quality of school environment (access to electricity, water, functioning toilets/latrines, availability of a school library)
- Classroom observation: to determine to extent to which teachers are applying SHRP teaching practices

B.4 SAMPLING

The standard approach to determining sample size for analytical surveys is to estimate the sample size required to achieve a specified level of power (probability), such as 90 percent, for detecting a change of a specific magnitude. This sample size depends on a number of factors including the evaluation design,

the impact estimate, design of the sample survey used to collect data, the statistical test, and the population under investigation.

An initial group of 11 districts located in 4 different language areas was selected by the IP and MoES to participate in the intervention. NORC selected a sample of 4 comparison districts. Each comparison district was individually matched on the basis of P3 and P6 NAPE literacy scores to each of the 4 language areas. Initially, within each area CCTs were randomly assigned to 4 arms of the intervention (3 treatment arms and 1 control arm). NORC calculated the number of schools needed in each language/arm cell (20 cells) and within each cell, RTI selected the requisite sample of N treatment schools and N control schools using random assignment, and N comparison schools in each comparison district. This "balanced" design is an efficient one, with high return of precision and power for survey resources expended.

For the impact evaluation of the Reading Program, we estimated the sample required to detect a double-difference measure of impact of magnitude D = 0.20 with a power of 90%. These calculations determined that 14 schools were necessary for each of the 20 cells, and that 30 P1 students would be sampled for each school, for a total of 420 students per cell. With 20 cells (3 arms and 2 controls, and 4 language subgroups), the total sample size required amounted to 8,400 students from 280 schools; i.e. 8,400 students at baseline and 8,400 students at endline. Of the 280 schools, 168 (5,040 students) would constitute the treatment group, and 112 schools (3,360) would be controls.

Based on these estimations, RTI randomly selected 168 schools for the evaluation sample, from 410 randomly selected intervention schools. Control schools within the treatment districts were selected from the schools in those districts that were not selected for the intervention.

Table 3. Baseline Target Sample Size

Language Group	Treatment Arm 1	Treatment Arm 2	Treatment Arm 3	Control	Comparison	Total
Luganda	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Ateso	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Leblango	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Runyankore /Rukiga	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	14 schools 420 students	70 schools 2100 students
Total	**168 schools 5040 students**			**56 schools 1680 students**	**56 schools 1680 students**	**280 schools 8400 students**

Table 4 shows the target sample size for the baseline data collection. However, the IP was unable to implement the three different treatment arms during the first year of the intervention. The sample size for the endline data collection was thus reduced to save resources. Given that only one treatment was implemented, the number of treatment schools was reduced to 56 schools (from 168 schools at baseline). As such, the endline sample size (and the sample size included for the impact evaluation) consisted of 56 treatment, 56 control and 56 comparison schools. The 56 treatment schools included in the endline data collection were selected randomly from the pool of 168 treatment schools from baseline.

Table 4. Endline Target Sample Size

Language Group	Treatment	Control	Comparison	Total
Luganda	14 schools 420 students	14 schools 420 students	14 schools 420 students	42 schools 1,260 students
Ateso	14 schools 420 students	14 schools 420 students	14 schools 420 students	42 schools 1,260 students
Leblango	14 schools 420 students	14 schools 420 students	14 schools 420 students	42 schools 1,260 students
Runyankore/Rukiga	14 schools 420 students	14 schools 420 students	14 schools 420 students	42 schools 1,260 students
Total	**56 schools** **1680 students**	**56 schools** **1680 students**	**56 schools** **1680 students**	**168 schools** **3,780 students**

C. DATA COLLECTION AND FINAL SAMPLE

Data collection was conducted at the beginning and at the end of the 2013 school year by the IP and its local partner Center for Social Research (CSR). The data collection was conducted using tablets in Tangerine software and consisted of EGRAs in English and 4 local languages (each student was assessed in his/her local language and in English), followed by a learner context interview. At each school, field teams also administered a teacher and head teacher interview and a school inventory. Finally classroom observations were also conducted for a subset of schools (~10%).

C.1 ASSESSOR TRAINING[7]

In preparation for data collection, RTI engaged over 100 potential assessors in a seven-day training program, and then selected the highest-performing 70 assessors for the final data collection. Assessors were selected based on inter-rater reliability tests (IRR) that were given throughout the week as well as interpersonal and leadership skills. Technical training, which was undertaken by School Health and Reading Program staff (RTI and CSR), included hands-on practice, where assessors spent one day in a school administering the tool to learners and teachers. Trainees were first trained to administer the tools on paper and then introduced to electronic data collection on Nexus tablets so that they could be prepared for both circumstances. Four Data Quality Assurance (DQA) Officers, who also acted as assessor trainers, were given an extra day of training on the classroom observation instrument. This included going to a school, observing a class, and comparing findings among one another.

C.2 DATA COLLECTION

Data were collected from February 15 through March 20, 2013 for the baseline round and from October 14 through November 1 2013 for the endline round. Seventy assessors in teams of four (each of these teams included one supervisor) were deployed to the four language areas. Each of the four language area teams was supported by a DQA Officer. This DQA Officer was responsible for

[7] From RTI EGRA Baseline Report

overseeing all aspects of data collection deployment, observing assessors and providing feedback and support, ensuring data were uploaded from the electronic tablets every evening, and observing the reading classes. Besides the team supervisors and the DQA Officers, data collection was overseen by MoES staff (from the Directorate of Education Standards and from Guidance and Counseling), School Health and Reading Program Staff, and staff from the external evaluation team from NORC.

C.3 SCHOOLS AND LEARNERS SAMPLED

The final sample included in the analysis consists of a panel of 163 schools. In those schools, a total of 4260 students were tested at baseline and 4278 students completed EGRA in the endline round. Table 5 presents the number of students included in the analysis; the number of schools is indicated in parenthesis.

Table 5. Final Sample Used in the Analysis

	Treatment		Control		Comparison	
	Baseline	**Endline**	**Baseline**	**Endline**	**Baseline**	**Endline**
Luganda	350	345	236	209	306	312
	(15)		(12)		(14)	
Lango	452	426	388	390	386	347
	(15)		(13)		(13)	
Ateso	392	391	418	421	424	420
	(13)		(14)		(14)	
Runyankore/Rukiga	282	315	337	379	289	323
	(13)		(14)		(13)	
Total	**1476**	**1477**	**1379**	**1399**	**1405**	**1402**
	(56)		**(53)**		**(54)**	

D. BALANCE AT BASELINE

In order to explore comparability of pre-treatment characteristics of the treatment, control and comparison groups, we conducted mean equality tests to test for balance. We compared pre-treatment scores (1) between treatment and control schools within treatment districts and (2) between treatment schools from treatment districts and comparison schools from comparison districts. Because comparisons between languages cannot be made given that no benchmarking and psychometric analysis has been performed to assess comparability between languages, mean equality tests of local language literacy scores are conducted within language subgroup. On the other hand, tests for English literacy scores can be conducted on the entire sample.

D.1 Treatment vs. Control Schools within Treatment Districts

Overall, our analysis indicates that pre-treatment literacy scores between treatment and control groups are balanced. Table 6 shows that the treatment and control groups are very similar in English literacy scores, demographics and other characteristics at baseline. None of the variables included in the means equality tests are statistically significantly different between treatment and control groups.

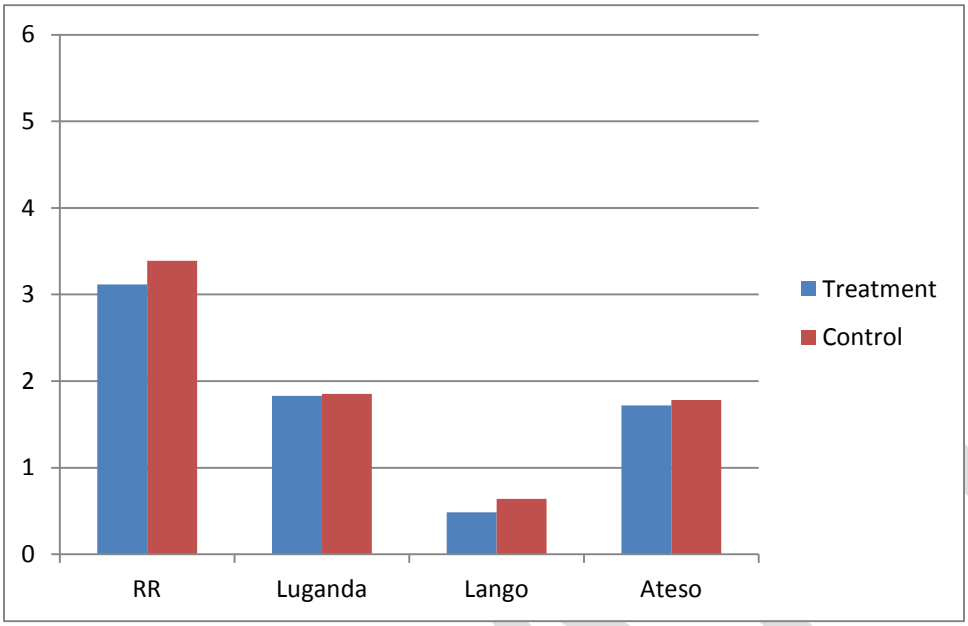
Table 6. English literacy scores, demographics and other characteristics at baseline for treatment and control groups. All language groups.

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C – T)
Age of student	7.317 (0.048)	7.325 (0.053)	0.008
% of cases with missing age information	0.125 (0.009)	0.156 (0.010)	0.031
Gender of student (female = 1)	0.499 (0.013)	0.494 (0.013)	-0.005
Number of assets (max = 8)	2.508 (0.033)	2.575 (0.034)	0.067
Lives with both parents (yes = 1)	0.628 (0.013)	0.653 (0.013)	0.025
Does not live with mother (yes = 1)	0.182 (0.010)	0.182 (0.010)	0.000
Reads at home (yes = 1)	0.388 (0.013)	0.414 (0.013)	0.026
Attended preschool (yes = 1)	0.480 (0.013)	0.415 (0.013)	-0.065
Student absent any day in the week prior to assessment (yes = 1)	0.508 (0.013)	0.507 (0.014)	-0.001
Teacher absent any day in the week prior to assessment (yes = 1)	0.414 (0.013)	0.430 (0.014)	0.016
Gender of teacher (female = 1)	0.745 (0.012)	0.615 (0.013)	-0.130
Letter sound score (max = 100)	1.331 (0.087)	1.513 (0.091)	0.182
Word segmenting score (max = 10)	0.175 (0.023)	0.188 (0.024)	0.013
Nonword decoding score (max = 50)	0.108 (0.025)	0.085 (0.026)	-0.023
Oral reading fluency (max = 68)	0.300 (0.044)	0.142 (0.032)	-0.157
Reading comprehension score (max = 5)	0.002 (0.001)	0.002 (0.002)	0.000
Receptive vocabulary score (max=20)	4.437 (0.097)	4.247 (0.092)	-0.190

Literacy scores in English and local language as well as demographics and other characteristics are similarly balanced between treatment and control groups within each language subgroup. Figure 2 provides an illustrative example; at baseline, letter sound knowledge scores in the local language are not statistically different between treatment and control groups for any of the four language subgroups. The only significant difference we see is in the Luganda subgroup, where Luganda treatment students scored significantly better than control students in Luganda reading comprehension, although the difference is

small and most students scored zero comprehension. None of the other literacy scores showed statistically significant differences between treatment and control groups. Annex A shows the results of all variables for all language subgroups.

Figure 2. Local language letter sound knowledge score, by language subgroup



D.2 Treatment Schools from Treatment Districts vs. Control Schools from Comparison Districts

By contrast, the baseline scores between treatment and comparison groups (i.e. treatment schools within treatment districts and control schools within comparison districts) were not as well balanced as it was expected given that comparison districts were selected through matching. This is probably the result of reducing the number of comparison districts to only 4, rather than the 11 planned in the original design. Despite some imbalance, most of the differences between these two groups were not statistically significant.

While there were no significant differences between treatment and control groups in basic demographic characteristics and English EGRA scores pooled from the entire sample, we note significant differences between treatment and comparison groups for three variables: (1) a higher proportion of treatment students had attended preschool than comparison students, (2) treatment students scored significantly higher on the alphabetic knowledge subtask (letter sound score) and (3) treatment students scored significantly better on the English passage reading subtask (oral reading fluency) than comparison students (Table7).

Table 7. English literacy scores, demographics and other characteristics at baseline for treatment and comparison groups. All language groups

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C – T)
Age of student	7.317 (0.048)	7.325 (0.048)	0.008
% of cases with missing age information	0.125 (0.009)	0.148 (0.009)	0.023
Gender of student (female = 1)	0.499 (0.013)	0.491 (0.013)	-0.008
Number of assets (max = 8)	2.508 (0.033)	2.373 (0.035)	-0.136
Lives with both parents (yes = 1)	0.628 (0.013)	0.661 (0.013)	0.033
Does not live with mother (yes = 1)	0.182 (0.010)	0.161 (0.010)	-0.021
Reads at home (yes = 1)	0.388 (0.013)	0.415 (0.013)	0.027
Attended preschool (yes = 1)	0.480 (0.013)	0.388 (0.013)	-0.091**
Student absent any day in the week prior to assessment (yes = 1)	0.508 (0.013)	0.532 (0.013)	0.025
Teacher absent any day in the week prior to assessment (yes = 1)	0.414 (0.013)	0.467 (0.014)	0.053
Gender of teacher (female = 1)	0.745 (0.012)	0.674 (0.013)	-0.071
Letter sound score (max = 100)	1.331 (0.087)	0.909 (0.067)	-0.422*
Word segmenting score (max = 10)	0.175 (0.023)	0.286 (0.036)	0.111
Nonword decoding score (max = 50)	0.108 (0.025)	0.057 (0.023)	-0.051
Oral reading fluency (max = 68)	0.300 (0.044)	0.129 (0.035)	-0.171*
Reading comprehension score (max = 5)	0.002 (0.001)	0.001 (0.001)	-0.001
Receptive vocabulary score (max = 20)	4.437 (0.097)	3.990 (0.081)	-0.447

Note: * Significant at 10% level, ** Significant at 5% level

Within each language subgroup as well, the scores were not as well balanced between treatment and comparison groups as they were between treatment and control groups. While the only significant difference between treatment and control groups was found for Luganda reading comprehension (for which Luganda treatment students scored significantly better than control students), we find significant differences between treatment and comparison groups in the following variables and for the following language subgroups: (1) R/R treatment students scored statistically significantly better than comparison students in R/R nonword decoding, R/R oral reading fluency, R/R reading comprehension, English nonword decoding, and English oral reading fluency; (2) Luganda treatment students scored statistically significantly better than comparison students in English receptive vocabulary; and (3) Lango treatment

students scored statistically significantly better than comparison students in English oral reading fluency. All other literacy scores were balanced between treatment and comparison groups. See Annex B for the complete set of results for all language subgroups.

Overall, the sample at baseline was well-balanced between treatment and control schools and between treatment and comparison schools, indicating that the randomization and matching approach generated equivalent groups in terms of baseline characteristics. The small differences found between treatment and comparison schools can be taken into account in the data analysis. We are therefore confident that differences detected between treatment and control groups or between treatment and comparison groups over time would stem from the impact of the SHRP intervention as opposed to the effect of other baseline characteristics of the various sampled groups.

E. IMPACT ANALYSIS RESULTS

This section summarizes the results of the impact analysis after one year of intervention[8]. Our analysis is conducted for each language subgroup for each of the outcome scores on subtasks conducted in the local language. On the other hand, for outcome scores on English subtasks, our analysis is applied to the entire sample (pooled sample from all four language subgroups). It is important to keep in mind that this evaluation does not allow for comparison of progress and impact between languages as the EGRA tools were developed independently for each local language and no psychometric analysis has been conducted to determine whether EGRA scores in one language can be compared to EGRA scores in another language. Therefore, while we can determine whether SHRP has had an impact for a particular language subgroup vs. another language subgroup, we cannot comment on the relative magnitude of these impacts between language subgroups, except for the English scores.

E.1 IMPACT OF SCHOOL-LEVEL INTERVENTION: TREATMENT VS. CONTROL SCHOOLS

First, we study the effect of the school-level intervention with difference-in-difference models using the treatment and control school sample (see Evaluation Design section, **Figure 1**). We present findings for scores in local language within each language subgroup and in English for the pooled sample. In the tables below, we report the average treatment effect on each outcome of interest and for each language subgroup using different model specifications to check for robustness. All regression models include individual controls (sex and age of the student, a dummy for whether age is missing, household asset index, dummies for living with both parents, having someone read to the student at home, and language spoken at home)[9] and district dummies. For the regressions on English scores using the pooled sample, we include dummies for local language subgroups. We also test results with and without school fixed effects. All standard errors are robust standard errors and allow for correlation in the unobservables between learners in the same class. All regressions are Ordinary Least Squares (OLS) regressions.

Impact on Letter Sound Knowledge

Each cell in Table 8 shows the average effect of the treatment (SHRP intervention) on Letter Sound Score for each language subgroup using 5 different model specifications. Since schools were randomly assigned to the treatment and control groups, and mean equality tests show that both groups are well-balanced at baseline, we start with the simplest model (model (1)) which does not include any controls. To reach more precision and to check for robustness, we add controls in models (2) through (5).

- Model (1) includes no controls.
- Model (2) is similar to model (1) but includes districts fixed effects.
- Model (3) is similar to model (2) but adds individual controls.
- Model (4) is similar to model (3) and also includes school fixed effects.
- Model (5) is similar to model (4) but only includes the subsample of students 6 years and older. We do this because a somewhat considerable proportion of students interviewed at baseline

[8] At the end of the first year, SHRP had implemented 2 teacher trainings. Teacher school support and distribution of materials did not start until August/September 2013.

[9] With the exception of models (1) and (2) in Table 8.

and endline reported being 4 and 5 years of age even though the official age for starting P1 in Uganda is 6 years old. At baseline, more than 9 percent of the students in our sample reported being 4 and 5 years of age and at follow-up, 4.5 percent reported being 5 years old and none reported being 4 years old. Although it may be possible that some students officially enroll in P1 before the age of 6, reports from the data collection teams suggest that these younger children present in P1 classes are in fact not formally enrolled in primary school. This issue was uncovered at the beginning of the baseline data collection when teams were already in the field;therefore, the issue of whether to include them in the data collection was not addressed during training (since it was assumed that all students present would be official P1 students). It is, therefore, possible that field teams sampled children who were too young to attend P1. In the follow-up round, in addition to the natural aging process of the cohort, this issue was handled in a more systematic way (i.e. excluding those children that did not formally belong to the P1 classes), thus explaining the drop in the proportion of students reporting being 4 and 5 years old. For these reasons, for this analysis, we decided to include a model that only analyzes the subsample of students who reported being 6 years and older, i.e. students officially enrolled in P1.

With the simplest model (model (1)), we find that, in its first year, the intervention did not have an effect on Letter Sound scores in any local language, nor in English, with the exception of Luganda. We find that results are very similar across all other models; the intervention, in its first year, did not have an impact on Letter Sound scores except for the Luganda subgroup for which treatment students scored on average 2.5 letters sounds higher than control students. Both groups started with baseline knowledge of around 1.85 letter sounds, while the Luganda control group scored only 3.5 correct letter sounds at the end of the academic year, the Luganda treatment group averaged 6 correct sounds. This difference is significant as a percentage however the improvement of both groups during the academic year is disappointing in absolute terms.

Table 8. SHRP School Level Effect - Letter Sound Score

Language	(1)	(2)	(3)	(4)	(5) 6 years and older
Runyankore/Rukiga	0.219	-0.145	-0.048	0.246	0.069
	(1.712)	(1.685)	(1.670)	(1.592)	(1.755)
Luganda	2.454**	2.539**	2.515**	2.455**	1.906*
	(1.070)	(1.080)	(1.096)	(1.038)	(0.956)
Lango	0.970	0.963	1.086	1.073	1.065
	(0.838)	(0.834)	(0.826)	(0.820)	(0.889)
Ateso	0.076	0.079	0.260	0.278	0.345
	(0.897)	(0.898)	(0.935)	(0.931)	(1.021)
English	1.098	0.874	0.930	1.256	1.385
	(1.239)	(1.223)	(1.186)	(1.110)	(1.200)
Districts Fixed Effects	no	yes	yes	yes	yes
Individual Controls	no	no	yes	yes	yes
School Fixed Effects	no	no	no	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include age and sex of

learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English scores include controls for local language.

Regressions in Table 9 include district fixed effects, individual controls and a constant term. Table 9 shows the average treatment effect on the percentage of students who could not identify any letters in the Letter Sound Knowledge subtask (percentage of zero scores). Columns (1) and (2) show the results of regressions without and with school fixed effects respectively. The intervention had an impact on the percentage of students who scored zero in Luganda letter sound knowledge, although the statistical significance is weak; the intervention led to approximately a 12 percentage point decrease in the proportion of students who could not identify any letters. We also find a weakly significant impact on the percentage of students who scored zero on the English letter sound knowledge subtask when we include school fixed effects; the intervention contributed to a 6.6 percentage point drop in the students who knew no English letter sounds. This result does not hold when school fixed effects are not included. There are no significant impacts of the treatment in any of the other languages.

Table 9. SHRP School Level Effect - Percentage of students who scored zero on letter sound subtask

Language	(1)	(2) With school fixed effects
R/R	-0.122	-0.093
	(0.082)	(0.083)
Luganda	-0.121*	-0.129*
	(0.073)	(0.073)
Lango	-0.092	-0.091
	(0.066)	(0.066)
Ateso	-0.055	-0.055
	(0.074)	(0.075)
English	-0.061	-0.066*
	(0.037)	(0.037)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Word Segmenting Score

Table 10 shows the average treatment effect on Word Segmenting scores (regressions include individual controls, district fixed effects and a constant term). Students received one point for every word that they could segment correctly for a maximum of 10 points. We find slightly significant impacts for the Luganda subgroup for which treatment students scored approximately 0.9 words better than the control students. No impact is found for any other local languages, nor English word segmenting.

Table 10. SHRP School Level Effect - Word Segmenting Score

Language	(1)	(2) With school fixed effects
R/R	0.443	0.229
	(0.426)	(0.430)
Luganda	0.863*	0.881*
	(0.468)	(0.478)
Lango	0.006	0.005
	(0.100)	(0.102)
Ateso	0.221	0.029
	(0.569)	(0.492)
English	0.018	0.021
	(0.174)	(0.174)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

In terms of the impact of the intervention on higher level literacy skills, namely non-word decoding, oral reading fluency, reading comprehension and English receptive vocabulary, we find no impact whatsoever for any of the languages. In **Error! Not a valid bookmark self-reference.** we show our findings using the school fixed effect models[10].

Table 11. SHRP School Level Effect - Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

	(1)	(2)	(3)	(4)
Dependent Variable	Non-Word Decoding	Oral Reading Fluency	Reading Comprehension	English Receptive Vocabulary
R/R	0.995	1.727	0.140	
	(1.014)	(1.263)	(0.128)	
Luganda	-0.028	-0.277	-0.035	
	(0.600)	(0.787)	(0.070)	
Lango	0.076	0.008	0.008	
	(0.064)	(0.101)	(0.008)	
Ateso	0.065	-0.030	0.010	
	(0.077)	(0.087)	(0.010)	
English	0.139	0.184	0.004	0.153
	(0.190)	(0.383)	(0.008)	(0.213)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district and school fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include controls for local language.

Impact on Boys vs. Girls

Finally, in general we do not find differences between the performance of boys and girls. However, when focusing on letter sound identification, girls show an advantage over boys in Luganda and English. The advantage is quite substantial in Luganda where girls are able to identify more than one letter sound than boys on average. In English, the lead is more modest – a quarter of a letter - but statistically significant. We do not observe differences in any of the other languages and we do not observe differences in the impact of SHRP by gender.

Discussion of impact of school-level intervention

In summary, these analyses suggest that the intervention, after one year of implementation, has only had an impact on students from one language subgroup (Luganda students) and only on lower level literacy

[10] Results are very similar when we do not include school fixed effects.

skills - letter sound knowledge and weak impacts on phonemic awareness as measured through the word segmenting subtask - but not on the higher level skills - reading words, oral reading fluency and reading comprehension. Given that the intervention has only been implemented for one year - with only two teacher trainings organized at the time of the follow-up data collection and with a delay of distribution of textbooks which only happened well into the school year in August/September 2013 - it may not be surprising that the intervention has had little impact.

Impact of school-level intervention:

- Positive impact on letter sound knowledge and reduction in proportion of students who could not read any letter sounds in Luganda subgroup
- Positive impact on word segmenting in Luganda
- Reduction in proportion of students who could not read any letter sounds in English (although only significant with school fixed effects model)
- No impact on higher level literacy skills for Luganda subgroup
- No impact for any EGRA subtask in any other local language (Runyankore/Rukiga, Lango, Ateso)

On the other hand, where we do see impacts, the impacts are on lower level literacy skills. Given that skills related to alphabetics (letter sound knowledge and phonemic awareness) are foundational for decoding words[11], it is expected that these are the skills that would show improvement first.

However, despite the positive impacts for the Luganda subgroup, it is also important to note that in absolute terms, results are concerning. On average, after one school year, despite the presence of a literacy project, students still only know a few letter sounds, 6 letter sounds for the Luganda treatment group, for example, while the proportion of students who cannot identify a single letter sound at the end of the school year is 40 percent for the same group (down from 70% at the beginning of the year).

As implementation continues to be rolled out, and students move up in grades, it is possible that impacts on higher level literacy could manifest themselves. Additionally, with time, we may see a gradual transfer of these skills from the local language to English, if the hypothesis that mother-tongue based early grade literacy with transition to English is the most effective bilingual literacy acquisition model holds. However, such higher level impacts must be preceded by substantial improvements in students' foundational reading fluency and comprehension skills.

Finally, as for why impacts were detected in Luganda but not other language subgroups, this will be something that the Performance Evaluation can shed light on. It is possible that implementation was rolled out differently in the various districts associated with each language subgroup. For instance, the effectiveness of the instructional materials developed by SHRP may vary from language to language. It could also be that the quality and effectiveness of teacher trainings differed by district or that teachers from the Luganda speaking districts were more receptive and eager to apply the techniques acquired through the trainings than teachers from other areas or that CCTs in these areas developed better support and monitoring systems.

It could also be that teachers in the Luganda trainings attended more trainings than teachers from other areas. We investigate this by looking at the number of SHRP trainings that teachers reported attending.

[11] Longitudinal studies have shown that phonemic awareness is highly predictive of decoding and that phonemic awareness is necessary, although not sufficient, for learning how to read. International Reading Association, 1998.

During the endline data collection, one P1 teacher from each sampled school was interviewed about their attendance to SHRP trainings. Recall that by the end of 2013, SHRP had organized two teacher trainings. The sample is small but suggests that the proportion of teachers that attended both teacher trainings - the May and September trainings - is higher in the Luganda speaking districts than other districts, although the differences are not statistically significant (83% of sampled Luganda district teachers had attended both trainings compared to 66-69% for other language subgroups). This may be a factor in explaining the fact that SHRP had significant positive impacts in the Luganda speaking districts but not in the other districts, although we remain cautious given the small sample sizes.

Table 12. Teacher attendance to SHRP trainings in the treatment schools

	No training	One training only	Both trainings (May and September)	Total
	%	%	%	%
Luganda	0	16.67	83.33	100
Lango	0	33.33	66.67	100
Ateso	7.69	23.08	69.23	100
Runyankore/Rukiga	0	33.33	66.67	100
Total	2.04	26.53	71.43	100

E.2 IMPACT OF DISTRICT AND SCHOOL-LEVEL INTERVENTION: TREATMENT VS. COMPARISON SCHOOLS

In addition to assessing the effect of the school-level intervention, we also assess the effect of the district and school-level interventions combined by comparing treatment schools to comparison schools (see Evaluation Design section, **Figure 1**). The regressions in the tables below use the same specifications as the ones presented to assess the effect of the school-level intervention in section E.1:

- In Table 13, we report the impact on Letter Sound scores. We start with the simplest model with no controls and add progressively more controls.
- In Table 14, we analyze the impact of the school+district-level intervention on the percentage of students who scored zero on the letter sound knowledge subtask, with and without school fixed effects.
- In**Table 15**, we report the average treatment effect on word segmenting scores, with and without school fixed effects.
- In **Table 16**, we report the average treatment effect for other subtasks; non-word decoding, oral reading fluency, reading comprehension and English receptive vocabulary. These models all include school fixed effects.

In addition, we have also estimated all these models using propensity-score matching (PSM) at the school level to further ensure comparability of treatment and comparison cases. These models yield similar results to the ones without use of school level PSM.

Impact on Letter Sound Knowledge and Word Segmenting

Similar to the impact of the school-level intervention discussed in Section E.1, we find impacts of the school/district-level intervention on the Luganda subgroup in Letter Sound scores and Word Segmenting. On average, treatment students in the Luganda subgroup could read 2.5 letters more than comparison students and could segment 0.8 words more than comparison students at the end of the first year of intervention (although the impact is not significant for the word segmenting subtask when we include school fixed effects). The size of the impact for these two subtasks is quite similar to what we found for the school-level intervention.

In addition to these impacts on Luganda scores, we also find significant impacts on Runyankore/Rukiga Letter Sound scores although these impacts are negative (for the analysis on impact of the school-level intervention, Runyankore/Rukiga Letter Sound coefficients were also negative in some models but not significant). On average Runyankore/Rukiga treatment students could identify 2.2 letter sounds fewer than comparison students. This finding is robust across all models, with the exception of the model that only includes 6 year-old and older students. We also find a significant impact on the percentage of students who could read no letter sounds at all in the Runyankore/Rukiga subgroup[12]; on average the intervention contributed to a 16 to 19 percentage point increase in the proportion of students who could not identify any letters.

We find no significant impacts on Letter Sound Knowledge or Word Segmenting for any other language subgroup.

[12] This is true in the model with students 6 years and older as well.

Table 13. SHRP School and District Level Effect - Letter Sound Score

Language	(1)	(2)	(3)	(4)	(5) 6 years or older
Runyankore/Rukiga	-2.194*	-2.331**	-2.128*	-2.204*	-2.170
	(1.091)	(1.109)	(1.082)	(1.168)	(1.365)
Luganda	2.569**	2.559**	2.503**	2.465**	2.279**
	(1.102)	(1.092)	(1.103)	(1.094)	(0.999)
Lango	0.173	0.164	0.250	0.111	0.311
	(0.995)	(0.993)	(0.977)	(0.944)	(0.941)
Ateso	1.029	1.032	0.883	0.927	0.693
	(0.942)	(0.943)	(0.993)	(0.987)	(1.126)
English	0.334	0.326	0.357	0.338	0.218
	(0.405)	(0.403)	(0.405)	(0.406)	(0.411)
Districts Fixed Effects	no	yes	yes	yes	yes
Individual Controls	no	no	yes	yes	yes
School Fixed Effects	no	no	no	yes	yes

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English scores include controls for local language.

Table 14. SHRP School and District Level Effect – Percentage of students who scored zero on Letter Sound subtask

Language	(1)	(2) With school fixed effects
R/R	0.162**	0.195**
	(0.073)	(0.075)
Luganda	-0.071	-0.070
	(0.069)	(0.068)
Lango	-0.050	-0.035
	(0.066)	(0.065)
Ateso	-0.047	-0.049
	(0.064)	(0.064)
English	0.004	0.010
	(0.037)	(0.037)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the

student and language spoken at home. Regressions for English language include control for local language.

Table 15. SHRP School and District Level Effect -- Word Segmenting Score

Language	(1)	(2) With school fixed effects
R/R	0.310	0.109
	(0.407)	(0.431)
Luganda	0.782*	0.727
	(0.443)	(0.431)
Lango	0.017	0.016
	(0.077)	(0.080)
Ateso	0.109	-0.209
	(0.689)	(0.579)
English	0.020	0.025
	(0.168)	(0.169)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

For the Luganda subgroup, the school-level intervention only impacted lower level literacy skills – letter sound knowledge and word segmenting. However, our analysis shows that in addition to impacting letter sound knowledge and word segmenting, the school+district-level intervention also had a weakly significant impact on all other literacy skills for that subgroup, namely non-word decoding, oral reading fluency and reading comprehension. On average Luganda treatment students could read 0.9 additional words, read 1 additional word per minute and answer 0.1 additional comprehension questions correctly than comparison students.

We find no significant impacts for these skills in any other language subgroup.

Table 16. SHRP School and District Level Effect - Non-Word Decoding, Oral Reading Fluency, Reading Comprehension and English Receptive Vocabulary

	(1)	(2)	(3)	(4)
Dependent Variable	Non-Word Decoding	Oral Reading Fluency	Reading Comprehension	English Receptive Vocabulary
R/R	-0.026	0.994	0.124	
	(1.118)	(1.306)	(0.130)	
Luganda	0.865*	1.006*	0.093*	
	(0.448)	(0.506)	(0.051)	
Lango	0.048	0.109	0.012	
	(0.077)	(0.092)	(0.012)	
Ateso	-0.047	-0.125	-0.008	
	(0.081)	(0.106)	(0.008)	
English	0.234	0.479	0.009	-0.192
	(0.183)	(0.343)	(0.008)	(0.234)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include district and school fixed effects, individual controls and a constant term. Individual controls include age and sex of learner, dummy for age missing, score on household assets index, dummies for student living with both parents, someone at home reading to the student and language spoken at home. Regressions for English language include control for local language.

Impact on Boys vs Girls

In terms of differential impacts on boys vs girls, the findings are similar to those of the impact of the school-level intervention. We do not find that the program has different impacts depending on the gender of the learner. In addition, in general, we do not find differences between the performance of boys and girls. However, girls show an advantage over boys on letter sound identification in Luganda and English.

Discussion of impact of school+district-level intervention

In summary, these results are similar to what was found in terms of the impact of the school-level intervention but differ with respect to the impact on higher level Luganda literacy skills and the impact on Runyankore/Rukiga letter sound knowledge. Below, we discuss possible explanations for these findings:

Impact of school+district-level intervention:

- Positive impact on Luganda letter sound knowledge, word segmenting, nonword decoding, oral reading fluency and reading comprehension
- Negative impact on letter sound knowledge and proportion of learners that know letter sounds in Runyankore/Rukiga subgroup
- No impact on other Runyankore/Rukiga scores
- No impact for any other local language EGRA scores (Lango, Ateso) or English

- *Why are there impacts on both higher level and lower level Luganda literacy skills for school+district-level interventions, while there were impacts on lower literacy skills for school level interventions?*

  At the district level, SHRP invited district education officials to attend trainings. It is possible that these district education officials provided an additional level of support to the teachers. SHRP also worked with Language Boards to develop orthographies and instructional materials in local languages. Therefore the district-level activities combined with the school-level activities could have had a stronger impact than the school-level activities alone, for the Luganda subgroup.

  However, it is also important to note that higher level literacy skills are very rare among the learners. The vast majority of learners, in SHRP and non-SHRP schools, is not able to complete any part of the subtasks such as decoding non-words, or reading any word correctly in a paragraph.

- *Why are there impacts for Luganda but not for the other language subgroups?*

  As mentioned in section E.1, it could be that the intervention was implemented differently in the Luganda areas than the other areas and that teacher trainings and/or instructional materials in Luganda are more effective. At the district level, there may also be differences in how district education officials support teachers after attending SHRP trainings between the Luganda areas and other areas. Our performance evaluation will investigate differences in implementation between districts.

- *Why are impacts on Runyankore/Rukiga letter sound knowledge negative?*

  In general, we would not expect the SHRP program to have a negative impact but it cannot be ruled out. Our analysis controls for learners and households characteristics, and we used districts and schools fixed effects which can absorb differences between the groups even when we cannot actually observe them. In addition, the results are robust to different specifications including the use of propensity score matching at the school level.

  However the analysis cannot take into account any district-level characteristic that we do not observe and that changes between baseline and endline, such as any additional intervention or policy that is implemented in comparison but not treatment districts or vice-versa. For instance, if the Runyankore/Rukiga comparison district, Ibanda, benefitted from a new district-level intervention during the course of SHRP's first year of implementation, while the

Runyankore/Rukiga treatment districts did not benefit from that intervention, the difference-in-difference model would not be able to adjust for it. Of course, this would only matter if the intervention or policy had affected early reading literacy skills. . It may be that during 2013, another intervention was implemented in Ibanda and not in treatment districts that would affect students' letter sound knowledge skills more than the SHRP intervention, such as another literacy-focused intervention or a health intervention that might be linked to education outcomes. If this is the case, the impact of SHRP would be underestimated. Again, our performance evaluation will investigate and try to uncover this issue.

CONCLUSION

This report describes the results of the first annual impact evaluation of the SHRP intervention, focusing on Result 1 activities, i.e. those related to Early Grade Reading. We have presented the results of two sets of analyses:

1. Impact of the school-level intervention by comparing treatment schools to control schools within treatment districts.
2. Impact of the school+district-level intervention by comparing treatment schools to comparison schools from comparison districts.

In general, the impact of the SHRP program on early reading skills has been very modest during the first year of implementation. The first set of analyses shows that the school-level intervention:

- Had no impact on any EGRA subtask for the Runyankore/Rukiga, Lango, and Ateso subgroups
- Had a positive impact letter sound knowledge and in the proportion of learners that can identify at least one letter sound in the Luganda subgroup.
- Had a positive impact in the proportion of learners that can identify at least one letter sound in English (although only significant with school fixed effects model)
- Had a positive impact on Luganda word segmenting scores
- Had no impact on higher level literacy skills for the Luganda subgroup or in English for any subgroup

The second set of analyses shows that the school+district-level intervention:

- Had no impact on any EGRA subtask for Lango, and Ateso subgroups
- Had no impact on any EGRA subtask in English for any subgroup
- Had a modest positive impact on Luganda letter sound knowledge, word segmenting, nonword decoding, oral reading fluency and reading comprehension
- Had a negative impact on letter sound knowledge and on the proportion of students who could not read any letter sounds in the Runyankore/Rukiga subgroup
- Had no impact on any other Runyankore/Rukiga subtask

The Performance Evaluation will investigate why positive impacts were only found in the Luganda language subgroup. As mentioned, it is possible that the intervention was implemented differently in Luganda speaking districts than in other areas (in terms of quality and effectiveness of teacher trainings and/or instructional materials for instance) which may explain the fact that the intervention was effective in these districts but not in other districts. The Performance Evaluation will also aim to explain why negative impacts were found for the Runyankore/Rukiga language subgroup for the letter sound knowledge subtask.

As implementation continues to be rolled out, and students move up in grades, it is possible that impacts on higher level literacy could manifest themselves. Additionally, with time, we may see a gradual transfer of these skills from the local language to English, if the hypothesis that mother-tongue based early grade literacy with transition to English is the most effective bilingual literacy acquisition model holds. However, such higher level impacts must be preceded by substantial improvements in students' foundational reading fluency and comprehension skills.

ANNEX A. BALANCE AT BASELINE – TREATMENT VS. CONTROL

Table A.1. Treatment vs. Control, R/R Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.858	7.839	-0.019
	(0.128)	(0.113)	
% of cases with missing age information	0.128	0.080	-0.048
	(0.020)	(0.015)	
Gender of student (female = 1)	0.475	0.499	0.023
	(0.030)	(0.027)	
Number of assets	2.511	2.383	-0.128
	(0.076)	(0.072)	
Lives with both parents (yes = 1)	0.695	0.724	0.029
	(0.027)	(0.024)	
Does not live with mother (yes = 1)	0.184	0.134	-0.051
	(0.023)	(0.019)	
Someone reads to student at home (yes = 1)	0.534	0.538	0.003
	(0.030)	(0.027)	
Attended preschool (yes = 1)	0.433	0.467	0.034
	(0.030)	(0.028)	
Student absent any day in the week prior to assessment (yes = 1)	0.549	0.512	-0.037
	(0.030)	(0.027)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.482	0.405	-0.077
	(0.032)	(0.028)	
Gender of teacher (female = 1)	1.000	0.852	-0.148
	(0.000)	(0.019)	
Teacher teaches in R/R (yes = 1)	0.777	0.763	-0.014
	(0.025)	(0.023)	
Speaks R/R at home (yes = 1)	0.883	0.896	0.013
	(0.019)	(0.017)	
R/R Letter sound score (max = 100)	3.117	3.389	0.272
	(0.335)	(0.271)	
R/R Word segmenting score (max = 10)	2.993	3.585	0.592
	(0.236)	(0.230)	
R/R Nonword decoding score (max = 50)	0.677	0.407	-0.271
	(0.158)	(0.123)	
R/R Oral reading fluency (max = 68)	0.582	0.303	-0.279
	(0.158)	(0.110)	
R/R Reading comprehension score (max = 5)	0.046	0.018	-0.028
	(0.014)	(0.009)	
R/R Listening comprehension	1.504	1.430	-0.073
	(0.059)	(0.055)	
English Letter sound score (max = 100)	2.064	2.920	0.856
	(0.259)	(0.249)	
English Word segmenting score (max = 10)	0.121	0.184	0.063
	(0.030)	(0.032)	
English Nonword decoding score (max = 50)	0.301	0.151	-0.150
	(0.093)	(0.070)	
English Oral reading fluency (max = 68)	0.549	0.191	-0.357
	(0.142)	(0.082)	
English Reading comprehension score (max = 5)	0.004	0.000	-0.004
	(0.004)	(0.000)	
English Receptive vocabulary score (max=20)	5.408	6.436	1.028
	(0.178)	(0.146)	

*p<0.10, **p<0.05, ***p<0.001

Table A.2. Treatment vs Control, Luganda Language subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	6.735	7.251	0.517*
	(0.090)	(0.138)	
% of cases with missing age information	0.117	0.123	0.006
	(0.017)	(0.021)	
Gender of student (female = 1)	0.463	0.462	-0.001
	(0.027)	(0.033)	
Number of assets	2.983	3.072	0.089
	(0.084)	(0.104)	
Lives with both parents (yes = 1)	0.423	0.419	-0.003
	(0.026)	(0.032)	
Does not live with mother (yes = 1)	0.291	0.339	0.048
	(0.024)	(0.031)	
Someone reads to student at home (yes = 1)	0.523	0.538	0.016
	(0.027)	(0.033)	
Attended preschool (yes = 1)	0.760	0.689	-0.071
	(0.023)	(0.030)	
Student absent any day in the week prior assessment (yes = 1)	0.504	0.557	0.052
	(0.027)	(0.033)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.454	0.538	0.084
	(0.027)	(0.033)	
Gender of teacher (female = 1)	1.000	0.911	-0.089
	(0.000)	(0.019)	
Teacher teaches in Luganda (yes = 1)	0.543	0.597	0.055
	(0.027)	(0.032)	
Speaks Luganda at home (yes = 1)	0.780	0.869	0.089**
	(0.022)	(0.022)	
Luganda Letter sound score (max = 100)	1.829	1.852	0.023
	(0.197)	(0.259)	
Luganda Word segmenting score (max = 10)	2.119	3.129	1.010
	(0.187)	(0.272)	
Luganda Nonword decoding score (max = 50)	0.209	0.140	-0.069
	(0.080)	(0.078)	
Luganda Oral reading fluency (max = 68)	0.165	0.140	-0.025
	(0.059)	(0.079)	
Luganda Reading comprehension score (max = 5)	0.017	0.000	-0.017**
	(0.008)	(0.000)	
Luganda Listening comprehension	0.809	0.987	0.179
	(0.046)	(0.063)	
English Letter sound score (max = 100)	1.963	1.547	-0.416
	(0.241)	(0.227)	
English Word segmenting score (max = 10)	0.306	0.538	0.232
	(0.063)	(0.114)	
English Nonword decoding score (max = 50)	0.186	0.178	-0.008
	(0.073)	(0.100)	
English Oral reading fluency (max = 68)	0.780	0.483	-0.297
	(0.139)	(0.136)	
English Reading comprehension score (max = 5)	0.006	0.013	0.007
	(0.004)	(0.009)	
English Receptive vocabulary score (max=20)	7.757	6.775	-0.982
	(0.206)	(0.248)	

*p<0.10, **p<0.05, ***p<0.001

Table A.3. Treatment vs. Control, Lango Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.601	7.535	-0.066
	(0.076)	(0.080)	
% of cases with missing age information	0.102	0.119	0.017
	(0.014)	(0.016)	
Gender of student (female = 1)	0.500	0.503	0.003
	(0.024)	(0.025)	
Number of assets	2.387	2.526	0.139
	(0.049)	(0.056)	
Lives with both parents (yes = 1)	0.728	0.657	-0.071
	(0.021)	(0.024)	
Does not live with mother (yes = 1)	0.131	0.162	0.032
	(0.016)	(0.019)	
Someone reads to student at home (yes = 1)	0.285	0.301	0.016
	(0.021)	(0.024)	
Attended preschool (yes = 1)	0.436	0.336	-0.100**
	(0.024)	(0.024)	
Student absent any day in the week prior assessment (yes = 1)	0.508	0.435	-0.073**
	(0.024)	(0.025)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.379	0.318	-0.061
	(0.024)	(0.025)	
Gender of teacher (female = 1)	0.498	0.508	0.011
	(0.024)	(0.026)	
Teacher teaches in Lango (yes = 1)	0.728	0.745	0.017
	(0.021)	(0.022)	
Speaks Lango at home (yes = 1)	0.892	0.881	-0.010
	(0.015)	(0.016)	
Lango Letter sound score (max = 100)	0.487	0.639	0.152
	(0.095)	(0.151)	
Lango Word segmenting score (max = 10)	0.004	0.005	0.001
	(0.003)	(0.004)	
Lango Nonword decoding score (max = 50)	0.002	0.016	0.013
	(0.002)	(0.016)	
Lango Oral reading fluency (max = 68)	0.029	0.031	0.002
	(0.010)	(0.013)	
Lango Reading comprehension score (max = 5)	0.000	0.000	0.000
	(0.000)	(0.000)	
Lango Listening comprehension	1.889	1.778	-0.111
	(0.048)	(0.053)	
English Letter sound score (max = 100)	0.456	0.397	-0.059
	(0.076)	(0.070)	
English Word segmenting score (max = 10)	0.040	0.013	-0.027
	(0.017)	(0.013)	
English Nonword decoding score (max = 50)	0.000	0.008	0.008
	(0.000)	(0.008)	
English Oral reading fluency (max = 68)	0.004	0.005	0.001
	(0.004)	(0.004)	
English Reading comprehension score (max = 5)	0.000	0.000	0.000
	(0.000)	(0.000)	
English Receptive vocabulary score (max=20)	3.274	3.441	0.166
	(0.141)	(0.130)	

*p<0.10, **p<0.05, ***p<0.001

Table A.4. Treatment vs. Control, Ateso Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.109	6.616	
	(0.091)	(0.090)	-0.492*
% of cases with missing age information	0.156	0.270	
	(0.018)	(0.022)	0.115*
Gender of student (female = 1)	0.546	0.500	
	(0.025)	(0.024)	-0.046
Number of assets	2.222	2.494	
	(0.058)	(0.051)	0.272***
Lives with both parents (yes = 1)	0.648	0.722	
	(0.024)	(0.022)	0.075
Does not live with mother (yes = 1)	0.140	0.151	
	(0.018)	(0.018)	0.010
Someone reads to student at home (yes = 1)	0.276	0.347	
	(0.023)	(0.024)	0.072
Attended preschool (yes = 1)	0.311	0.287	
	(0.024)	(0.022)	-0.024
Student absent any day in the week prior assessment (yes = 1)	0.481	0.543	
	(0.025)	(0.025)	0.062
Teacher absent any day in the week prior to assessment (yes = 1)	0.371	0.491	
	(0.025)	(0.026)	0.120*
Gender of teacher (female = 1)	0.615	0.356	
	(0.025)	(0.023)	-0.258
Teacher teaches in Ateso (yes = 1)	0.857	0.864	
	(0.018)	(0.017)	0.006
Speaks Ateso at home (yes = 1)	0.946	0.967	
	(0.011)	(0.009)	0.020
Ateso Letter sound score (max = 100)	1.719	1.782	
	(0.171)	(0.196)	0.063
Ateso Word segmenting score (max = 10)	1.332	1.380	
	(0.149)	(0.147)	0.049
Ateso Nonword decoding score (max = 50)	0.013	0.067	
	(0.013)	(0.042)	0.054
Ateso Oral reading fluency (max = 68)	0.015	0.050	
	(0.015)	(0.030)	0.035
Ateso Reading comprehension score (max = 5)	0.000	0.010	
	(0.000)	(0.006)	0.010
Ateso Listening comprehension	1.594	1.672	
	(0.048)	(0.045)	0.078
English Letter sound score (max = 100)	1.247	1.395	
	(0.123)	(0.153)	0.147
English Word segmenting score (max = 10)	0.253	0.156	
	(0.059)	(0.033)	-0.097
English Nonword decoding score (max = 50)	0.023	0.050	
	(0.021)	(0.031)	0.027
English Oral reading fluency (max = 68)	0.033	0.038	
	(0.020)	(0.022)	0.005
English Reading comprehension score (max = 5)	0.000	0.000	
	(0.000)	(0.000)	0.000
English Receptive vocabulary score (max=20)	2.115	1.801	
	(0.110)	(0.102)	-0.313

*p<0.10, **p<0.05, ***p<0.001

ANNEX B. BALANCE AT BASELINE – TREATMENT VS. COMPARISON

Table 17. Treatment vs. Comparison, R/R Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.858	7.746	-0.111
	(0.128)	(0.116)	
% of cases with missing age information	0.128	0.073	-0.055
	(0.020)	(0.015)	
Gender of student (female = 1)	0.475	0.484	0.009
	(0.030)	(0.029)	
Number of assets	2.511	2.190	-0.320
	(0.076)	(0.080)	
Lives with both parents (yes = 1)	0.695	0.734	0.039
	(0.027)	(0.026)	
Does not live with mother (yes = 1)	0.184	0.145	-0.039
	(0.023)	(0.021)	
Someone reads to student at home (yes = 1)	0.534	0.534	-0.001
	(0.030)	(0.030)	
Attended preschool (yes = 1)	0.433	0.486	0.053
	(0.030)	(0.030)	
Student absent any day in the week prior to assessment (yes = 1)	0.549	0.535	-0.014
	(0.030)	(0.030)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.482	0.496	0.014
	(0.032)	(0.031)	
Gender of teacher (female = 1)	1.000	0.702	-0.298**
	(0)	(0.027)	
Teacher teaches in R/R (yes = 1)	0.777	0.723	-0.053
	(0.025)	(0.026)	
Speaks R/R at home (yes = 1)	0.883	0.830	-0.053
	(0.019)	(0.022)	
R/R Letter sound score (max = 100)	3.117	1.612	-1.505
	(0.335)	(0.219)	
R/R Word segmenting score (max = 10)	2.993	3.609	0.616
	(0.236)	(0.254)	
R/R Nonword decoding score (max = 50)	0.677	0.000	-0.677*
	(0.158)	(0)	
R/R Oral reading fluency (max = 68)	0.582	0.000	-0.582*
	(0.158)	(0)	
R/R Reading comprehension score (max = 5)	0.046	0.000	-0.046**
	(0.014)	(0)	
R/R Listening comprehension	1.504	1.415	-0.088
	(0.059)	(0.056)	
English Letter sound score (max = 100)	2.064	1.311	-0.752
	(0.259)	(0.185)	
English Word segmenting score (max = 10)	0.121	0.138	0.018
	(0.030)	(0.030	
English Nonword decoding score (max = 50)	0.301	0.010	-0.291*
	(0.093)	(0.010)	
English Oral reading fluency (max = 68)	0.549	0.007	-0.542**
	(0.142)	(0.005)	
English Reading comprehension score (max = 5)	0.004	0.000	-0.004
	(0.004)	(0)	
English Receptive vocabulary score (max=20)	5.408	5.367	-0.041
	(0.178)	(0.157)	

*p<0.10, **p<0.05, ***p<0.001

Table 18. Treatment vs Comparison, Luganda Language subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	6.735	7.121	0.386**
	(0.090)	(0.110)	
% of cases with missing age information	0.117	0.108	-0.009
	(0.017)	(0.018)	
Gender of student (female = 1)	0.463	0.487	0.024
	(0.027)	(0.029)	
Number of assets	2.983	2.802	-0.181
	(0.084)	(0.086)	
Lives with both parents (yes = 1)	0.423	0.490	0.067
	(0.026)	(0.029)	
Does not live with mother (yes = 1)	0.291	0.284	-0.007
	(0.024)	(0.026)	
Someone reads to student at home (yes = 1)	0.523	0.532	0.009
	(0.027)	(0.029)	
Attended preschool (yes = 1)	0.760	0.671	-0.089
	(0.023)	(0.027)	
Student absent any day in the week prior assessment (yes = 1)	0.504	0.513	0.009
	(0.027)	(0.029)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.454	0.505	0.051
	(0.027)	(0.029)	
Gender of teacher (female = 1)	1.000	0.850	-0.150
	(0)	(0.020)	
Teacher teaches in Luganda (yes = 1)	0.543	0.719	0.176**
	(0.027)	(0.026)	
Speaks Luganda at home (yes = 1)	0.780	0.735	-0.045
	(0.022)	(0.025)	
Luganda Letter sound score (max = 100)	1.829	1.454	-0.374
	(0.197)	(0.204)	
Luganda Word segmenting score (max = 10)	2.119	2.467	0.348
	(0.187)	(0.227)	
Luganda Nonword decoding score (max = 50)	0.209	0.176	-0.032
	(0.080)	(0.076)	
Luganda Oral reading fluency (max = 68)	0.165	0.196	0.031
	(0.059)	(0.077)	
Luganda Reading comprehension score (max = 5)	0.017	0.003	-0.014
	(0.008)	(0.003)	
Luganda Listening comprehension	0.809	0.716	-0.093
	(0.046)	(0.048)	
English Letter sound score (max = 100)	1.963	1.075	-0.888
	(0.241)	(0.159)	
English Word segmenting score (max = 10)	0.306	0.294	-0.012
	(0.063)	(0.069)	
English Nonword decoding score (max = 50)	0.186	0.206	0.020
	(0.073)	(0.098)	
English Oral reading fluency (max = 68)	0.780	0.471	-0.309
	(0.139)	(0.151)	
English Reading comprehension score (max = 5)	0.006	0.003	-0.002
	(0.004)	(0.003)	
English Receptive vocabulary score (max=20)	7.757	6.062	-1.695**
	(0.206)	(0.184)	

*p<0.10, **p<0.05, ***p<0.001

Table 19. Treatment vs. Comparison, Lango Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.601	7.697	0.096
	(0.076)	(0.088)	
% of cases with missing age information	0.102	0.127	0.025
	(0.014)	(0.017)	
Gender of student (female = 1)	0.500	0.492	-0.008
	(0.024)	(0.025)	
Number of assets	2.387	2.122	-0.265***
	(0.049)	(0.058)	
Lives with both parents (yes = 1)	0.728	0.619	-0.109
	(0.021)	(0.025)	
Does not live with mother (yes = 1)	0.131	0.145	0.015
	(0.016)	(0.018)	
Someone reads to student at home (yes = 1)	0.285	0.277	-0.008
	(0.021)	(0.023)	
Attended preschool (yes = 1)	0.436	0.311	-0.125**
	(0.024)	(0.024)	
Student absent any day in the week prior assessment (yes = 1)	0.508	0.495	-0.013
	(0.024)	(0.026)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.379	0.408	0.030
	(0.024)	(0.026)	
Gender of teacher (female = 1)	0.498	0.301	-0.197
	(0.024)	(0.027)	
Teacher teaches in Lango (yes = 1)	0.728	0.707	-0.021
	(0.021)	(0.023)	
Speaks Lango at home (yes = 1)	0.892	0.886	-0.006
	(0.015)	(0.016)	
Lango Letter sound score (max = 100)	0.487	0.816	0.329
	(0.095)	(0.145)	
Lango Word segmenting score (max = 10)	0.004	0.005	0.001
	(0.003)	(0.004)	
Lango Nonword decoding score (max = 50)	0.002	0.026	0.024
	(0.002)	(0.019)	
Lango Oral reading fluency (max = 68)	0.029	0.106	0.077
	(0.010)	(0.030)	
Lango Reading comprehension score (max = 5)	0.000	0.005	0.005
	(0)	(0.005)	
Lango Listening comprehension	1.889	1.912	0.023
	(0.048)	(0.052)	
English Letter sound score (max = 100)	0.456	0.448	-0.008
	(0.076)	(0.086)	
English Word segmenting score (max = 10)	0.040	0.013	-0.027
	(0.017)	(0.008)	
English Nonword decoding score (max = 50)	0.000	0.034	0.034
	(0)	(0.029)	
English Oral reading fluency (max = 68)	0.004	0.048	0.044**
	(0.004)	(0.017)	
English Reading comprehension score (max = 5)	0.000	0.000	0.000
	(0)	(0)	
English Receptive vocabulary score (max=20)	3.274	3.363	0.088
	(0.141)	(0.110)	

*p<0.10, **p<0.05, ***p<0.001

Table 20. Treatment vs. Comparison, Ateso Language Subgroup

Variable	Treatment Mean (SE)	Control Mean (SE)	Difference (C - T)
Age of student	7.109	6.752	-0.356
	(0.091)	(0.062)	
% of cases with missing age information	0.156	0.248	0.092
	(0.018)	(0.021)	
Gender of student (female = 1)	0.546	0.498	-0.048
	(0.025)	(0.024)	
Number of assets	2.222	2.418	0.197*
	(0.058)	(0.060)	
Lives with both parents (yes = 1)	0.648	0.774	0.126***
	(0.024)	(0.020)	
Does not live with mother (yes = 1)	0.140	0.097	-0.044*
	(0.018)	(0.014)	
Someone reads to student at home (yes = 1)	0.276	0.377	0.101
	(0.023)	(0.024)	
Attended preschool (yes = 1)	0.311	0.192	-0.119**
	(0.024)	(0.019)	
Student absent any day in the week prior assessment (yes = 1)	0.481	0.578	0.098
	(0.025)	(0.024)	
Teacher absent any day in the week prior to assessment (yes = 1)	0.371	0.471	0.100
	(0.025)	(0.025)	
Gender of teacher (female = 1)	0.615	0.788	0.173
	(0.025)	(0.020)	
Teacher teaches in Ateso (yes = 1)	0.857	0.899	0.041
	(0.018)	(0.015)	
Speaks Ateso at home (yes = 1)	0.946	0.981	0.035
	(0.011)	(0.007)	
Ateso Letter sound score (max = 100)	1.719	1.318	-0.401
	(0.171)	(0.176)	
Ateso Word segmenting score (max = 10)	1.332	1.495	0.164
	(0.149)	(0.162)	
Ateso Nonword decoding score (max = 50)	0.013	0.000	-0.013
	(0.013)	(0)	
Ateso Oral reading fluency (max = 68)	0.015	0.000	-0.015
	(0.015)	(0)	
Ateso Reading comprehension score (max = 5)	0.000	0.000	0.000
	(0)	(0)	
Ateso Listening comprehension	1.594	1.660	0.066
	(0.048)	(0.045)	
English Letter sound score (max = 100)	1.247	0.934	-0.313
	(0.123)	(0.117)	
English Word segmenting score (max = 10)	0.253	0.630	0.377
	(0.059)	(0.106)	
English Nonword decoding score (max = 50)	0.023	0.000	-0.023
	(0.021)	(0)	
English Oral reading fluency (max = 68)	0.033	0.038	0.005
	(0.020)	(0.027)	
English Reading comprehension score (max = 5)	0.000	0.000	0.000
	(0)	(0)	
English Receptive vocabulary score (max=20)	2.115	2.127	0.013
	(0.110)	(0.117)	

*p<0.10, **p<0.05, ***p<0.001