ALL CHILDREN READING:
GRAND CHALLENGE FOR DEVELOPMENT

EVALUATION REPORT

GraphoGame™ Teacher Training Service

Agora Center, University of Jyväskylä, Finland

Sub-Grantee: Centre for the Promotion of Literacy in Sub-Saharan Africa, University of Zambia, Zambia

APRIL 2017

Prepared by
School-to-School International (STS)
For All Children Reading: A Grand Challenge for Development

USAID FROM THE AMERICAN PEOPLE
World Vision
Australian Aid
SCHOOL-TO-SCHOOL INTERNATIONAL


NKh

EVALUATION REPORT

GraphoGame™ Teacher Training Service

Agora Center, University of Jyväskylä, Finland
Sub-Grantee: Centre for the Promotion of Literacy in Sub-Saharan Africa, University of Zambia, Zambia

Table of Contents

List of Acronyms

ACR GCD	All Children Reading: A Grand Challenge for Development
CAPOLSA	Centre for Promotion of Literacy in Sub-Saharan Africa
CLSPM	Correct Letter Sounds per Minute
CNWPM	Correct Nonwords per Minute
CWPM	Correct Works per Minute
EGRA	Early Grade Reading Assessment
EOP	End-of-Project
FOI	Fidelity of Implementation
GG	GraphoGame™
GG-TTS	GraphoGame™ Teacher Training Service
IRB	Institutional Review Board
ICT	Information and Communications Technologies
MoGE	Ministry of General Education (Zambia)
NLF	National Literacy Framework (Zambia)
ORF	Oral Reading Fluency
RTI	Research Triangle Institute International
SACMEQ	Southern and Eastern African Consortium for Monitoring Educational Quality
STS	School-to-School International
USAID	United States Agency for International Development

I. Executive Summary

All Children Reading: A Grand Challenge for Development (ACR GCD)—a joint partnership between the United States Agency for International Development (USAID), World Vision, and the Australian Government—is an ongoing series of grant and prize competitions that leverage science and technology to source, test, and disseminate scalable solutions to improve literacy skills of early grade learners in developing countries. Round 2 of ACR GCD, which started in 2014 and continues through 2017, supports technology-based innovations to improve early grade reading outcomes in developing countries.[1] These technology-based innovations feature three focus areas:

1. Mother tongue instruction and reading materials
2. Family and community engagement
3. Children with disabilities

ACR GCD increased its focus on the assessment of early grade reading skills to understand the ability of the technology-based innovations to improve the literacy skills of early grade learners. To measure this, ACR GCD uses the Early Grade Reading Assessment (EGRA) to systematically assess reading skills across all Round 2 grantees. The EGRA is an oral assessment of students designed to measure the most basic foundational skills for literacy acquisition in the early grades—specifically, recognizing letters of the alphabet, reading simple words, understanding sentences and paragraphs, and listening with comprehension. The EGRA methodology was developed under EdData II and has been applied in more than 30 countries and 60 languages. The EGRA instruments used by ACR GCD grantees were adapted to reflect the specific context of each grantee's project.

Agora Center at the University of Jyväskylä—an ACR GCD Round 2 grantee—implemented the GraphoGame™ Teacher Training Service (GG-TTS) project with their local partner, Centre for the Promotion of Literacy in Sub-Saharan Africa (CAPOLSA) at the University of Zambia. GraphoGame™ (GG), a literacy game for children researched in 20 countries, is being used as a sustainable, cost-effective mobile intervention to improve students' basic reading skills. The GG-TTS project combines GG with a training website to improve teachers' use of technology to support literacy instruction. GG-TTS began in January 2016 and concluded implementation in schools in September 2016. The project's overall goal was to improve the ciNyanja literacy skills of struggling Grade 2 students. In order to understand how these students' literacy outcomes may have changed as a result of the project, School-to-School International (STS), Agora Center, and CAPOLSA conducted EGRA assessments three times throughout the project: baseline data was collected in January 2016, midline data was collected in June 2016, and endline data was collected in September 2016.

Following the endline data collection, STS conducted end-of-project (EOP) interviews with GG-TTS project managers from the Agora Center and CAPOLSA, Grade 2 teachers, Grade 2 students, and other stakeholders. The interviews sought to determine any lessons learned from project implementation, better understand how the project impacted students and teachers, and assess the potential for scalability of the GG-TTS project.

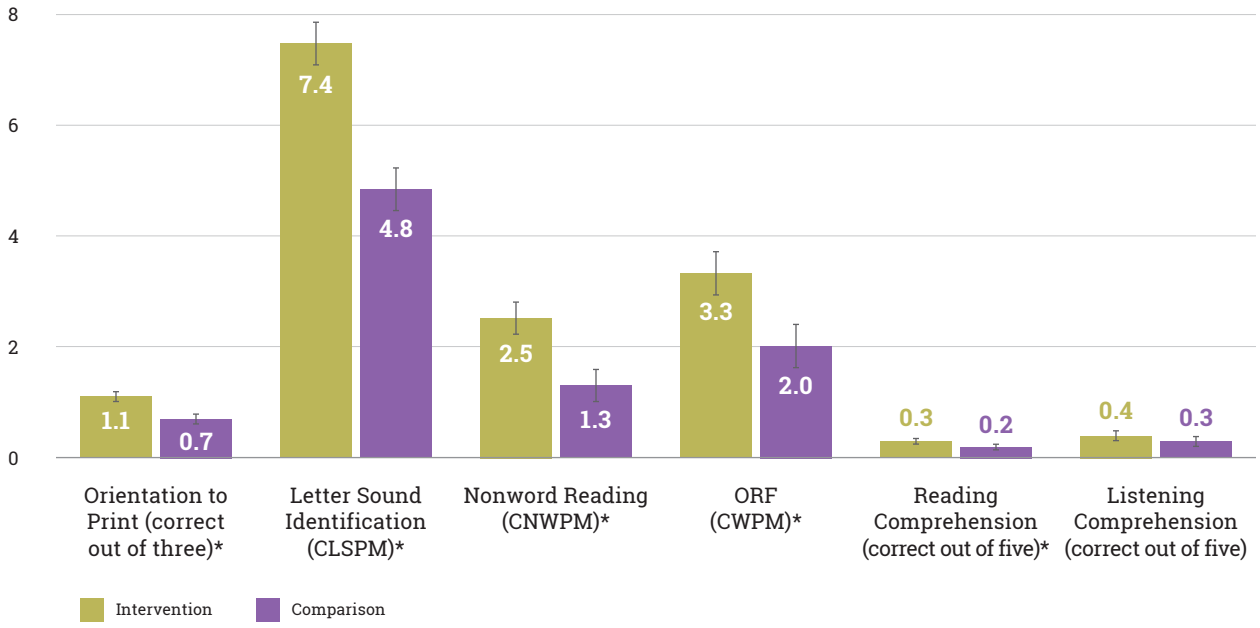
The following report presents a summary of lessons learned from project implementation, comprehensive EGRA results, comprehensive GG assessment results, and scalability assessment results.

1 Retrieved from: http://allchildrenreading.org/about-us/

Key Findings

- **Overall, students who were in the intervention group had significantly larger gains than their peers in the comparison group on five of six EGRA subtasks**, namely: orientation to print, letter sound identification, nonword reading, oral reading fluency (ORF), and reading comprehension (see Figure 1). The gains between the intervention and comparison groups on listening comprehension were not statistically different.
- **The GG-TTS project faced two key challenges in implementation that likely impacted student outcomes:** (1) due to GG programming challenges, most students were unable to progress beyond the letter sound module of the game and therefore were not exposed to word recognition content; and, (2) the teacher training component of the project was introduced with only three months remaining in the intervention rather than made available at the start of the implementation.
- **The GG component of the GG-TTS project was rolled-out inconsistently across schools.** Although the GG-TTS project management team provided extensive monitoring and technical assistance to the schools and teachers, implementation of the project was variable: four schools did not provide the recommended dosage to students, while two schools were able to deliver at least double the recommended dosage.
- **Girls appear to have benefitted more from the GG-TTS project than boys.** The average gains made by girls in the intervention group were significantly larger than the average gains for girls in the comparison group on five out of six EGRA subtasks: orientation to print (1.2 versus 0.7 additional questions), letter sound identification (7.0 versus 4.9 additional letter sounds), nonword reading (2.2 versus 1.1 additional nonwords), ORF (3.1 versus 1.5 additional words), and reading comprehension (0.3 versus 0.1 additional questions). In comparison, the average gains for boys in the intervention group were statistically larger than the average for boys in the comparison group on just two subtasks: orientation to print (1.1 versus 0.7 additional questions) and letter sound identification (8.1 versus 4.6 additional letter sounds).
- **For both the intervention and comparison groups, the proportion of students who were unable to correctly identify a single item or answer a single question, decreased from baseline to endline on nearly all tasks (see Figure 3).** The decreases in students unable to answer a single item correctly were generally larger for the intervention group. The largest decreases from baseline to endline in the intervention group were observed on the ORF (from 90.1 percent to 63.4 percent), orientation to print (from 34.2 percent to 7.8 percent), and nonword reading (from 92.3 percent to 68.1 percent) subtasks. This indicates that students who were exposed to GG-TTS were less likely to receive zero scores than their peers in the comparison group.
- Although students in the intervention group showed significantly larger increases in early grade reading skills compared to students in the comparison group, the gains were not practically significant. **The gains were not large enough to ensure that struggling readers had mastered the skills necessary to read and comprehend words and passages.** Only the gains observed on the letter sound identification subtask were large enough to be considered impactful to students' pre-literacy skills.
- Although the teacher training website was implemented late in the project, **teachers were impacted by its content.** In interviews and questionnaires conducted after the project, teachers in the intervention group reported that they used the internet and a variety of information and communications technologies (ICT) for educational purposes more frequently than before and that their teaching methods better addressed student reading struggles. They also reported spending more hours on literacy instruction at the project's endline than at its baseline. **However, there was insufficient evidence to determine if teachers integrated content learned from the website into their classrooms.**

Figure 1: **Average Gain Scores from Baseline to Endline by Subtask and Group**[2]



II. Project Description

The GG-TTS project was designed to provide a cost-effective, scalable, mobile phone-based teacher training model to improve learning outcomes of Grade 2 struggling readers in Zambia. The model aimed to support the national literacy framework used by the Ministry of General Education (MoGE) in Zambia through use of phonics-based instruction.

GG is a child-friendly digital literacy game that helps students learn letter sounds, syllables, and words. It has been used to support early grade reading skills in Zambia since 2005 when an early version of the game was piloted on computers.[3] In a study conducted by graduate students in Lusaka, the early computer version was found to improve students' knowledge of orthography and spelling if they played the game for two hours during a single month of intervention.[4] In 2010, the Reading Support for Zambia project, which administered a phone-based game with basic features, showed that GG was most effective when the teachers as well was students used the game.[5] In 2013, three different tablet-based GG implementation models were studied; participants from all treatment groups increased their basic reading skills.[6, 7]

2 An asterisk (*) indicates the gain score for the intervention group was significantly higher than the gain score for the comparison group at $p<0.05$. *N* sizes: *N* Intervention=222; *N* Comparison=211.

3 Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology*, 10 (1), 39-60. doi:10.17011/ht/urn.201405281859.

4 Ojanen, E., Kujala, J., Richardson, U., & Lyytinen, H. (2013). Technology-Enhanced Literacy Learning in Zambia: Observations from a Multilingual Literacy Environment. *Insights on Learning Disabilities: From Prevailing Theories to Validated Practices*, 10 (2), 103-127.

5 Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J., Sampa, F., Yalukanda, C., Westerholm, J., Richardson, U., Serpell, R., Lyytinen. H. (2014): The effect of using a mobile literacy game to improve literacy levels of grade one students in Zambian schools. *Educational Technology Research and Development*, August 2014, Volume 62, Issue 4, 417-436.

6 Walubita, G., Nieminen, L., Serpell, R., Ojanen, E., Lyytinen, H., Choopa, M., ... Nakawala-Maumbi, M. (2015). Ensuring Sufficient Literacy Practice with Tablet Technology in Zambian Schools. In P. Cunningham, & M. Cunningham (Eds.), *IST-Africa 2015 Conference Proceedings* (pp. 421-430). IEEE. doi:10.1109/ISTAFRICA.2015.7190560

7 Kauppinen, K-P. (2014) *Investigating language acquisition in Zambia: mapping vowel confusion of a, e and i between English and ciNyanja*. Unpublished Master's thesis. Department of Psychology. University of Jyväskylä. Jyväskylä. Available online https://jyx.jyu.fi/dspace/bitstream/handle/123456789/45157/URN:NBN:fi:jyu-201501271180.pdf;sequence=1

In 2014, GG was implemented via smartphones in rural schools in the Eastern province of Zambia. In that study, teachers from 24 schools participated in a three-day workshop which covered the new curriculum, methods for supporting struggling readers, and also how to use GG. Teachers used GG themselves and provided GG for their students to use at school or on phones at home. For the at-home portion of the project, adult family members also had the opportunity to play GG. The study did not include external pre- and post-testing of the students' literacy skills, but internal GG assessments showed increases on the letter sound and word recognition among children and adult users, as well as small increases among teachers.[8]

Using lessons-learned from previous GG research studies in Zambia, the GG-TTS project was designed to benefit Grade 2 Zambian students who performed at or below literacy standards at the end of Grade 1. It sought to improve their ciNyanja literacy skills so that they were reading on par with their classmates. The program consisted of two primary components:

1. Grade 2 teachers received in-person training on GG and how to use it in their classroom and were asked to administer GG to struggling readers in their classroom using project-provided phones.
2. Grade 2 teachers attended a two-day, in-person training on the GG-TTS teacher training website; they then used the website to complete an online training on techniques to teach literacy in mother tongue languages and support struggling readers.

Implementation of the project started at the end of February 2016 following the baseline data collection. GG-TTS was implemented in 30 schools in a rural district of the Eastern province of Zambia. Teachers were instructed to provide two 15-minute GG playing-sessions per week to their students and to continue the playing-sessions until: (1) their students reached the minimum recommended exposure time of 240 minutes or (2) their student reached 100 percent progress on all the GG learning modules—letter sounds, syllables, and words. Teachers were asked to ensure that students logged in with the correct player account and to keep track of each student's progress through the GG analytics tool on the smartphones. Furthermore, teachers were required to send game logs, which were automatically saved through the phone, to GG-TTS staff each Friday to ensure consistent project monitoring.

In addition to GG-equipped phones, teachers in the intervention group received in-person training and online training materials accessible through smartphones. The teacher training component of GG-TTS sought to improve teachers' capacity to learn using ICTs, enhance their ability to teach literacy in ciNyanja, and to support struggling readers. Specific modules on the website included: GG instructions; supporting struggling readers; story reading and telling as literacy instruction; designing literacy games; and singing as a literacy tool.

During the in-person training, which was held on June 23–24, 2016, teachers received smartphones and internet bundles to access the website. They also received paper handouts with instructions on how to access the website, including usernames and passwords. Following the in-person training, teachers were expected to access the training website regularly and asked to complete all training modules and quizzes by the end of September 2016.

To incentivize full participation in GG-TTS, teachers who completed their website modules and submitted their GG game logs received a certificate. The GG-TTS project concluded in October 2016; smartphones were retrieved from teachers, and endline data collection commenced.

8 Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., ... Lyytinen, H. (2015). GraphoGame – a catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, 6, 671. doi:10.3389/fpsyg.2015.00671

III. Research Purpose and Design

The goal of the GG-TTS project was to improve the ciNyanja literacy skills of Grade 2 students by providing struggling readers with GG and teachers with online training focused on literacy instruction in mother tongue languages and techniques to support struggling readers. The research conducted by STS, Agora Center, and CAPOLSA sought to answer the following questions specific to the GG-TTS project:

1. Does GG-TTS help improve early grade literacy instruction in ciNyanja, as measured through literacy acquisition among students?
2. Can the online GG-TTS materials (accessed via smartphones) help rural Zambian teachers utilize new methods for ciNyanja literacy instruction?

In addition, EOP research was conducted to answer the following ACR GCD supplemental questions:

1. How successful was the rollout of the GG-TTS project?
2. How did the project influence or impact adults' (teachers, parents, community members) knowledge, skills, or attitude regarding their role in helping children read?
3. How did the project influence certain subsets of the student population more than others based on identifiable contextual factors?
4. How much did the development, implementation, and management aspects of the project cost?
5. Is this project and technology suitable to be considered for scaling?

To answer these research questions, STS, Agora Center, and CAPOLSA collected EGRA assessment data at three time points throughout the project. Baseline data was collected in January 2016, midline data was collected in June 2016, and endline data was collected in September 2016. Qualitative and cost data were also collected to answer ACR GCD's supplemental questions.

Sample

Thirty schools in a rural district in the Eastern province of Zambia were selected to participate in the research study. The district office of MoGE provided a list of 135 schools in the district, and a sample frame was constructed using the following selection criteria:

1. School offers reading instruction in ciNyanja for first and second grade students.
2. School did not participate in 2014 national EGRA.
3. School did not participate in the 2014 GG study conducted by Ojanen in the Eastern province.
4. School had cellular coverage by Airtel.

Out of the original 135 schools on the list, 38 schools were eligible for selection. Among those, 30 were randomly selected to participate as either intervention or comparison schools in the ACR GCD project. After the initial 30 schools were selected, the research team randomly assigned schools into the intervention and comparison groups. 15 schools were assigned to the intervention group and received the GG-TTS project; 15 schools were assigned to the comparison group. In total, 37 Grade 2 teachers were included in the research study.

Within Grade 2 classrooms, only struggling readers were eligible for inclusion in the research study. Following Grade 1, the MoGE instructed schools to place students in the following categories according to reading ability using established criteria: outstanding; desirable; minimum; and below minimum. Students who were categorized as minimum or below minimum were designated as "red-level" students, and schools were instructed to focus their efforts on these students. Proportional sampling was conducted to determine how many red-level students per classroom to include in the research sample. This included, totaling the number of red-level students in each classroom and dividing that number by the total number of red-level students in all research study schools. This proportion was then multiplied by the target sample size (300[9] for both the intervention and comparison groups). Once the target number of red-level students per classroom was determined, red-level students were randomly selected to be included in the GG-TTS project. All students who participated in the project were assessed.

Table 1 shows the total number of students assessed in intervention and comparison groups by gender.

Table 1: **Total Number of Students Assessed by Group and Gender**

Assessment	Intervention			Comparison			Total		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
Baseline	112	183	295	123	157	280	235	340	**575**
Midline	83	131	214	80	99	179	163	230	**393**
Endline	88	144	232	96	123	219	184	267	**451**

In total, 575 students were assessed at baseline, 393 students were assessed at midline, and 451 students were assessed at endline. Students' transfers between schools and absenteeism during data collection periods were the primary reasons for attrition.[10]

STS also conducted EOP interviews on November 29–December 9, 2016 to explore the contextual factors that may have impacted implementation and EGRA gains, as well as to determine lessons learned from project implementation and considerations for scalability of the project. EOP interview details are provided in Table 2.

Table 2: **EOP Interview Sample**

Type of Interview	*N*	Description
Project Management	11	2 Agora staff, 8 CAPOLSA staff including research assistants, and 1 BongoHive developer
Stakeholder	4	1 USAID representative and 4 MoGE representatives
Teacher	13	9 teachers and 4 head teachers
Student	48	30 girls and 18 boys
Total	**76**	

9 The sample size of 300 per intervention group was determined based on statistical analysis and the number of mobile phones available for distribution.

10 Because only students with data from baseline and endline were used for gain score analyses, the number of students assessed may differ from the number of students included in results analysis (see Data Analysis and EGRA Results sections).

At EOP, project management interviews were conducted with Agora Center and CAPOLSA staff members. A developer from BongoHive, the Zambian technology and innovation hub that was contracted to design the teacher training website, was also interviewed. Stakeholders were purposively selected based on their knowledge of the GG-TTS project as well as their understanding of funding and policy priorities of USAID/Zambia and MoGE.

Schools included in the EOP interview sample were selected using baseline and endline EGRA results. An average gain score for all subtasks was computed for each of the intervention schools, and schools were ranked from largest to smallest gains on each subtask. Particular attention was paid to gains on the letter sound identification and nonword reading subtasks, as these were most related to the GG-TTS intervention. The eight schools that were selected for the EOP interview sample represented a variety of gains: consistently high across all subtasks, consistently low across all subtasks, and mix of high and low amongst subtasks.

Students within each school were selected for interviews based on: the student's participation in the GG-TTS intervention; the student's attendance at school on the day of interviews; the student's gains on the letter sound identification EGRA subtask; and the amount of time the student spent playing GG throughout the intervention period. Students were categorized as having high, average, or low gains on the letter sound identification subtask, and as receiving a high, average, or low GG dosage based on the number of minutes they spent playing. After determining which participating students were in attendance on the day of the interviews, STS selected an average of five students per school with a variety of EGRA gain and GG dosage categorizations. All teachers who participated in the GG-TTS project at the selected schools were interviewed, and, when available, head teachers were also interviewed.

IV. Fieldwork Preparation and Data Collection

Early Grade Reading Assessment Instrument

The ciNyanja EGRA instrument used for the GG-TTS project was developed and adapted by the Research Triangle Institute International (RTI) in 2014.[11] The EGRA instrument used at baseline and endline consists of six subtasks: orientation to print, letter sound identification, nonword reading, ORF, reading comprehension, and listening comprehension.

Additional Instruments

Supplementing the EGRA, several additional instruments were administered to participants at baseline and/or at endline: a student demographic questionnaire; two teacher questionnaires; and an ACR GCD student questionnaire. Results from the GG ciNyanja letter sound and word recognition assessments were also collected from students at baseline and endline. Finally, the GG-TTS team conducted classroom observations in September 2016 to understand teacher instructional practices.

The student demographic questionnaire was administered at baseline to intervention and comparison schools to collect background information on students and to provide a better understanding of the student population. Questions were related to the students' socio-economic status, ICT use in the home, and the home learning environment.

Following the endline EGRA, assessors administered the ACR GCD student questionnaire to students in the intervention and comparison groups. This questionnaire contains questions related to nine key themes that are standard across all ACR GCD Round 2 projects: language consistency, socio-economic status, parental literacy,

11 The ciNyanja EGRA is also being used by ACR GCD grantee Creative Associates International for the Makhalidwe Athu project.

parental reading support, reading materials access, teacher reading support, disposition to reading, technology use, and engagement in the program. The GG-TTS ACR GCD student questionnaire was adapted to provide information on factors that may have influenced a student's EGRA scores at baseline or endline.

Both teacher questionnaires—a general questionnaire and an ICT questionnaire—were administered at baseline and at endline to teachers in the intervention and comparison groups.[12] The questionnaires were used to better understand teachers' knowledge and attitudes towards teaching and their use of ICT. The general teacher questionnaire collected information about teachers' educational background, teaching methods in the classroom, and views on literacy instruction for struggling readers. The ICT teacher questionnaire contained questions related to ICT exposure and training as well as internet and social media use in school and at home. Because the GG-TTS project required use of ICT—specifically smartphones and internet—this questionnaire provided contextual information to help understand implementation.

GG ciNyanja letter sound and word recognition assessments are built into the game and track student data over time. The assessments are dynamic and allow learning content within GG to be modified based on students' answers. Unlike the EGRA, GG assessments do not require students to verbally state the response; instead, through the auditory and visual stimuli from the game, students point to their response on the mobile phone. Students in the intervention and comparison groups played GG at baseline and endline in order to collect data for the letter sound and word recognition assessments. GG assessments were conducted for both intervention and comparison group students to: a) measure the gains of students who played GG versus those who did not, and b) understand the relationship between the foundational literacy skills measured by EGRA and GG.

Finally, the GG-TTS team conducted classroom observations of teachers in the intervention group to assess how the GG-TTS project, particularly the teacher training website component, may have impacted their teaching practices.

Institutional Review Board

Institutional review boards (IRBs) are responsible for ascertaining the acceptability of proposed research in terms of institutional commitments and regulations, applicable laws, standards of professional conduct and practice, and ethical and societal norms. IRBs examine subject recruitment procedures, proposed remuneration, and the informed consent process. IRBs also evaluate the potential risks and benefits to participants outlined in each protocol.

The IRB application to conduct this research study was submitted to the University of Zambia, School of Humanities and Social Sciences Research Ethics Committee. The committee reviewed the application at the end of 2015 and gave written approval in February 2016 for a duration of one year spanning the length of the research design.

Baseline EGRA

The baseline EGRA training and operational data collection took place January 14–February 4, 2016 (see Table 3). Twelve assessors were selected by CAPOLSA to conduct the baseline. All assessors had experience as enumerators and some had previous experience administering GG or conducting EGRAs. All 12, regardless of their past experience, received training on both GG and EGRA and underwent assessor accuracy testing. Assessor accuracy testing[13] is conducted to ensure consistency in scoring between assessors, and it measures the degree to which different assessors agree in their assessment decisions. At least 90 percent consistency is considered the minimum requirement; this means that at least 90 percent of assessors' ratings must be consistent with the list of acceptable responses. All baseline assessors met the 90 percent threshold.

12 See the GG-TTS baseline report for full teacher questionnaires.

13 Assessor accuracy testing is similar to interrater reliability testing. According to the EGRA Toolkit (2nd Edition), assessor accuracy refers to the testing conducted during training, while interrater reliability is conducted during operational data collection.

Table 3: **Fieldwork Preparation and Data Collection Timeline**

Task	Dates
GG training	January 14–15, 2016
Baseline assessor training incl. pilot test and assessor agreement	January 18–22, 2016
Baseline EGRA operational data collection	January 25–February 4, 2016
Midline EGRA refresher training and operational data collection	June 13–24, 2016
Endline EGRA refresher training and operational data collection	September 19–30, 2016
EOP interviews	November 29–December 9, 2016

Following assessor training and pilot testing of the EGRA, assessors collected operational baseline EGRA data between January 25 and February 4, 2016.

Midline and Endline EGRA

A midline EGRA data collection was conducted in all 30 schools June 13–24, 2016.[14] Assessors received a one-day refresher training on the EGRA and GG assessment. During this training, assessor accuracy data were collected to ensure the consistent and accurate scoring of EGRAs across assessors. Assessors made appointments with teachers participating in the project in advance to maximize attendance on the days during which assessments were to be conducted. The midline EGRA differed from the baseline and endline instrument: it did not include the orientation to print subtask, and items in the other subtasks were different. The two EGRA instruments were not equated, so the results from the baseline and endline EGRAs are not directly comparable to the results from the midline EGRA (see Midline EGRA results).[15]

The endline EGRA was conducted September 19–30, 2016 in all project schools. Prior to operational data collection, assessors participated in a one-day refresher training, including assessor accuracy testing and review sessions on the EGRA instrument and administration and on the GG assessment. During the endline operational data collection, assessors also administered general teacher questionnaires and ICT teacher questionnaires. Assessors conducted classroom observations in 11 intervention schools October 3–5, 2016 to collect information on teacher practices.

End-of-Project Interviews

STS, with assistance from CAPOLSA, conducted EOP interviews several weeks after the endline EGRA data collection. The purpose of the interviews was to explore the variation in implementation and results between schools and students. EOP interviews were conducted with four key populations: project management, stakeholders, students, and teachers and head teachers. These interviews took place between November 29 and December 9, 2016.

14 The midline EGRA was originally intended to measure durability. Because it was expected that students would reach the recommended dosage of GG playing time by June or July 2016, the gains from baseline to midline and from midline to endline could be compared in order to understand if gains made while playing GG lasted after students stopped playing GG. However, because of implementation challenges, a majority of students played GG until the end of the project (see Project Implementation section). As a result, midline data is used as an additional point of comparison rather than a measure of durability.

15 When EGRA instruments are not the same, subtasks are equated using a statistical process to ensure that student scores on each subtask are comparable even though the instruments are different. In the case that the instruments are not equated, results should not be directly compared.

Project management interviews consisted of open-ended questions related to general project information and the intervention timeline, characteristics of the implementing organizations, perceptions of project design and implementation quality, and considerations for scalability. Agora Center and CAPOLSA staff members directly involved in the implementation of the GG-TTS project were interviewed. Additionally, the BongoHive consultant responsible for the creation and design of the teacher training website was interviewed. Stakeholder interviews were guided by key questions related to the scalability of the GG-TTS project, particularly in relation to the relevance of the project to education policy priorities and the relative advantage of the project in comparison with existing policies or programs.

Students and teachers were interviewed November 29–December 2, 2016. Students were asked 23 open-ended questions related to their engagement in the GG-TTS project, their disposition to reading, their access to reading materials, and whether they like learning with GG more than other classroom teaching practices. Teachers were asked 28 open-ended questions related to their use of GG in the classroom, challenges they faced in implementing the project with fidelity, the training they received on GG and on the teacher training website, their teaching practices in the classroom, and the project's potential for scalability. When present, head teachers were also interviewed about the GG-TTS project, challenges faced by their teachers in implementing the project, and the scalability of the project.

V. Project Implementation

The GG-TTS project began at the end of February 2016, and ended in October 2016. This section presents implementation challenges, solutions, and successes that help answer the ACR GCD research question: *How successful was the rollout of the intervention?*

Development

An initial iteration of GG was used in Zambia in 2005, and since then GG has been in use as an ICT-based literacy tool for many years. Because GG was a well-established tool that had already been used by ciNyanja-speaking students in Zambia, few financial or personnel resources were required for the development of GG.

The teacher training website was developed specifically for the GG-TTS project, and its development presented several challenges. Challenges at the beginning of the project included a lack of clarity about the website design and platform—specifically, how teachers should access the website and what content should be included. Agora Center and CAPOLSA's program managers also expressed challenges in conceptualizing the website's training content, with particular concerns surrounding what topics should be included and how to structure that content in a progressive manner. Reviewing, proofreading, and editing the website content required a significant investment of time from Agora Center and CAPOLSA staff.

Project management faced additional challenges selecting and finalizing a website platform where the training content would be displayed. Ultimately, the project decided to contract BongoHive, who recommended using Moodle—an open-source learning platform—for the teacher training website. Once developed, BongoHive conducted a focus group with teachers outside of Lusaka to pilot test the website and identify any challenges teachers faced with its usability. Major takeaways were incorporated prior to rolling-out the website to teachers participating in the GG-TTS project.

Although the website was initially intended to launch at the start of implementation, its development challenges significantly delayed this component of the GG-TTS project. The website was not completed until early June 2016, and teachers began using it at the end of June 2016, allowing roughly three months for use before the project's end.

Implementation

Although the original project design called for the two key components—GG and the teacher training website—to be rolled-out simultaneously, during implementation they ended up staggered because of delays in the website development. The GG component had previously been extensively piloted in schools in Zambia, so there were no significant delays between the baseline assessment and the roll-out of the GG component. However, because the GG-TTS model differed in its approach from previous GG projects—it required teachers to provide more intensive exposure and GG dosage to students—there is evidence that some teachers struggled to find adequate time to offer GG to their students outside of classroom hours. During EOP interviews, many teachers expressed that it was a challenge to find time before, during, or after school to deliver GG to their students. The GG-TTS model did not dictate to teachers a specified timetable for GG application in the classroom to allow teachers to have the autonomy to decide what worked best for their students; however, because of differing engagement and contextual challenges, it is clear that this variability created differing levels of GG playing dosage across schools.

An additional implementation challenge was related to the advancement of students through GG's different modules. According to the project design, students should have advanced through GG's letter sounds to syllables to word recognition modules within a six- to eight-week time period. This meant that students should

16 By the end of implementation, all teachers had advanced through all modules.

have finished all GG modules several months before the end of implementation. For this reason, the research design included a midline EGRA, intended to measure student reading outcomes in the time period immediately following completion of the modules. The endline EGRA, then, was expected to measure the durability and retention of reading gains several months after students had finished playing GG. It became evident late in project implementation that students were not advancing through the modules within the anticipated period of time. In fact, by the end of the project, less than eight percent of students had advanced to the syllables module; and less than one percent had advanced to the word reading module. This lack of advancement was attributed to glitches in GG's internal assessments: students were not able to advance through modules because of GG programming. As a result, almost all students were unable to finish GG within the implementation period, and most were not able to access the modules beyond letter sounds. Because students received most of their GG dosage on the letter sounds module, it is likely that EGRA results were significantly affected by this implementation challenge.

As previously discussed, the implementation of the teacher training website component was significantly delayed due to challenges in the development of the technology. As a result, teachers were not exposed to training content throughout the life of the project. The actual exposure to the teacher training website—measured as the amount of time between the roll-out of the website and the endline EGRA—was a maximum of three months. Teacher progress through the modules varied significantly, and the length of exposure may not have allowed teacher sufficient time to incorporate learned content into their classrooms, both in terms of teacher behavior change as well as student learning.[16]

Management

The GG-TTS project management did not face any notable challenges; this was most likely due to their experience working together in the past. However, staff from both the Agora Center and CAPOLSA expressed that they faced a learning curve on managing the requirements of the project. In particular, monitoring, evaluation, and reporting requirements—as well as the rigor of the research study—were new to both organizations. Regardless, the two organizations divided their work according to pre-established distributions of labor, and there were no major complications in completing work, according to project management interviews.

The design of the teacher training website did present a challenge to the established divisions of labor. Initially, the website's content and programming was overseen by Agora Center, but later it was managed by the CAPOLSA team who worked with BongoHive. Together, CAPOLSA and BongoHive developed, pilot tested, and finalized the website, with remote feedback provided by Agora Center and significant support for website content development from World Vision and STS. This lack of clarity over who managed development of the website may have added to delays in website development.

Fidelity of Implementation

Fidelity of implementation (FOI) is the accurate and consistent application of an agreed upon procedure; FOI research is used to assess the degree to which a project is implemented as intended. Measuring FOI helps implementers and researchers understand and differentiate between what was supposed to happen and what actually happened during the life of a project. When FOI is high and gains in the intervention group are significantly greater than gains in the comparison group, it is possible to attribute student learning impact to the project. This, in turn, makes it possible to recommend scaling the project for future implementation. FOI also makes it possible to identify which components of an intervention are most strongly associated with outcomes. When FOI is low, implementers and researchers are unable to attribute any impacts observed to the project or to assess the quality of the design of the project. Beyond attribution at the end of the project, FOI can also be

16 By the end of implementation, all teachers had advanced through all modules.

coupled with monitoring and evaluation exercises to provide feedback to implementers during the project cycle. In the case of low FOI, this intermediary feedback allows implementers to improve adherence to project design moving forward.[17]

As part of their projects, all ACR GCD Round 2 grantees conducted FOI research during the implementation period. The primary objectives of FOI for grantees were to:

1. Understand what FOI is and why it is important throughout the life of the project;
2. Identify essential components, activities, and questions for each phase of project implementation; and,
3. Create relevant, project-specific FOI tools to monitor participant adherence to the intervention plan.

STS held a series of FOI meetings with each ACR GCD grantee to develop project-specific FOI tools and an implementation plan for FOI research. After finishing the FOI sessions, ACR GCD grantees were expected to pilot test their FOI tools and collect data. Grantees were advised to collect a minimum of one round of FOI data; two or more rounds of data collection was ideal. The data collected served several purposes:

1. To indicate where revisions in data collection tools were necessary;
2. To highlight where improvements in implementation were needed; and,
3. To attribute impact when combined with assessment data.

The Agora Center and CAPOLSA project management teams participated in a series of FOI calls, developed FOI tools, and collected FOI data on the GG-TTS project. Because of the project timeline, FOI data was collected immediately prior to the endline EGRA. As a result, findings were not used to improve implementation. Where possible, results from FOI tools have been incorporated into project findings (see the EGRA Findings and Teaching Method Results Sections).

VI. Data Analysis

Before data analysis began, EGRA results from baseline and endline were matched with GG assessment results to ensure student data was available for both assessments. Records in the EGRA dataset were matched with the anonymous GG game log data using the school code and individual GG ID code. Student data were matched in order to calculate gain scores over time; only students who had data at baseline and endline were used in gain score and zero score analyses.[18] Additionally, midline mean scores are only presented for students with data at baseline and endline.

For each subtask, decision rules were applied to assess whether outliers would need to be removed. For example, if the time remaining for a timed subtask resulted in a fluency rate that was outside a reasonable range, then that student's fluency rate was not included in the analyses. Reasonable ranges were based on multiple factors, including the rate at which letters or words in the language tested are typically read and the mean fluency rate with and without the outlier data point(s). After consideration of the reasonable ranges in the data, no outliers were removed.

Data were analyzed using Microsoft Excel and IBM SPSS Statistics. Gain scores were computed as the difference between endline and baseline for each subtask, and student reading performance was calculated by comparing the gain scores for students in the intervention group to gain scores for students in the comparison group. Zero scores

17 Creative Associates International, Inc. (2015). *Fidelity of Implementation (FOI) How-to Guide* (unpublished). United States Agency for International Development (USAID). Washington, D.C.

18 Students receive a zero score if they are unable to correctly identify a single item on a subtask. In this report, zero scores are shown as the number of students and as the percentage of the total students unable to correctly identify a single item on a subtask.

were also calculated for all subtasks. The results were reported with standard errors or confidence intervals,[19] where appropriate. Differences between the intervention and comparison groups were tested for significance—average scores on each subtask were compared using analysis of variance (ANOVA) and differences in the proportion of zero scores were compared using the chi-square test for significance.[20] Results with statistically significant differences are reported throughout with an asterisk. Where results are not statistically significant, it is not possible to assume that there is any difference between results for intervention and comparison groups.

Tables 4 and 5 provide details on the EGRA subtasks and GG assessments, including how results were calculated.

Table 4: **EGRA Subtask and Data Analysis Method**

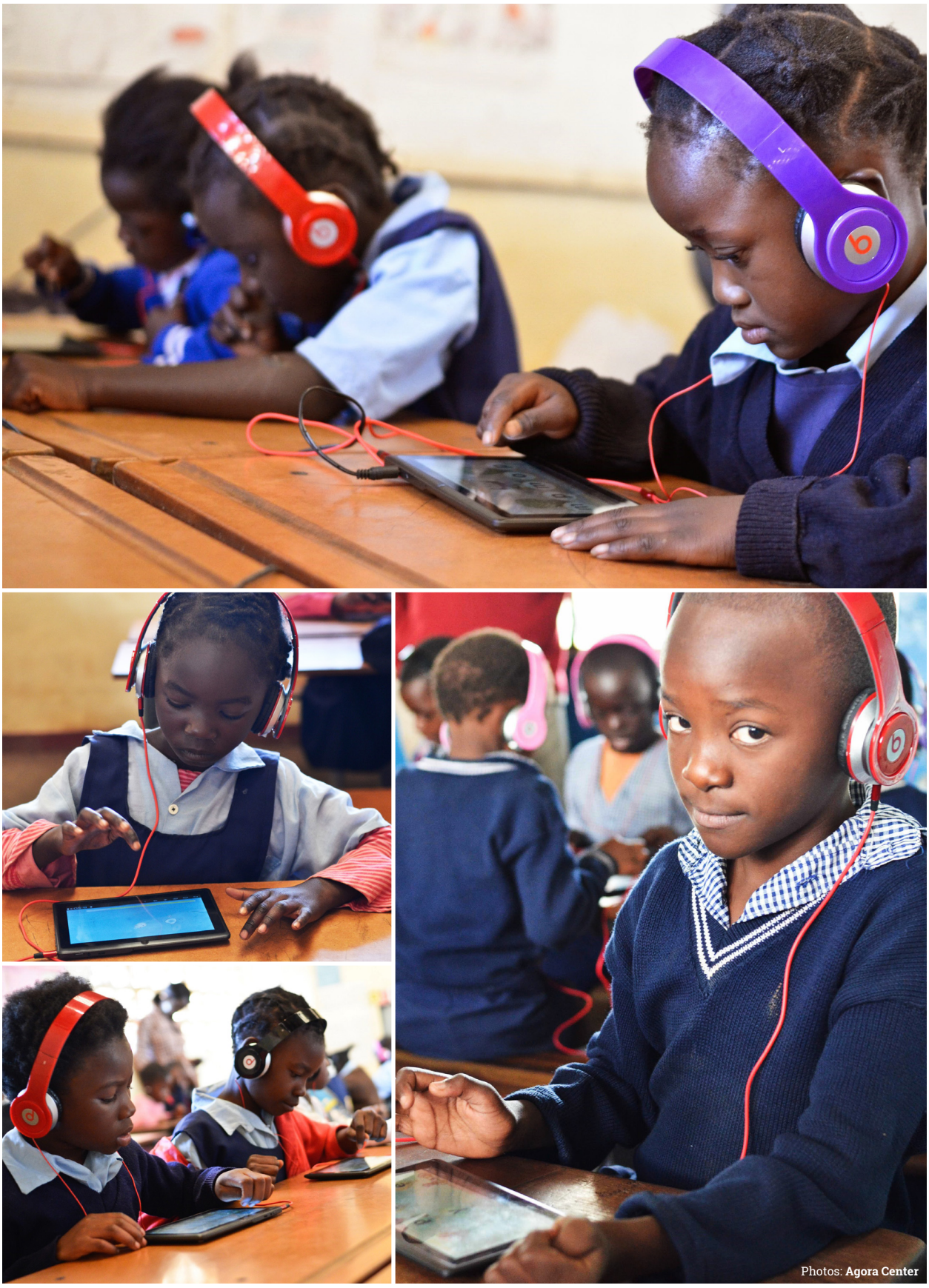
Subtask	Type	Analysis
Orientation to print	Untimed	Orientation to print is measured as the number of questions a student can correctly answer regarding text direction, the concept of a word, or basic knowledge of printed material. Students had the opportunity to answer three questions on this subtask.
Letter sound identification	Timed	Letter sound identification is measured as correct letter sounds read in one minute (CLSPM). Letter sound identification is a measure of alphabet knowledge. Each student had the opportunity to read up to 100 upper and lower case letters.
Nonword reading	Timed	Nonword reading is measured as correct nonwords read in one minute (CNWPM). Nonword reading measures decoding. Each student had the opportunity to read up to 50 one and two syllable nonwords.
Oral reading fluency	Timed	ORF is measured as correct words read in one minute (CWPM). ORF is a decoding and reading fluency measure. Each student had the opportunity to read 40 words. The ORF passage formed the textual basis for the Reading Comprehension subtask.
Reading comprehension	Untimed	Reading comprehension is measured as the number of correct answers verbally delivered to the assessor based on questions asked about the passage read as part of the ORF subtask. Each student had the opportunity to answer five factual questions.
Listening comprehension	Untimed	Listening comprehension is measured as the number of correct answers verbally delivered to the assessor. Listening comprehension is a measure of vocabulary. Each student had the opportunity to answer five questions based on a passage read to them by the assessor.

Table 5: **GG Subtask and Data Analysis Method**

Subtask	Type	Analysis
GG letter sound	Untimed	The GG letter sound assessment is measured as number of correctly identified letter sounds on a mobile phone screen after hearing the phoneme sound in headphones. The maximum score is 24.
GG word recognition	Timed	The GG word recognition assessment is measured as the number of correctly identified syllables and words on a mobile phone screen after hearing the syllable or word in headphones. The maximum score is 24.

19 Standard errors are represented as bars in graphs, indicating the possible range of values of mean scores within an established level of confidence. The confidence interval is noted by the lines at the top of each bar and indicates a range of values that is likely to encompass the true value.

20 Analysis of Variance (ANOVA) is a statistical model that is used to analyze the differences between group means, which helps identify differences in the sample that can be generalized to the population. The chi-square test is a statistical test comparing proportions of zero-score students that were observed in the data against what was expected.

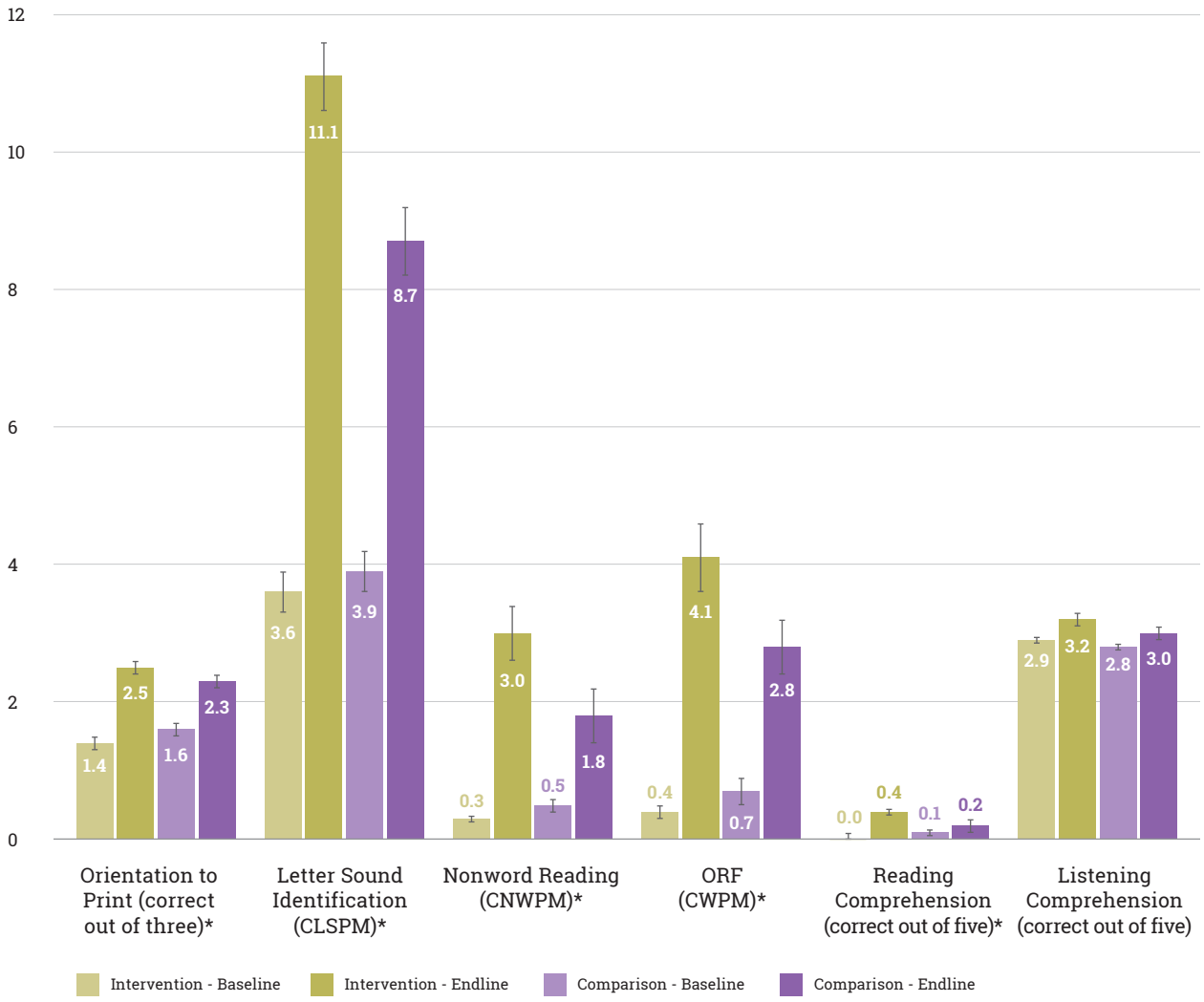

Photos: **Agora Center**

VII. EGRA Findings

The following section presents EGRA[21] results in order to answer the research question: does GG-TTS improve early grade literacy instruction in ciNyanja as measured through literacy acquisition among students?

Overall, the results presented in Figure 2 show that students in the intervention group made significantly greater gains on five of the six subtasks than did students in the comparison group (see also Annex Table E.1). Students in the intervention and comparison groups made gains from baseline to endline on all tasks; however, **gain scores for students in the intervention group were significantly greater than students in the comparison group** for five out of six subtasks: orientation to print, letter sound identification, nonword reading, ORF, and reading comprehension.

Figure 2: **Mean Results by EGRA Subtask and Group at Baseline and Endline**[22]

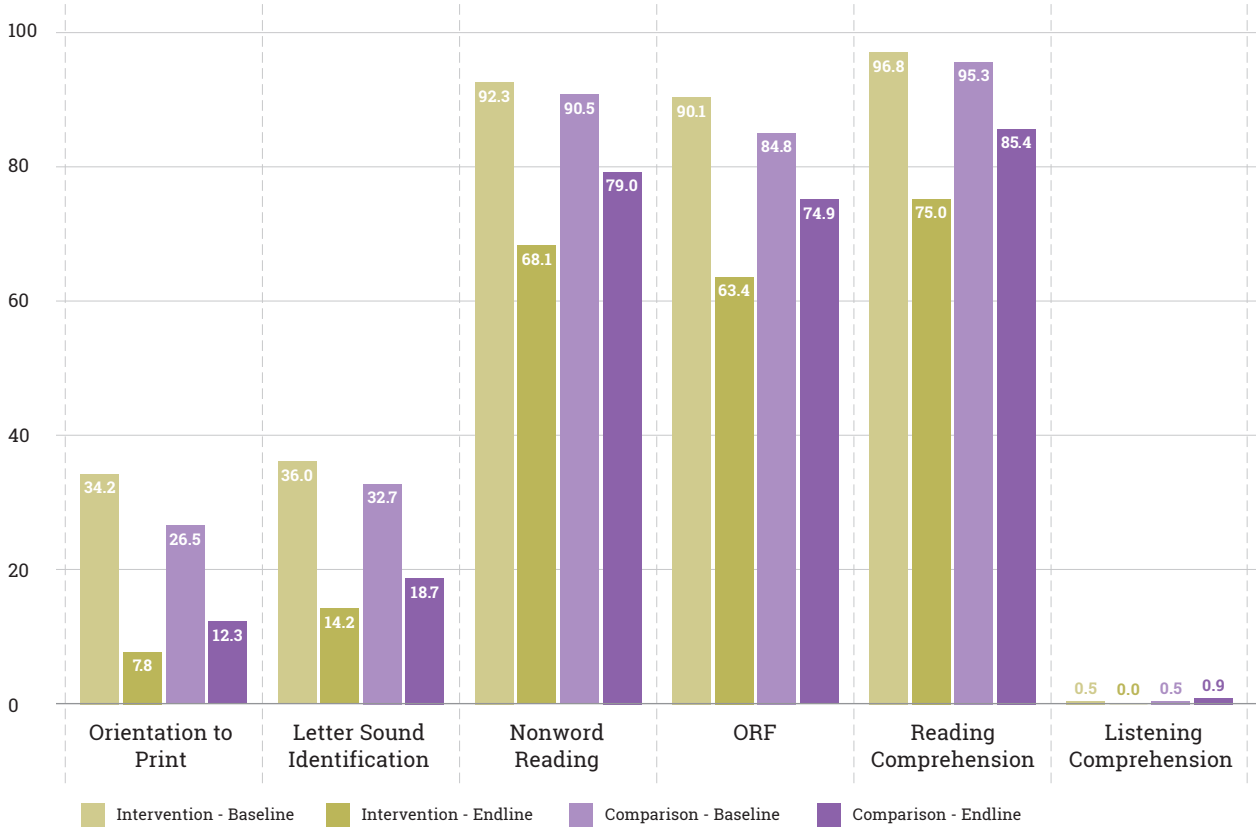


21 EGRA subtask mean scores/fluencies and proportion of zero scores may vary slightly from those reported in the GG-TTS baseline report. Students who did not have data at both baseline and endline were not used for analysis purposes in this report in order to calculate change over time per student.

22 An asterisk (*) indicates the gain score for the intervention group was significantly higher than the gain score for the comparison group at $p<0.05$. *N* sizes: Intervention Group—Baseline *N*=222, Endline *N*=232; Comparison Group—Baseline *N*=211, Endline *N*=219.

Figure 3 shows the percentage of students who received zero scores at baseline and endline. At endline, the proportion of students who received zero scores in the intervention group was significantly lower than in the comparison group for four of the six subtasks: letter sound identification, nonword reading, ORF, and reading comprehension. For the orientation to print and listening comprehension subtasks at endline, there was no statistical difference in the proportion of students who received zero scores in the intervention and comparison groups.

Figure 3: **Percentage of Students Receiving Zero Scores by EGRA Subtask and Group at Baseline and Endline (%)**[23]



EGRA gains and zero scores by subtask and gender are detailed in the following section. Full results detailing the percentage of zero scores by subtask and gender can be found in Annex Table E.4.

EGRA Results by Subtask

Orientation to Print

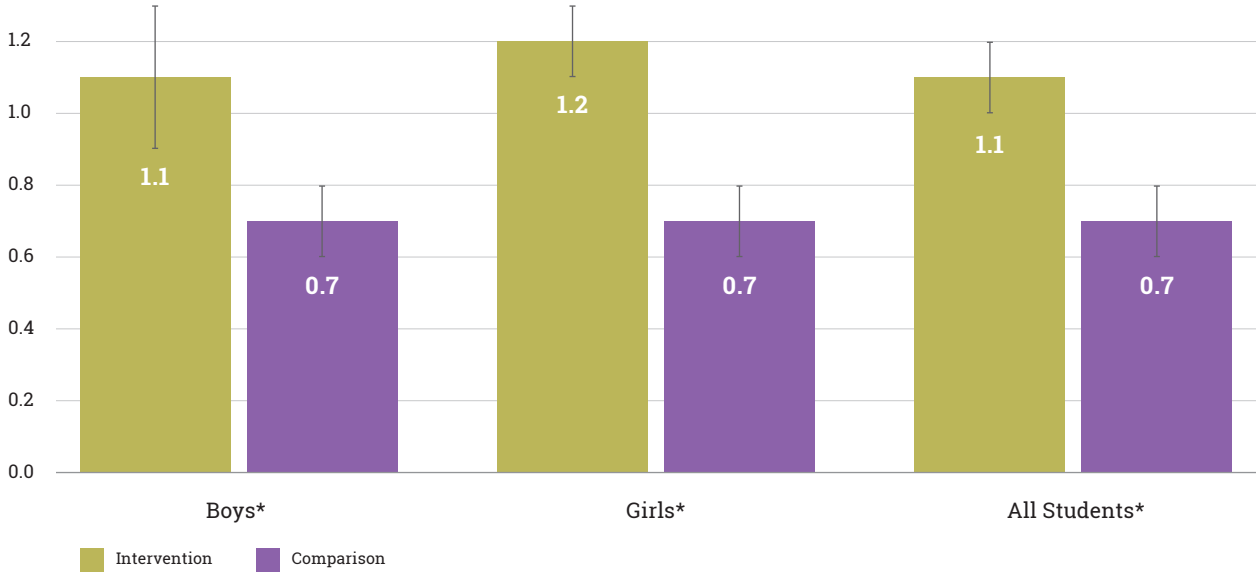
The orientation to print subtask measures students' knowledge of how words are organized on a page, the direction of print (e.g., left to right), and how print materials are organized. In this subtask, students were given a text and asked a series of questions that measured their understanding of how words on a page were organized and read. Students indicated their response to the three questions asked by pointing to the correct part of the page or by indicating the correct direction of reading. This is an untimed task; results are presented in terms of the average questions answered correctly out of three.

23 An asterisk (*) indicates the percentage of students receiving zero scores during the assessment period was significantly different between the intervention and comparison groups at $p<0.05$. *N* sizes: Intervention Group—Baseline *N*=222, Endline *N*=232; Comparison Group—Baseline *N*=211, Endline *N*=219.

Results of the orientation to print subtask are presented in Figure 4. The average number of correct responses on orientation to print increased significantly from baseline to endline for both the intervention group and the comparison group, and students in the intervention group made significantly greater gains than did comparison students. **On average, students in the intervention group answered about 1.1 more questions correctly at endline than at baseline, compared to students in the comparison group** who answered, on average, 0.7 more questions correctly at endline than at baseline. Further, gain scores in the intervention group were significantly higher for both genders.

Girls in the intervention group correctly answered, on average, 1.2 more questions at endline than at baseline, while boys correctly answered, on average, 1.1 more questions on endline than at baseline. Girls and boys in the comparison groups correctly answered, on average, 0.7 additional questions at endline.

Figure 4: **Average Gain Scores by Gender and Group – Orientation to Print (correct out of three)**[24]



Zero scores for the orientation to print subtask from baseline to endline are presented in Figure 3. Both the intervention and comparison groups saw a decrease in the number of students who were unable to answer a single question correctly over time. For the intervention group, the percentage of students who received zero scores decreased from about 34 percent at baseline to nearly 8 percent at endline—equivalent to 58 students who were able to answer at least one question correctly at endline who received zero scores at baseline. For the comparison group, the percentage of students who received zero scores decreased from about 27 percent at baseline to 12 percent at endline, equal to 29 students who were able to answer at least one question correctly at endline who received zero scores at baseline. **Girls in the intervention group experienced the greatest decrease in the percentage of students who received zero scores from baseline to endline**—about a 27 percentage point reduction (see Annex Table E.3). At endline, the difference in the proportion of students who received zero scores between the intervention and comparison groups was not statistically significant, and there was no difference in the proportion of zero scores between boys and girls across groups.

24 An asterisk (*) indicates the average gain score among students in the intervention group was significantly larger than the gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Letter Sound Identification

The letter sound identification subtask measures students' understanding of the alphabetic principle, which states that each letter of the alphabet corresponds to a specific sound. To demonstrate letter sound identification, students must identify the appropriate sounds for each letter. The ability to match letters with correct sounds is critical to reading fluency and comprehension. For this subtask, each student was presented with a stimulus of 100 letters and asked to read as many of the sounds as they could in one minute.[25] Results for this subtask are presented as a fluency rate per minute.

Average gain scores for the letter sound identification subtask are presented in Figure 5. On average, letter sound fluency increased from baseline to endline for students in both groups, and the gains of students in the intervention group were significantly larger than students in the comparison group. Specifically, students in the intervention group were able to read, on average, 7.4 additional letter sounds at endline than at baseline, compared to 4.8 additional letter sounds for students in the comparison group. The difference in fluency gains between the two groups was statistically significant. Both boys and girls in the intervention group had significantly higher gain scores than their peers in the comparison group.

Figure 5: **Average Gain Scores by Gender and Group – Letter Sound Identification (CLSPM)**[24]

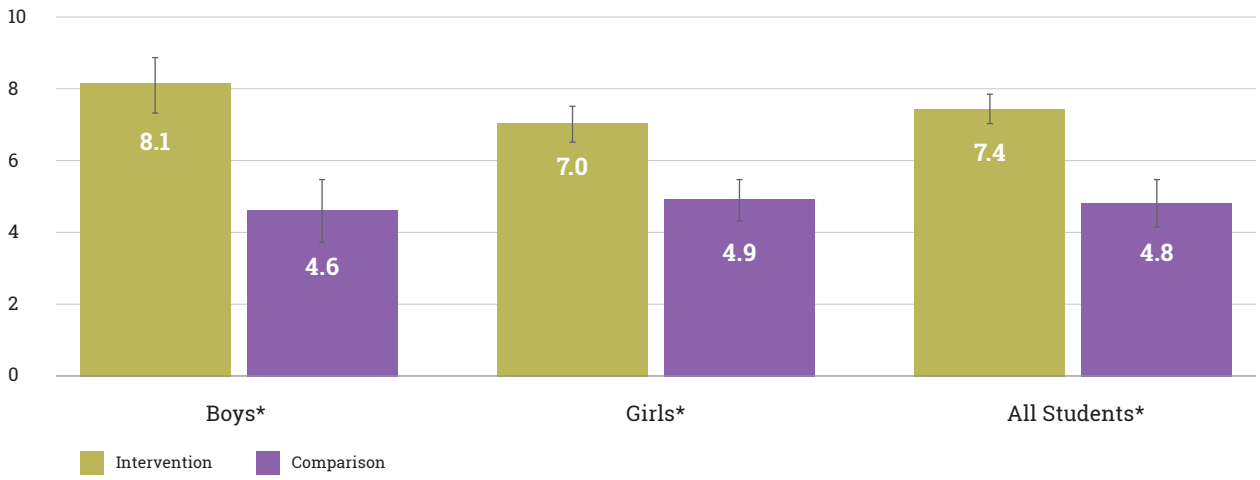
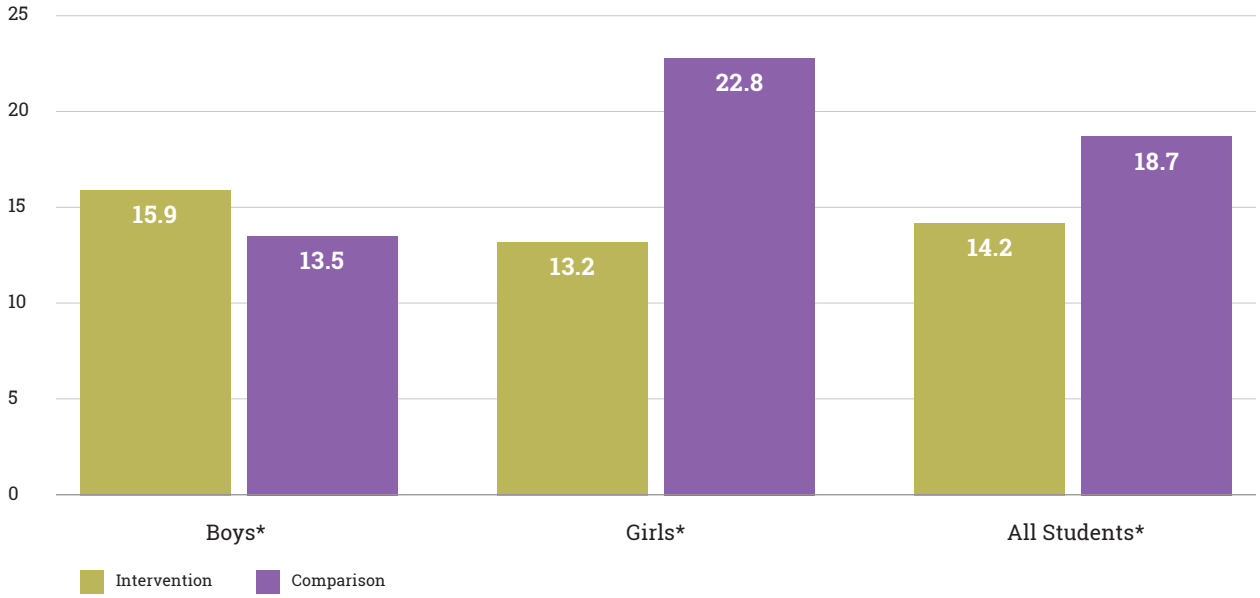


Figure 6 presents the percentage of boys and girls receiving zero scores at endline. Results show that for boys, zero scores on the letter sound identification subtask were comparable at endline, while for girls, the proportion of zero scores was significantly lower in the intervention group than the comparison group. Overall, about 14 percent of students in the intervention group received zero scores on the letter sound identification subtask compared to about 19 percent in the comparison group; this difference is statistically significant. The percentage of students who received zero scores at baseline and endline are presented in Figure 3. For both groups, there was a reduction in the percentage of students who received zero scores on the letter sound identification subtask between baseline and endline—from about 36 percent at baseline to about 14 percent at endline for the intervention group and from nearly 33 percent at baseline to nearly 19 percent at endline for the comparison group. Boys in the intervention group experienced the greatest decrease in the percentage of students receiving zero scores from baseline to endline—about 24 percentage points (see Annex Table E.3).

25 There is an auto stop rule in all the timed EGRA subtasks. In this case, the test was discontinued if a student was unable to correctly name any the first ten letters on the stimulus.

26 An asterisk (*) indicates the average gain score among students in the intervention group was significantly larger than the gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Figure 6: **Percentage of Students Receiving Zero Scores by Gender and Group at Endline – Letter Sound Identification (%)**[27]



Nonword Reading

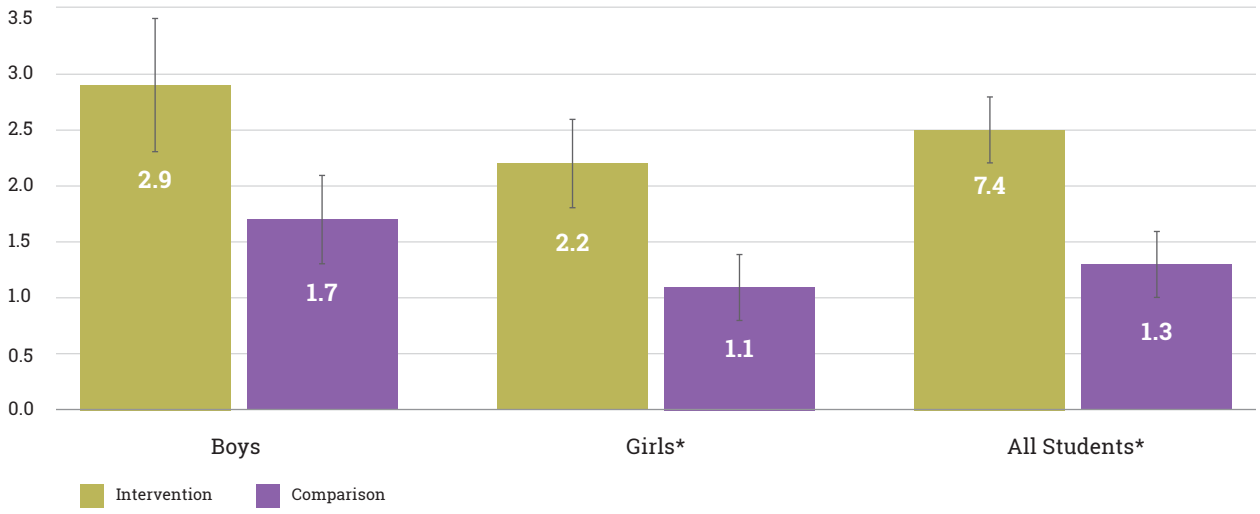
The nonword reading subtask measures students' decoding ability and is designed to present them with words that they would not be able to recognize by sight through familiarity. Many students in the early grades learn to memorize or recognize a range of familiar words by sight alone. Thus, to assess students' decoding skills, they are presented with invented (nonsense) words, which require them to sound out each letter and syllable to decode a word. During this timed subtask, the assessor presented each student with 50 nonwords and asked them to read as many as possible in one minute.[28] Results for this subtask are presented as a fluency rate.

Results for the nonword reading subtask, as measured by correct nonword identification per three minutes, are presented in Figure 7. On average, nonword reading fluency increased significantly from baseline to endline for students in both groups, with students in the intervention group showing significantly larger gains than students in the comparison group. Specifically, students in the intervention group made an average fluency gain of 2.5 CNWPM at endline over baseline compared to a gain of 1.3 CNWPM for students in the comparison group. Girls in the intervention group saw significantly larger gains in nonword reading fluency than girls in the comparison group: about 2.2 additional nonwords per minute for girls in the intervention group compared to 1.1 additional nonwords per minute for girls in the comparison group. No significant differences in average nonword fluency gains were observed for boys on this subtask.

27 An asterisk (*) indicates the percentage of students receiving zero scores in the intervention group was significantly smaller than the percentage of students receiving zero scores in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

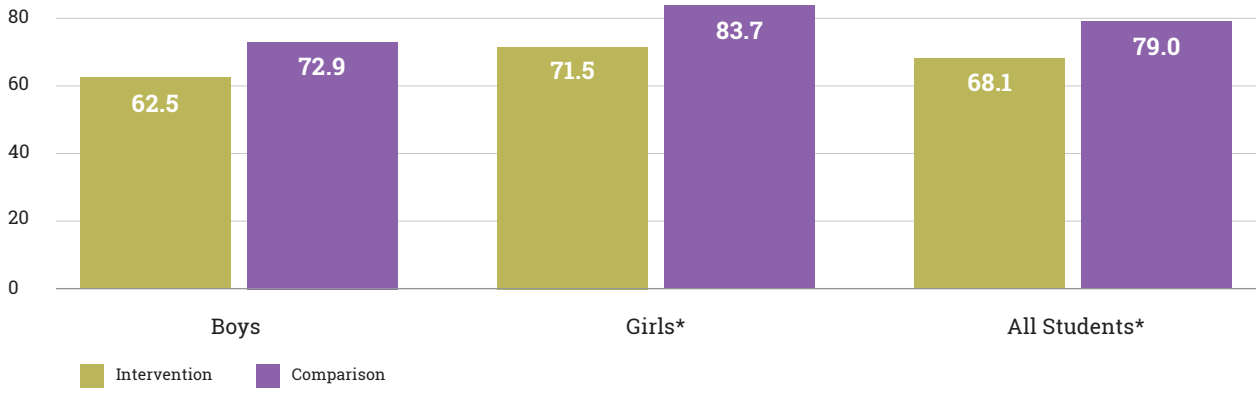
28 After one minute, the student was asked to stop. The subtask was discontinued if a student was unable to correctly read any the first ten nonwords.

Figure 7: **Average Gain Scores by Gender and Group – Nonword Reading (CNWPM)**[29]



Zero score results at endline are presented in Figure 8 and show that, as with the letter sound identification subtask, the proportion of students receiving zero scores was significantly lower for students in the intervention group at endline on the nonword reading subtask. **Overall, the proportion of students who received zero scores at endline was about 68 percent in the intervention group compared to 79 percent for the comparison group.** At endline, the difference in the proportion of students who received zero scores between intervention and comparison groups were statistically significant for girls—about 72 percent versus about 84 percent, respectively—but the difference was not statistically significant for boys.

Figure 8: **Percentage of Students Receiving Zero Scores by Gender and Group at Endline – Nonword Reading (%)**[30]



The greatest reduction in zero scores from baseline to endline on this subtask was observed among boys in the intervention group, for whom the proportion of students receiving zero scores dropped by over 30 percentage points from baseline to endline (see Annex Table E.3).

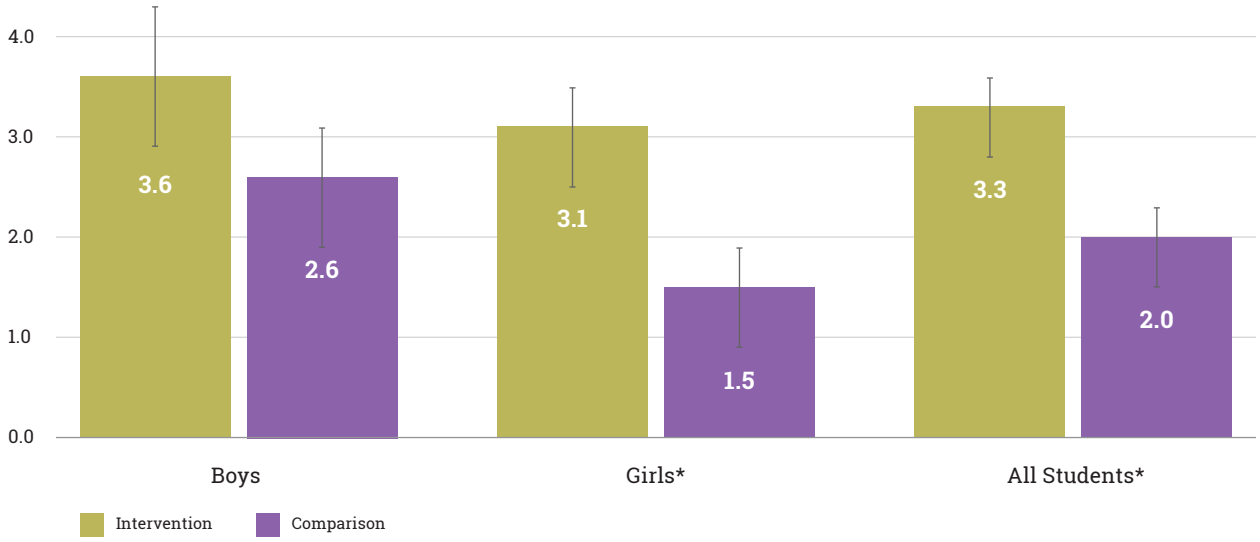
29 An asterisk (*) indicates the average gain score among students in the intervention group was significantly larger than the gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

30 An asterisk (*) indicates the percentage of students receiving zero scores in the intervention group was significantly smaller than the percentage of students receiving zero scores in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Oral Reading Fluency

The ORF subtask measures students' overall reading competence. It is the culmination of translating letters into sounds, merging sounds to become words, linking words to become sentences, relating text to meaning, and making inferences to fill in missing information. A student's ORF score is dependent on the skills in previous subtasks, since children need to have some mastery of orientation to print, letter sounds, and decoding of nonwords to read fluently. Results for this subtask are presented as a fluency rate.

Figure 9: **Average Gain Scores by Gender and Group – ORF (CWPM)**[31]

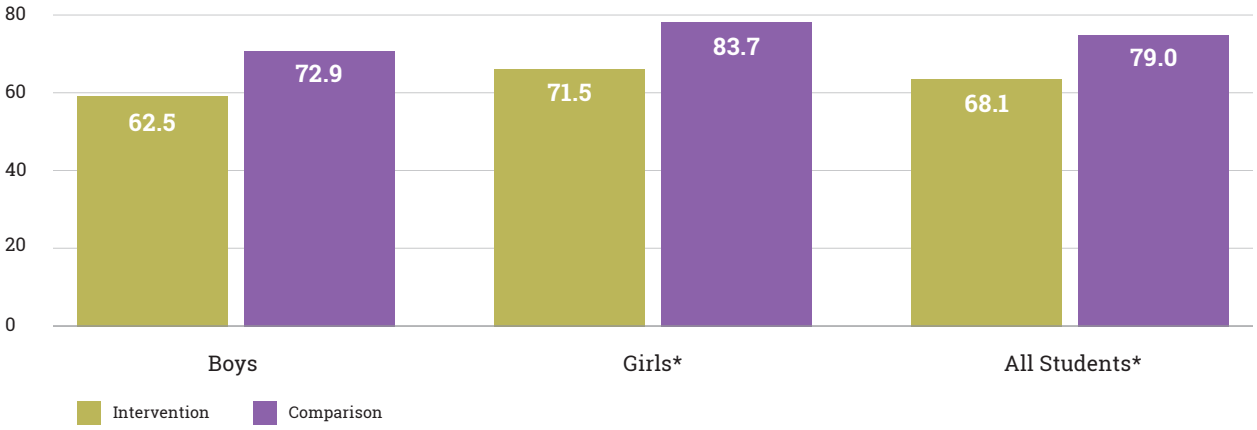


Average gain scores for ORF are presented in Figure 9. On average, ORF increased significantly from baseline to endline for students in both groups; students in the intervention group showed significantly larger gains than students in the comparison group. **Specifically, students in the intervention group made an average fluency gain of about 3.3 CWPM at endline over baseline compared** to a gain of about 2.0 CWPM for students in the comparison group. Girls in the intervention group also saw significantly larger gains in ORF than their counterparts in the comparison group—about 3.1 CWPM compared to about 1.5, respectively. There were no significant differences in gain scores for boys across groups.

The proportion of students who received zero scores at endline for ORF are presented in Figure 10. **At endline, about 63 percent of students in the intervention group received zero scores, while nearly 75 percent of students in the comparison group received zero scores.** As with the letter sound identification and nonword reading subtasks, the difference in the proportion of girls who received zero scores at endline between the intervention and comparison groups was statistically significant, while the difference in the proportion of boys who received zero scores at endline was not. **Further, as with the nonword reading subtask, the greatest reduction in the proportion of students receiving zero scores was observed in intervention group boys:** 26 boys who were unable to read a single word in the ORF passage correctly at baseline were able to read at least one correctly at endline, equal to a 36 percentage point drop (see Annex Table E.3).

31 An asterisk (*) indicates the average gain score among students in the intervention group was significantly larger than the gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Figure 10: **Percentage of Students Receiving Zero Scores by Gender and Group at Endline – ORF (%)**[32]

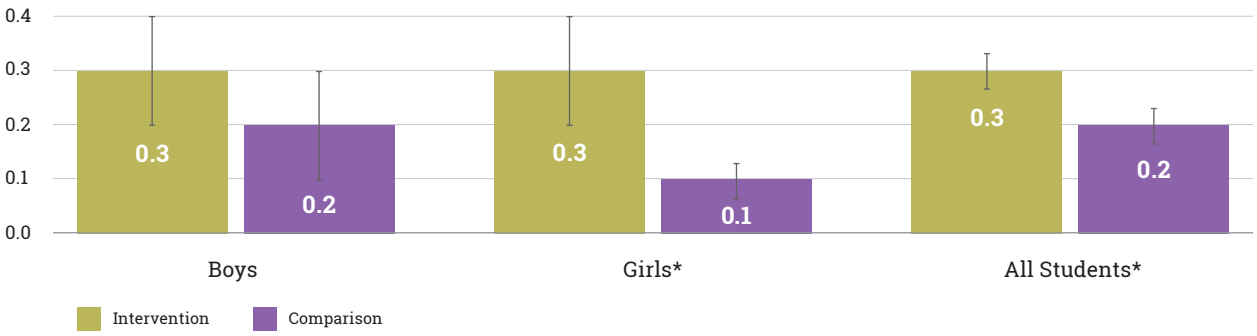


Reading Comprehension

The reading comprehension subtask measures students' understanding of the ciNyanja passage that they read in the ORF subtask. Upon completion of the ORF subtask, assessors read aloud up to five comprehension questions based on the text and asked students to respond verbally. The number of questions asked by assessors was contingent on the amount of text the students read in the ORF subtask.[33] Results for this subtask are presented as the number of correct questions out of five.

On average, reading comprehension scores increased from baseline to endline for students in both groups, with students in the intervention group showing significantly larger gains than students in the comparison group. Results are presented in Figure 11 and indicate that **students in the intervention group made an average gain of 0.3 correct questions at endline over baseline compared to a gain of 0.2 correct questions for students in the comparison group.** Girls in the intervention group also had significantly larger gains in reading comprehension scores than their peers in the comparison group: 0.3 versus 0.1, respectively. While the difference in reading comprehension gain scores for girls was statistically significant, the difference in reading comprehension gain scores for boys was not.

Figure 11: **Average Gain Scores by Gender and Group – Reading Comprehension (correct out of five)**[34]



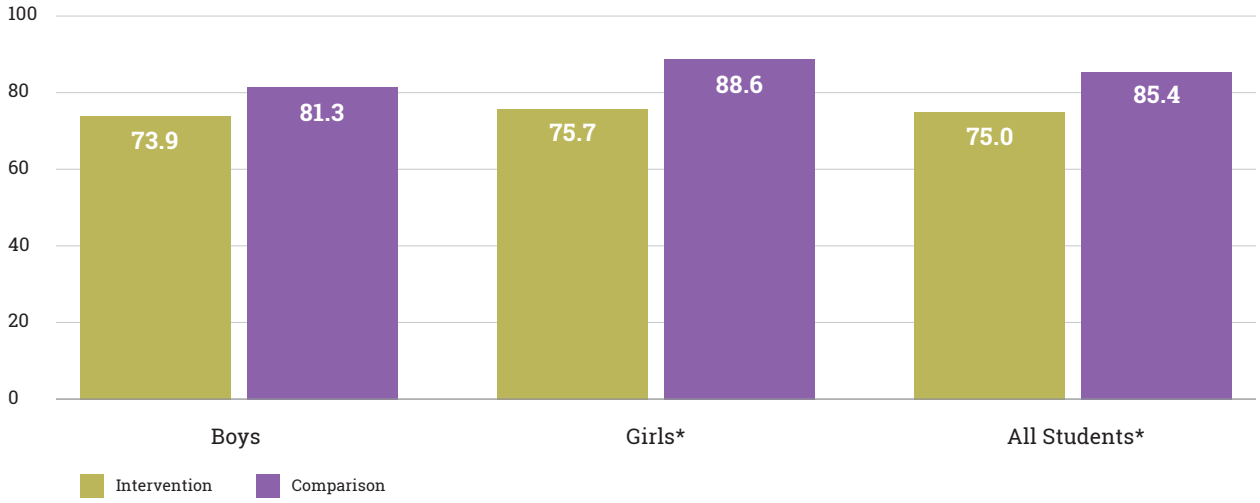
32 An asterisk (*) indicates the percentage of students receiving zero scores in the intervention group was significantly smaller than the percentage of students receiving zero scores in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

33 For example, if a student read the first sentence of text (five words), s/he would be asked the first comprehension question. Similarly, if a student read the whole text (40 words), s/he would be asked all five questions.

34 An asterisk (*) indicates the average gain score among students in the intervention group was significantly larger than the gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

The proportions of students receiving zero scores on the reading comprehension subtask at endline are presented in Figure 12.[35] **Overall, the proportion of students unable to correctly answer a single reading comprehension question correctly was 75 percent for the intervention group compared to about 85 percent for the comparison group;** this difference is statistically significant. Additionally, the proportion of girls who received zero scores in the intervention group was significantly lower than in the comparison group—over 75 percent in the intervention group compared to over 88 percent in the comparison group. The difference in the proportion of zero scores between boys in the intervention and the comparison groups at endline was not statistically significant.

Figure 12: **Percentage of Students Receiving Zero Scores by Gender and Group at Endline – Reading Comprehension (%)**[36]



The proportion of students unable to answer a single reading comprehension question correctly decreased among all groups from baseline to endline (see Annex Table E.3). **The greatest reduction in zero scores was observed among boys in the intervention group, who experienced more than a 22 percentage point decrease in the proportion of zero scores**—equal to 14 boys who were able to answer at least one item correctly at endline compared to zero items at baseline.

Annex Table E.5 presents the number of questions attempted at baseline and endline for the reading comprehension subtask. Frequencies show that in the intervention group there were large differences in the number of students who attempted one to two items at baseline versus at endline. **The number of students in the intervention group who attempted one to two questions increased from 32 to 87 from baseline to endline.** This represents an increase in the number of intervention group girls attempting one to two items from 23 at baseline to 52 at endline and in the number of intervention group boys attempting one to two items from nine at baseline to 35 at endline. In contrast, in the comparison group, the number of boys who attempted one to two items increased from 19 to 28, and the number of girls who attempted one to two items increased from 23 to 26. Overall in the comparison group, the number of students who attempted one to two items increased from 42 to 54.

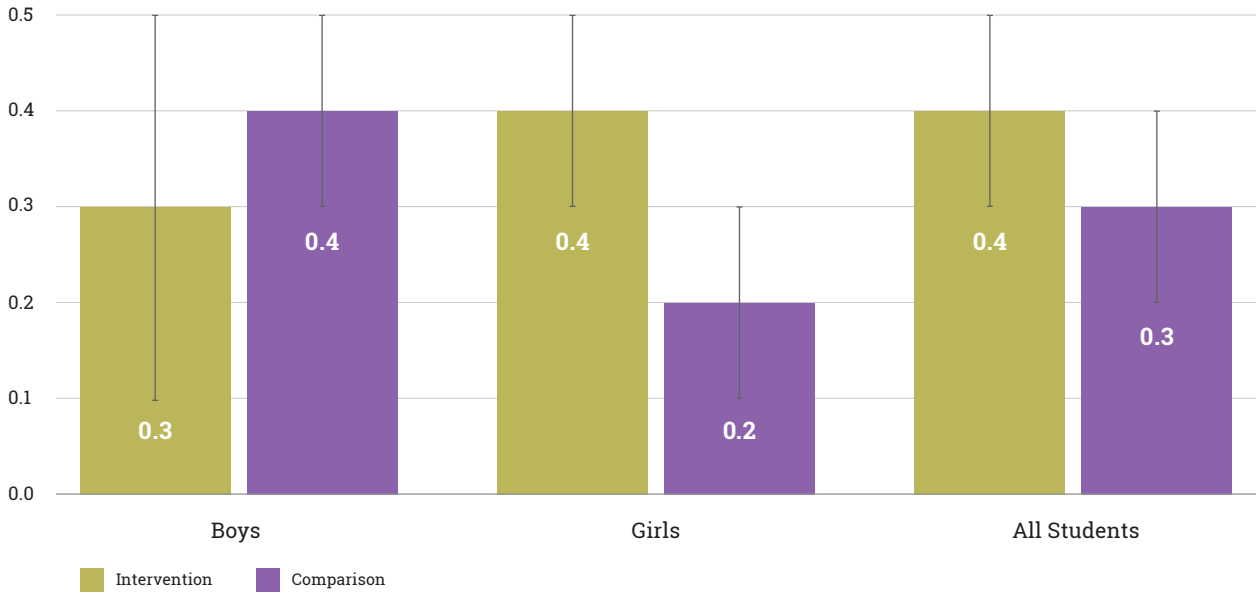
35 Zero scores on reading comprehension reflect two types of students: (1) those who did not read enough of the passage to be asked a single question, and (2) those who read enough to be asked at least one comprehension question but answered all questions incorrectly.

36 An asterisk (*) indicates the percentage of students receiving zero scores in the intervention group was significantly smaller than the percentage of students receiving zero scores in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Listening Comprehension

The untimed listening comprehension subtask measures students' ability to comprehend the meaning of a story read to them orally. Students do not need to know how to read to answer listening comprehension questions, and, as a result, this subtask is an important measure of students' pre-reading abilities because it helps detect obstacles to learning to read such as limited language proficiency, auditory problems, attention deficit, and other difficulties. In this subtask, the assessor read a short passage to the student and asked them to answer five comprehension questions[37] about what they heard. Results for this subtask are presented as the number of questions answered correctly out of five.

Figure 13: **Average Gain Scores by Gender and Group – Listening Comprehension (correct out of five)**[38]



Average gains on listening comprehension are presented in Figure 13. Results showed that on average, listening comprehension scores increased from baseline to endline for all students: students in the intervention group answered an average of 0.4 additional questions at endline, and students in the comparison group answered an average of 0.3 additional questions at endline. **Unlike the other subtasks, gain scores for students in both the intervention and comparison group were comparable across students and genders, meaning that the difference in gain scores for students in the intervention and comparison groups was not statistically significant.**

At baseline and endline, nearly all students across the intervention and comparison groups answered correctly at least one listening comprehension question. In the intervention group, the number of students who received zero scores decreased from one to zero; while the number of students who received zero scores in the comparison group increased from one to two (see Annex Table E.3). At endline, the difference in the proportion of students receiving zero scores in the intervention and comparison groups was not statistically significant.

37 The first three were direct questions (answers found explicitly in the story). The fourth and fifth questions were inferential.

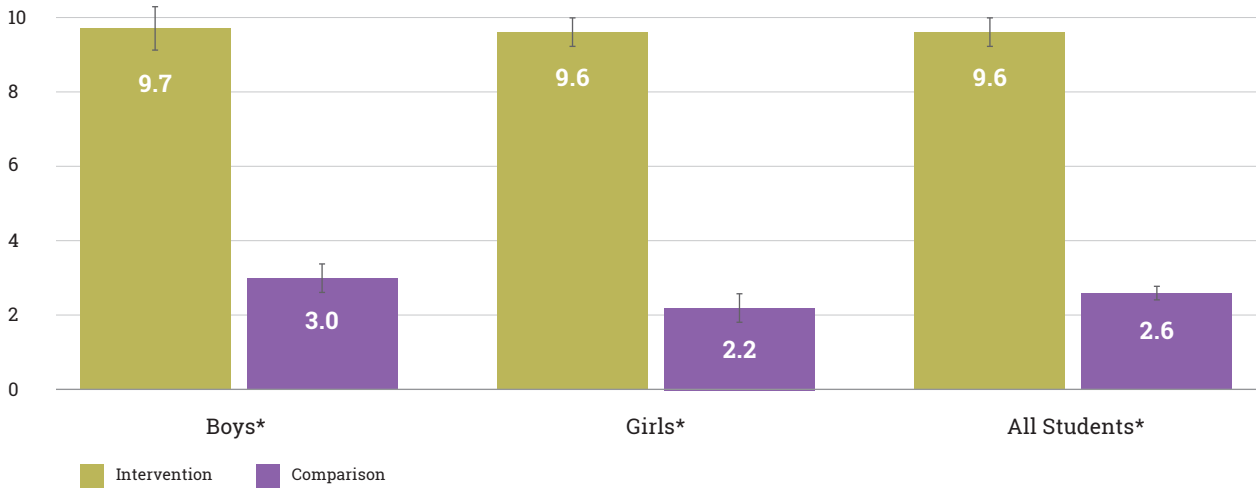
38 *N* sizes: Boys—Intervention Group *n*=82, Comparison Group *n*=92; Girls—Intervention Group *n*=140, Comparison Group *n*=119; All Students—Intervention Group *N*=222, Comparison Group *N*=211.

GraphoGame™ Results by Subtask

GraphoGame™ Letter Sound

On the GG letter sound assessment, 24 letters appear on a screen, and the sounds are heard on headphones one by one. The assessment is presented as three sets of stimuli, always shown in the same order. Students completed the GG letter sound assessment at baseline and endline, and average gain scores are presented in Figure 14. Overall, students in both the intervention and comparison groups made gains on the GG letter sound assessment, and the gain scores from baseline to endline were significantly larger for students in the intervention group than for students in the comparison group. **On average, students in the intervention group were able to identify 9.6 additional letter sounds on the assessment, while students in the comparison group were able to identify 2.6 additional letter sounds.**

Figure 14: **Average Gain Scores by Gender and Group – GG Letter Sound (correct out of 24)**[39]



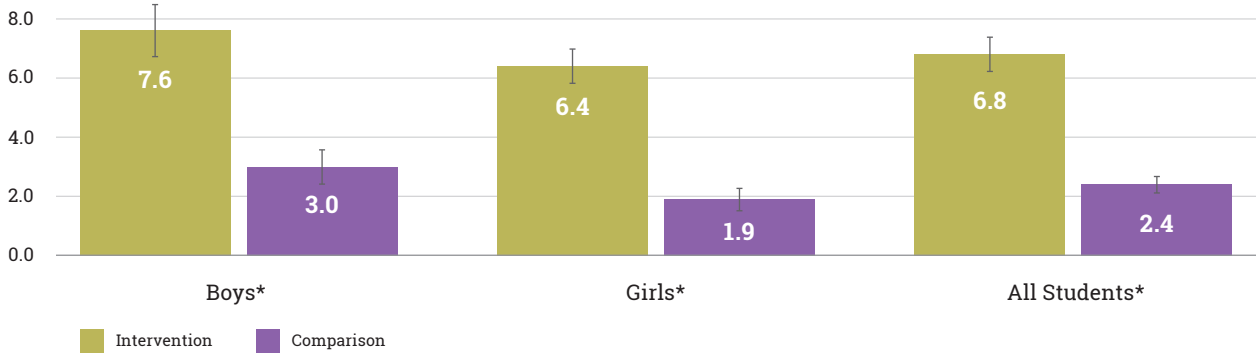
GraphoGame™ Word Recognition

On the GG word recognition assessment, the first eight items presented are two-letter syllables in a consonant-vowel pattern—bu, ka, li, se, yo, co, me, and gi. If a student makes three mistakes within the first eight items, the assessment ends, and, as a result, they are not presented with any words.

Average gain scores between baseline and endline on the GG word recognition assessment are presented in Figure 15. Similar to the GG letter sound assessment, students in both groups improved their scores from baseline to endline, and the gain scores from baseline to endline were significantly larger for students in the intervention group than for students in the comparison group. **On average, students in the intervention group were able to recognize 6.8 additional words on the assessment, while students in the comparison group were able to recognize 2.4 additional words.**

39 An asterisk (*) indicates the average gain score for the intervention group was significantly larger than the average gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

Figure 15: **Average Gain Scores by Gender and Group – GG Word Recognition (correct out of 24)**[40]

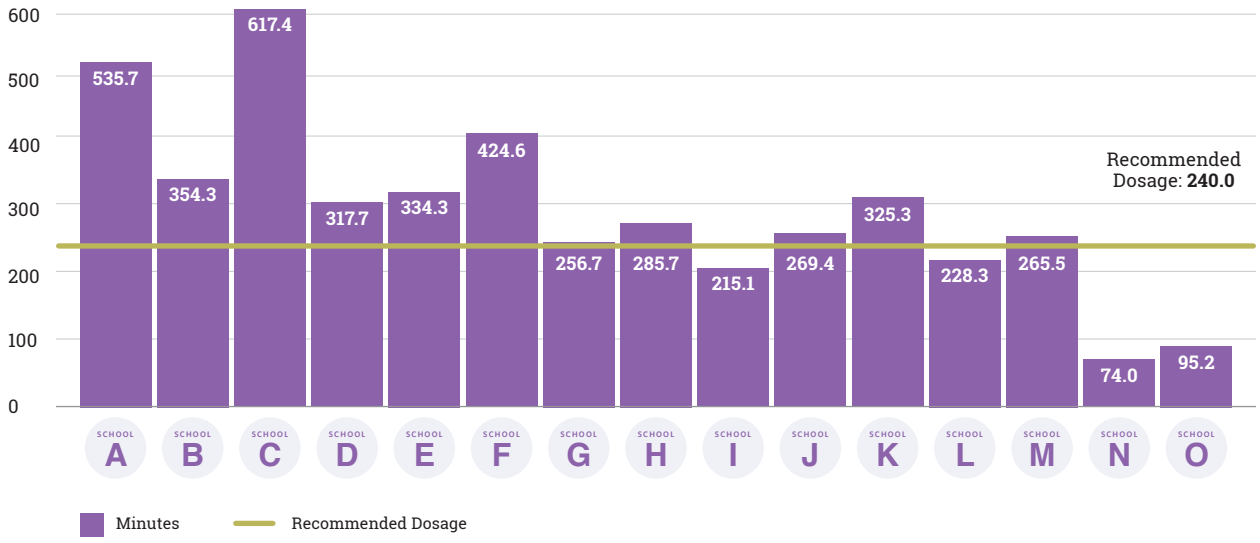


Results by Level of GraphoGame™ Dosage

To better understand how the total amount of GG playing time may have influenced student EGRA gains, results were analyzed against an established recommended GG dosage threshold. Because comparison group students did not play GG, results in this section are only for students in the intervention group.

The level of GG dosage recommended for each student was 240 minutes; this value was established by the Agora Center through their research on learning gains with GG across a variety of projects and countries in which GG has been used. The average number of minutes played per student by school is presented in Figure 16. **Notably, there was wide variation in the average level of GG dosage that students received across schools**—the maximum average dosage by school was about 617 minutes and the minimum average dosage by school was 74 minutes. This variation indicates that a students' exposure to GG was strongly influenced by which school they attended. In fact, in four out of the 15 schools, the average amount of GG usage by students was below the recommended project dosage. This variation was also found when analyzing the average number of days of GG use by school (see Annex Table F.1).

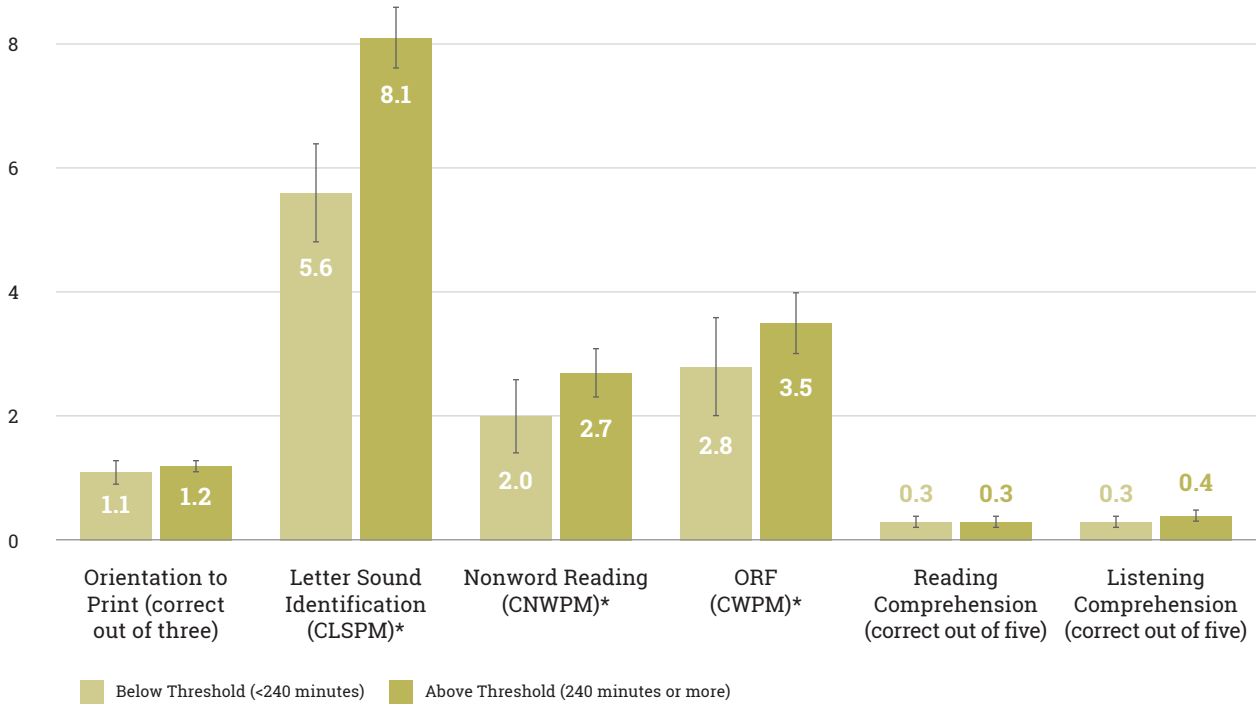
Figure 16: **Average Minutes of Student GG Use by School**



40 An asterisk (*) indicates the average gain score for the intervention group was significantly larger than the average gain score for students in the comparison group at $p<0.05$. *N* sizes: Boys—Intervention Group $n=82$, Comparison Group $n=92$; Girls—Intervention Group $n=140$, Comparison Group $n=119$; All Students—Intervention Group $N=222$, Comparison Group $N=211$.

To better understand how the recommended dosage of GG may have impacted a student's reading gains, students in the intervention group were classified as either "above threshold" (meaning they had played at least 240 minutes of GG during the project) or "below threshold" (meaning they played less than 240 minutes of GG during the project). EGRA gain scores were then compared for students in the two groups (see Figure 17) and the comparison showed that of all the subtasks, student performance on the letter sound identification subtask was most influenced by GG dosage. Students who used GG for at least 240 minutes throughout the project were able to identify an average of 8.1 more letter sounds correctly at endline than at baseline. In comparison, students who used GG for less than the recommended 240 minutes were able to identify an average of 5.6 more letter sounds at endline over baseline. **The difference in gain scores between above threshold and below threshold students on the letter sound identification subtask was statistically significant.** However, GG dosage did not significantly influence student performance on the other EGRA subtasks.

Figure 17: **Average Gain Scores from Baseline to Endline by GG Dosage Threshold**[41]



To understand what may have driven the variation in dosage by school, composite scores were calculated by GG dosage threshold (see Annex Table E.10). Results indicate that students who used GG for 240 minutes or more also reported significantly more teacher reading support and engagement in the program. **This suggests that teacher and student motivation were strongly related to a student receiving enough GG dosage during the project.**

41 An asterisk (*) indicates the average gain score for the intervention group were significantly different at $p<0.05$. *N* sizes: Below threshold $n=71$; above threshold $n=158$.

Key Factors for Success

To better understand the factors that may have influenced changes in EGRA scores from baseline to endline, items from the student questionnaire were combined into nine composites, or groups of items related to each other. This also helped answer the original research question: How did the project influence certain subsets of the student population more than others based on identifiable contextual factors? Each composite consisted of a series of items related to a specific theme that may have affected students' early grade reading skill acquisition, and each composite was assigned a maximum score equal to the total number of items in the composite.

Table 6 provides the mean composite score for the intervention and comparison groups (full composite descriptions and frequencies by item are listed in Annex D). The items in the engagement in program composite were not administered to the comparison group students, nor were technology use items specifically related to GG. The differences in the average composite scores between the two groups were not statistically different, meaning that the average composite scores for intervention and comparison students were comparable.

Table 6: **Descriptive Statistics for Composite Scores by Group**

Composite Category	Intervention (*N* = 232)		Comparison (*N* = 219)	
	Mean	SD	Mean	SD
Language consistency (out of 6.0)	5.44	0.84	5.47	0.68
Socio-economic status (out of 10.0)	4.47	1.73	4.24	1.65
Parental literacy (out of 3.0)	2.12	0.71	2.18	0.64
Parental reading support (out of 3.0)	2.21	0.75	2.09	0.78
Reading materials access (out of 3.0)	2.35	0.81	2.35	0.78
Teacher reading support (out of 6.0)	4.71	1.02	4.75	1.05
Disposition to reading (out of 3.0)	2.82	0.39	2.76	0.46
Technology use – common items (out of 8.0)	4.59	1.14	4.59	1.14
Technology use – all items (out of 11.0)	7.53	1.25	N/A	N/A
Engagement in program (out of 7.0)	6.49	0.75	N/A	N/A

To understand how the results from the nine key composites related to reading abilities—particularly which ones were the most important factors in improvements in reading outcomes—composite scores and EGRA results were analyzed using correlation and regression analyses. Because the intervention sought to primarily affect student abilities in letter sound identification and nonword reading, student gains for these two subtasks were used as the outcomes of interest.

First, the influences of three composite scores were examined on letter sound identification gains: teacher reading support, disposition to reading, and parental reading support composites. Results, presented in Annex Table E.8, showed that the three composites significantly and positively influenced student performance on the letter sound identification subtask, as did inclusion in the intervention group. In other words, students in the intervention group had significantly higher letter sound identification fluency gains, and for those students in the intervention group who had higher composite scores in disposition to reading, teacher reading support, and parental reading support, they also had greater gains on the same subtask.

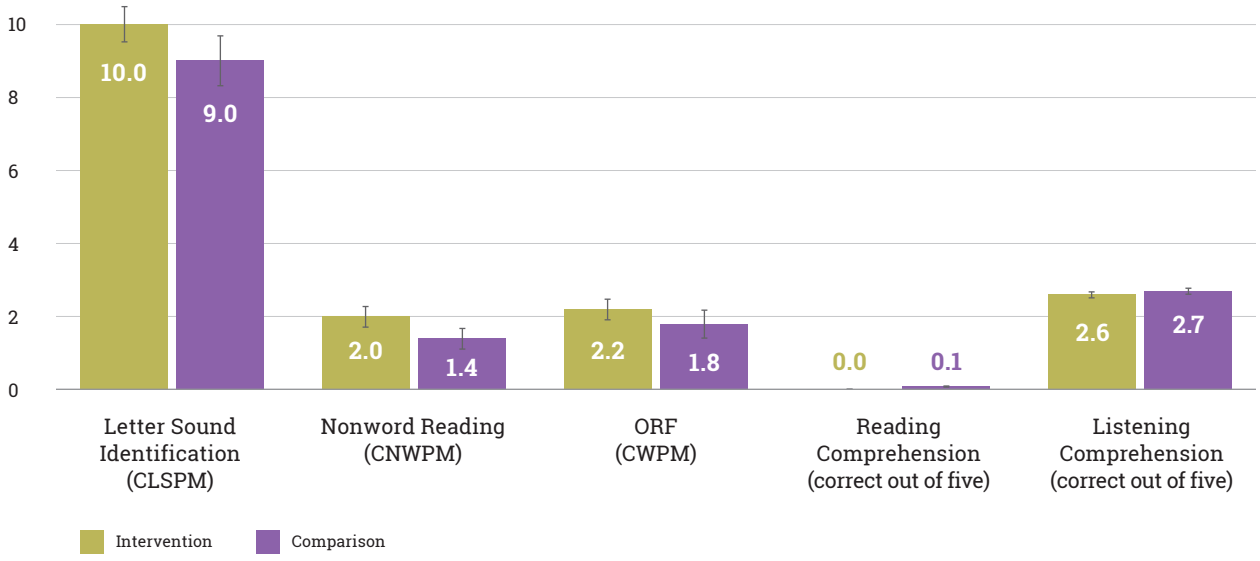
Next, the influence of the disposition to reading and parental reading support composites were examined on nonword reading gains. Results in Annex Table E.9 showed that as with letter sound identification, being in the intervention group predicted higher nonword reading fluency gains, as did higher scores on disposition to reading and parental support.

For both subtasks, while the composites listed above mattered, being in the intervention group was still the most significant predictor of a higher gain score on both subtasks, followed closely by a student's disposition to reading. In other words, the most important driver of students' gains on letter sound identification and nonword reading was being part of the intervention group. For students who were in the comparison group, there was a positive relationship between letter sound identification gains and teacher reading support, and there was a positive relationship between nonword reading gains and parental reading support. This indicates that even if students are not given exposure to GG or a similar educational game, reading support from teachers and parents still matters.

Midline EGRA Results

An EGRA assessment was administered in June 2016, at the midline point of the GG-TTS project. Because the midline EGRA instrument was not equated to the instrument used at baseline and endline, results are presented separately in Figure 18. **At midline, there were no significant differences between the intervention and comparison groups, and no significant differences across gender** (see Annex Table E.11).

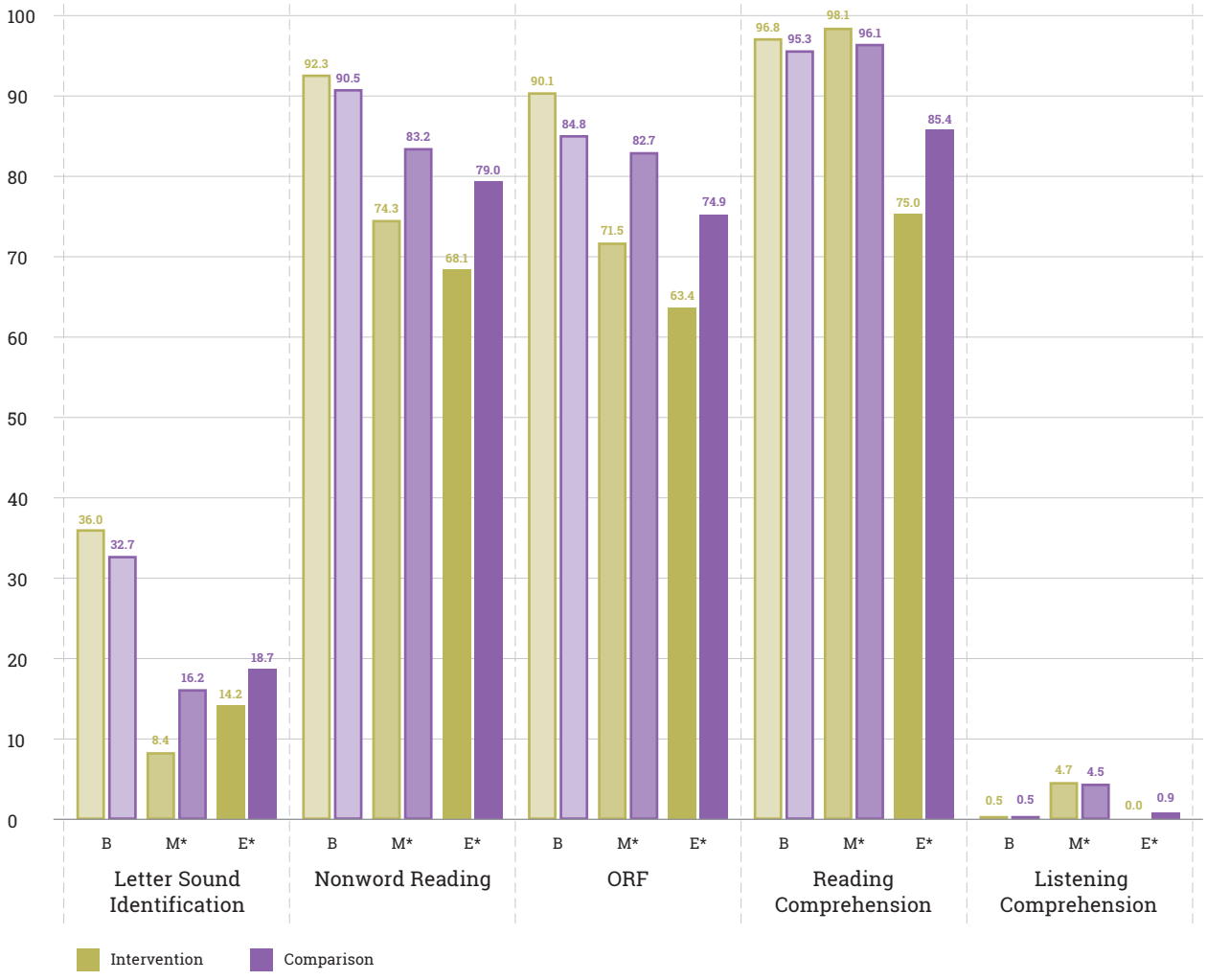
Figure 18: **Midline EGRA Results by Subtask and Group**[42]



42 Intervention *n*=232; Comparison *n*=219

Although results for the midline EGRA are not comparable with the baseline and endline EGRA results, the proportion of zero scores may be compared across the three assessments. Figure 19 shows the proportion of students receiving zero scores by subtask, assessment period, and group (see Annex Table E.4 for results by gender). The percentage of students receiving zero scores was significantly lower in the intervention group at midline and endline for letter sound identification, nonword reading, and ORF. On the reading comprehension subtask, the intervention group had a significantly lower proportion of students who received zero scores at endline.

Figure 19: **Percentage of Students Receiving Zero Scores by EGRA Subtask and Group at Baseline, Midline, and Endline by Group (%)**[43]



43 An asterisk (*) indicates the percentage of students receiving zero scores was significantly different between intervention and comparison groups at $p<0.05$.

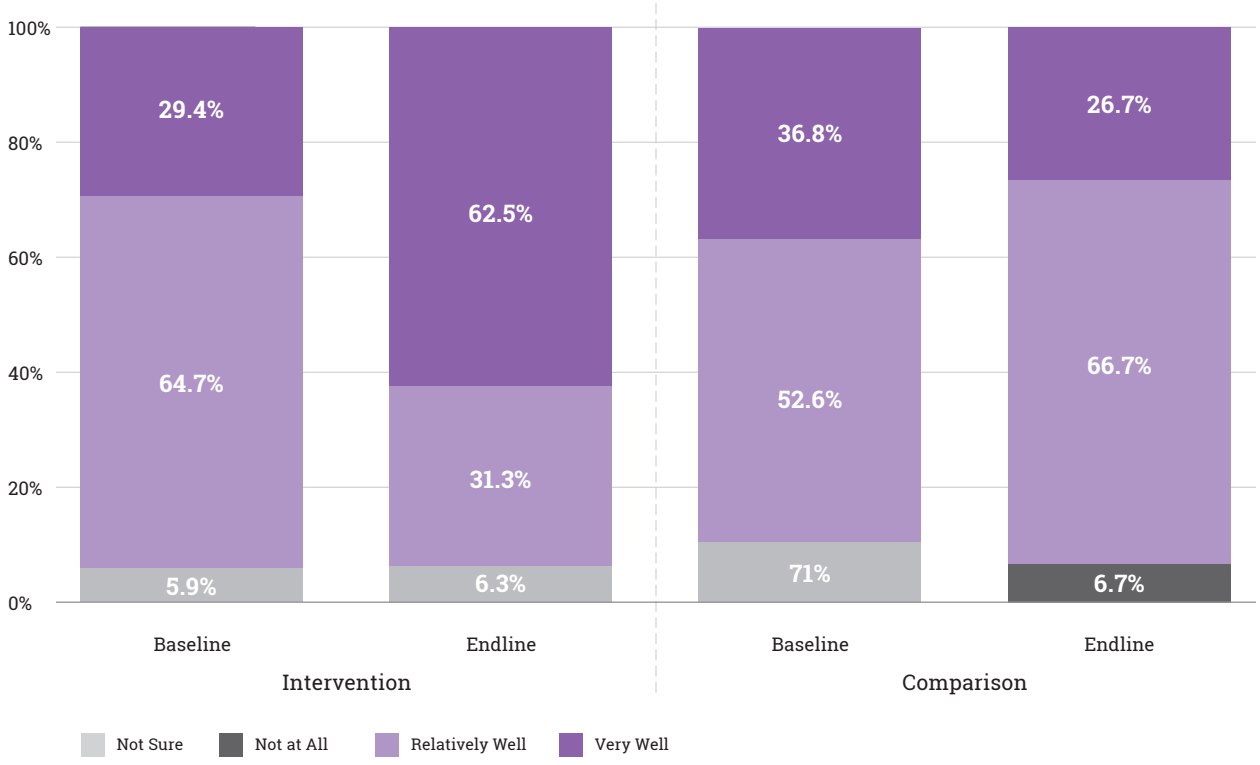
VIII. Teaching Methods Results

The following section presents results related to the following research questions: 1) can the online GG-TTS materials (accessed via smartphones) help rural Zambian teachers to utilize new methods for ciNyanja literacy instruction; and 2) how did the project influence or impact adults' (teachers, parents, community members) knowledge, skills, or attitude regarding their role in helping students read?

To help answer these questions, teachers responded to a general questionnaire and an ICT questionnaire during the baseline operational data collection and during the endline operational data collection. The goal of these questionnaires was to assess teacher behavior change on teaching methods and on ICT use.

Figure 20 presents teacher responses to the following question: how well are current teaching methods addressing student reading problems? Teachers in intervention schools appear to feel that their teaching methods are better addressing student reading problems at endline, as over 62 percent say their methods are addressing problems "very well." Further, this proportion increased from the baseline period, during which only about 29 percent of teachers in intervention schools reported that their teaching methods were "very well" addressing student reading problems. In contrast, teachers at comparison schools reporting "very well" on this question was reduced from about 37 percent at baseline to about 27 percent at endline.

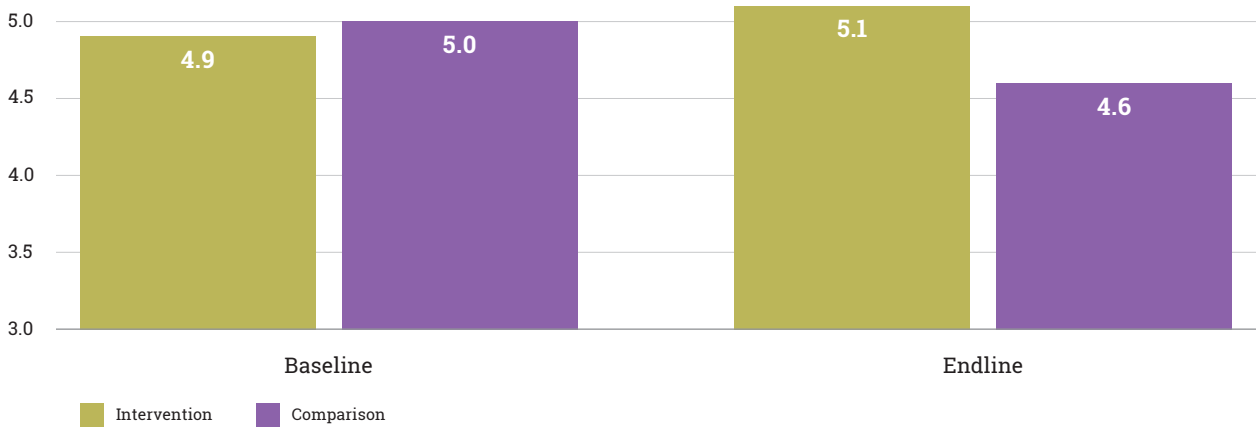
Figure 20: **How Well Are Current Teaching Methods Addressing Student Reading Problems by Group (%)**[44]



44 Because of rounding, totals may not add up to exactly 100 percent.

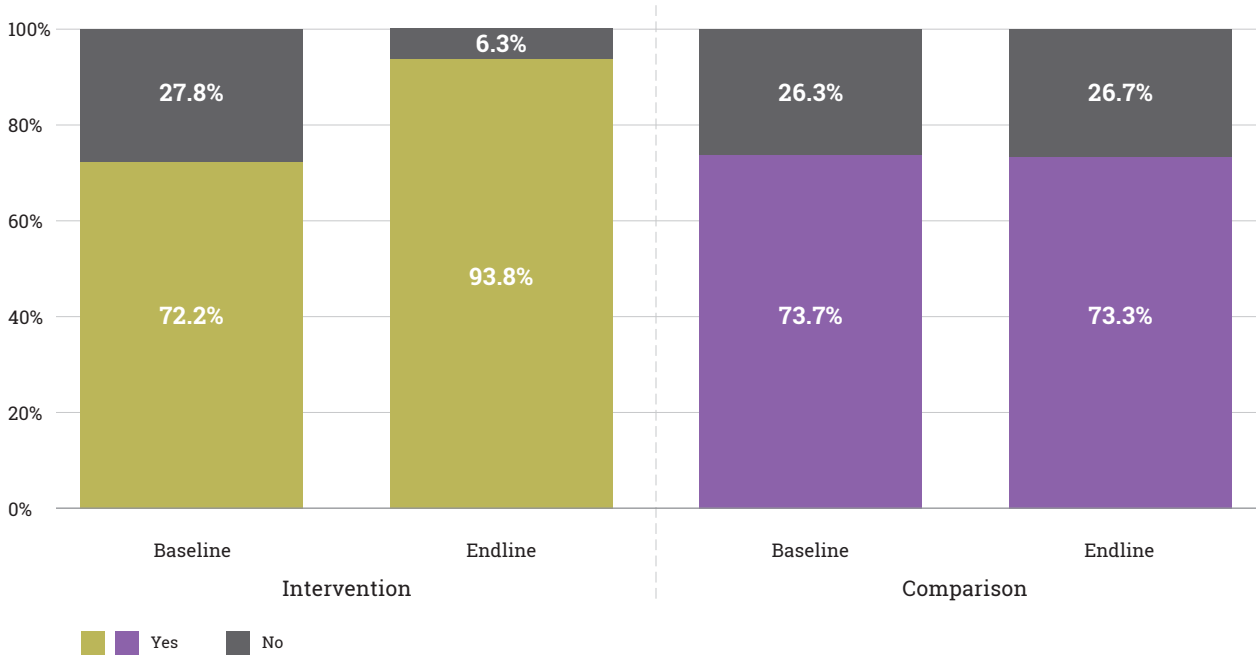
Teachers were also asked about the number of hours dedicated to literacy teaching in their school timetable each week, and results are presented in Figure 21. Teachers who participated in GG-TTS reported an increase in the amount of time they spent teaching literacy to their students from 4.9 hours at baseline to 5.1 hours at endline, while teachers in the comparison group reported spending less time on literacy teaching at endline than at baseline.

Figure 21: **Average Hours Dedicated to Literacy Teaching by Group**



On the ICT questionnaire, teachers were asked whether they had ever used the internet (Figure 22). Whereas the number of teachers who said they had used the internet slightly decreased for teachers in the comparison group from about 74 percent at baseline to 73 percent at endline, the percentage of teachers in the intervention group who reported using the internet was much higher at endline than at baseline, increasing from 72 percent at baseline to over 93 percent at endline.

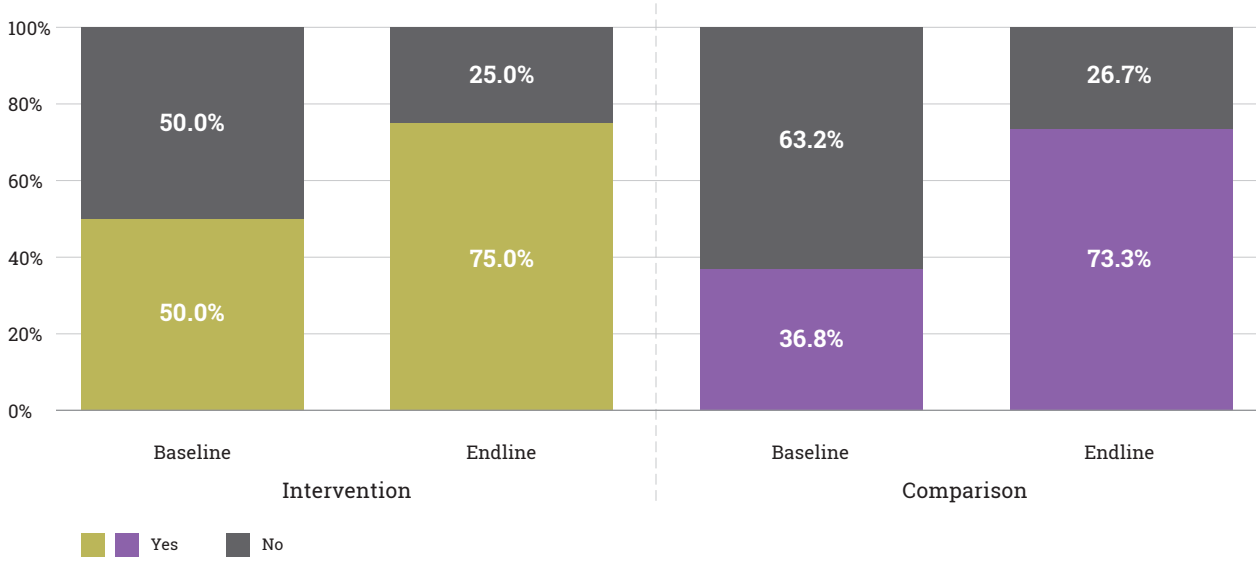
Figure 22: **Use of the Internet by Group (%)**[45]



45 Because of rounding, totals may not add up to exactly 100 percent.

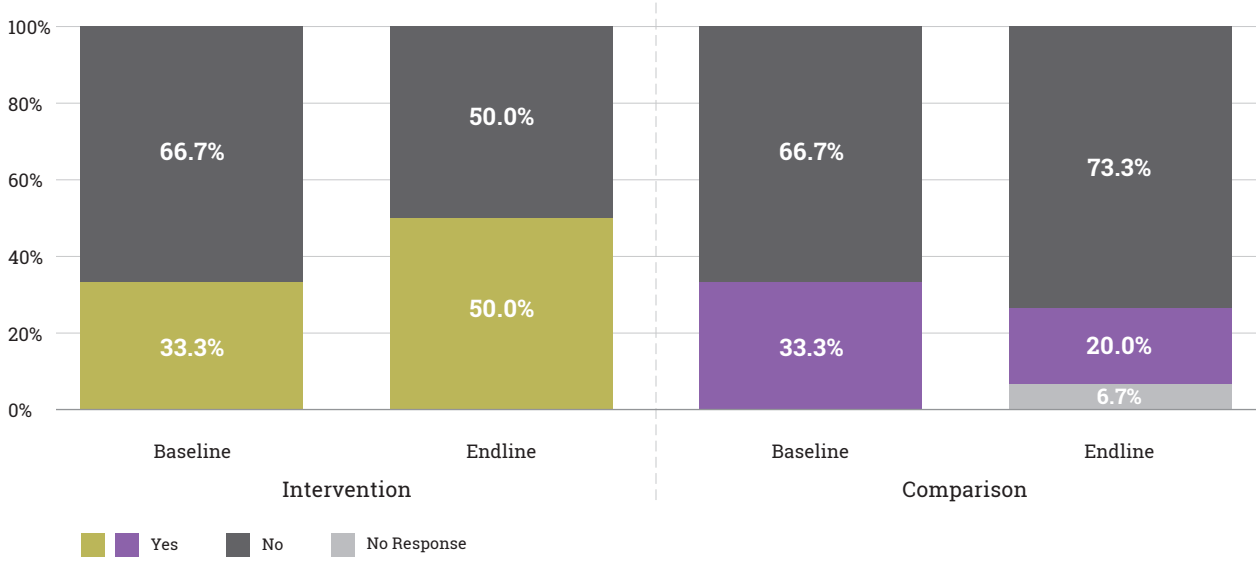
Teachers were also asked if they had ever used ICT for educational purposes, such as for lesson preparation, and results are presented in Figure 23. The proportion of teachers who used ICT for educational purposes increased across time for both intervention and comparison groups: at endline, 75 percent of intervention teachers reported having used ICT for educational purposes, compared to about 73 percent for comparison teachers.

Figure 23: **ICT Use for Educational Purposes (%)**



To determine if there were any school-level changes in regards to ICT use, teachers were asked if their school had arranged for any trainings for teachers on computer use, and results are provided in Figure 24. While fewer comparison teachers reported that their schools had arranged for computer trainings at endline—about 67 percent at baseline compared to 73 percent at endline—the percentage of teachers in the intervention group who said that their school had arranged for computer trainings increased from about 33 percent at baseline to 50 percent at endline. This may indicate that not only did GG-TTS raise awareness about the use of ICT among teachers, but also it raised awareness in school administration about the importance of ICT.

Figure 24: **School Training for Teachers on Computer Use (%)**



Classroom Observations

In addition to teacher questionnaires, GG-TTS staff observed teachers during a regular lesson with the intent of understanding if teachers were implementing learned teaching practices in the classroom after the rollout of the teacher training website (see Annex C for the full observation checklist). They made note of teacher classroom practices related to the content on the teacher training website. In total, 12 out of 17 teachers were observed from 11 out of 15 schools. It should be noted that because classroom observations were taken only once with intervention teachers and after the teacher training website was rolled out, it is not possible to determine if practices changed as a result of the website. The data presented is indicative of practices at one point in time and should not be interpreted as causal.

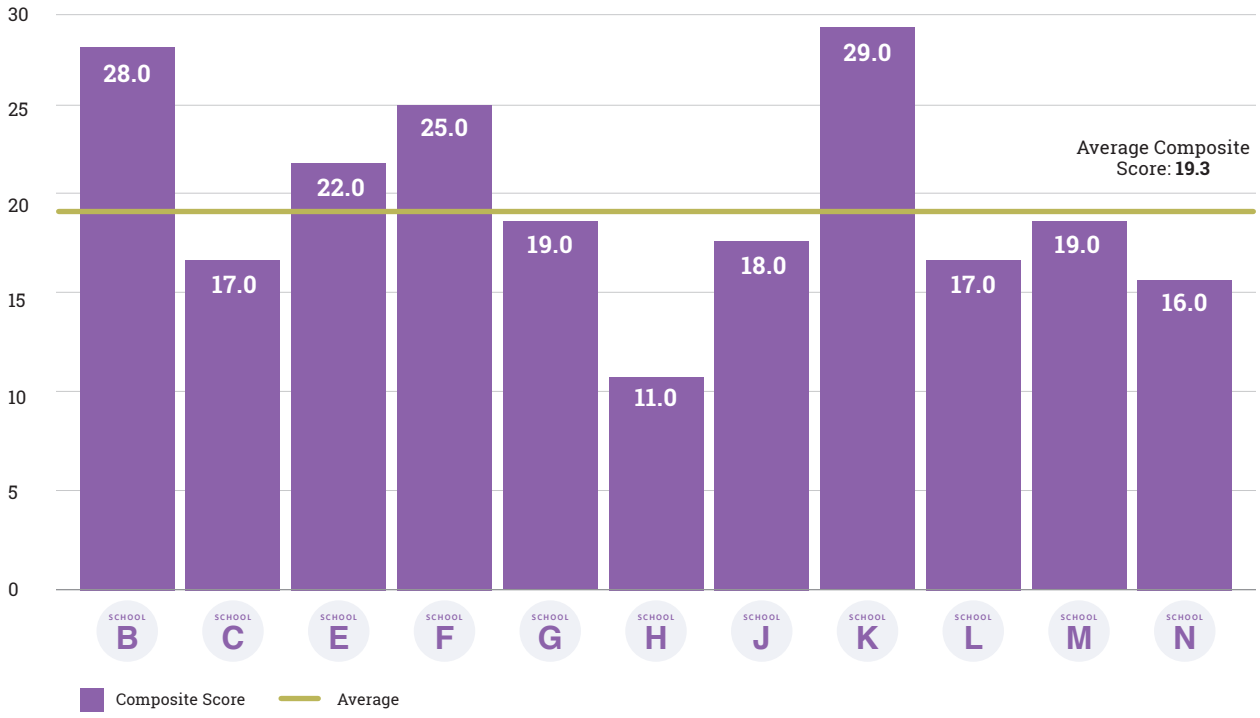
Results from the five composite categories are presented in Table 7. Overall, teaching practices varied across classrooms, and only on the designing your own literacy game composite was the maximum observational score equal to the maximum score possible. This indicates that although teachers may have been exposed to content through the website, it is unclear if teachers effectively incorporated what they learned into their classroom practices.

Table 7: **Classroom Observation Composites and Scores**

Composite Category	Average Score	Min. Score	Max. Score
Teaching literacy in mother tongue language – ciNyanja (out of 12.0)	7.3	5.0	10.0
Supporting struggling readers (out of 10.0)	6.0	2.0	9.0
Story reading and telling (out of 13.0)	4.9	1.0	8.0
Designing your own literacy games (out of 3.0)	0.8	0.0	3.0
Singing as a literacy tool (out of 3.0)	0.3	0.0	1.0
Total	**19.3**	**10.0**	**29.0**

Figure 25 shows the total composite score on the classroom observation form received by school. There was wide variation in how well schools incorporated website content into their classroom practices, and there also does not appear to be a consistent trend between the average level of GG dosage by school (Figure 16) and the classroom practices by school. This is further corroborated by results in Annex Table H.1, which show that there is not a conclusive relationship on EGRA gains between teachers who had above average classroom observation composites—in other words, teachers who incorporated website content effectively into their classrooms—and those who had below average classroom observation composites.

Figure 25: **Average Classroom Observation Total Composite Score by School**



IX. Scalability

Stakeholders are increasingly interested in assessing the scalability of interventions. To scale up a project means to expand, replicate, adapt, and sustain a successful project in a new geographic area and to reach more beneficiaries over time.[46] ACR GCD grantees have implemented small-scale pilot projects with technology-based solutions for improving early grade reading skills. An important consideration at the conclusion of each project is the feasibility of replicating or expanding the technology-based innovations and project models to a different or larger population or area.

To inform this decision, STS conducted a scalability assessment guided by the following research question: Is this intervention and/or innovation suitable to be considered for scaling? STS used an indirect approach that relies on qualitative descriptions of project performance around seven parameters of sustainability:

- Credibility
- Observability
- Relevance
- Relative Advantage
- Ease of Transfer and Adoption
- Testability
- Sustainability of Funding

46 Cooley, L., & Linn, J. F. (2014). *Taking Innovations to Scale: Methods, Applications and Lessons*. Results for Development Institute. Washington, D.C. Retrieved from: https://www.usaid.gov/sites/default/files/documents/1865/v5web_R4D_MSI-BrookingsSynthPaper0914-3.pdf

The seven parameters were adapted from the USAID-funded Scalability Assessment Tool developed by Management Systems International.[47] The tool includes seven parameters and 28 questions. STS used data from EOP interviews, EGRA assessment results, literature reviews, and project monitoring and evaluation to assess scalability parameters. These results are meant to inform local program staff, stakeholders, and donors of key considerations to be taken into account before scaling the GG-TTS project model and technology to a larger beneficiary population.

Credibility

An intervention or innovation must be credible to be taken to scale through either replication or expansion. This aspect of scalability assesses if various stakeholders—including potential adopters, funders, implementers, and beneficiaries—believe that the model has a strong evidence base that may include existing empirical research or anecdotal information.

Key Considerations:

1. What evidence was used to develop the intervention?
2. What evaluations have been conducted on the intervention?
3. In what social contexts does the intervention work?
4. What individuals and institutions support the intervention?

GG was developed in Finland as a technology-based intervention to help struggling readers,[48] and it has since been adapted to support students in over 20 countries.[49] In addition to Zambia, where it was introduced in 2005, GG has also been utilized in Kenya and Tanzania. GG was readily adapted to African languages with transparent orthographies[50] and had been used with both teachers and students alike to improve knowledge of letter sounds. A number of evaluations have been conducted on GG use in different social contexts in Zambia, including: exploring which delivery method is most effective for improving student outcomes,[51] the use of GG in rural homes to engage parents and students in letter sound knowledge,[52] and impact on teachers' beliefs and knowledge and literacy practices for Grade 1 learners.[53]

The GG-TTS project was developed on a strong evidence base, particularly in regards to the GG component of the project. In particular, the project expanded upon previous study designs to further test the impacts of GG when applied with different implementation parameters. Although numerous, previous studies of GG in Zambia had small sample sizes, short implementation periods, and lacked comparison groups. The research design for GG-TTS sought to fill those research gaps by exploring the impacts of GG when implemented for a longer period of time and measuring durability of reading gains after students completed GG. Additionally, the GG-TTS research design included a larger sample size and a comparison group, both of which allow for better isolation of the impacts of GG.

47 Cooley, L., & Linn, J. F. (2014). *Taking Innovations to Scale: Methods, Applications and Lessons*. Results for Development Institute. Washington, D.C. Retrieved from: https://www.usaid.gov/sites/default/files/documents/1865/v5web_R4D_MSI-BrookingsSynthPaper0914-3.pdf

48 Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., ... Lyytinen, H. (2015). GraphoGame – a catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, 6, 671. doi:10.3389/fpsyg.2015.00671

49 http://info.graphogame.com/

50 Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., ... Lyytinen, H. (2015). GraphoGame – a catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, 6, 671. doi:10.3389/fpsyg.2015.00671

51 Ibid.

52 Ibid.

53 Jere-Folotiya, J. (2014). Influence of grade one Zambian teachers and GraphoGame on initial literacy acquisition: Lusaka district. *Jyväskylä Studies in Education, Psychology and Social Research*, 404. https://jyx.jyu.fi/dspace/handle/123456789/44114

Another departure from existing studies of GG in Zambia was the use of EGRA as a tool for measuring change in reading skills. Most previous GG studies used the internal GG assessments to measure gains in student early grade reading skills rather than an external assessment, such as EGRA. As a result, there was limited- to no-evidence base upon which to understand how playing GG could impact a student's performance on letter sound or word or passage reading outside of the game.

The teacher training website was a new component for the GG-TTS team, and there is limited evidence on the impact access to ICT has on supporting teacher learning. Prior to the GG-TTS project, there were no published reports on using ICT for teacher training or learning. The use of ICT for teacher learning is a nascent area of research in Zambia; there is little conclusive evidence on what works when considering content, hardware, duration of exposure, and social-context considerations.

According to EOP interviews conducted by STS, MoGE representatives expressed support for the GG-TTS project and curiosity about the potential of positive outcomes. MoGE is eager to learn new ways to support improvements to teacher classroom practices; they are specifically interested in the use of mobile phones for capacity building. MoGE representatives also express the desire to see more individualized reading support for students, but there are currently no materials or resources available to teachers to encourage this practice.

Credibility Conclusion

Credibility for the GG component of the intervention is high. The GG intervention was developed using a reliable and strong evidence base within the Zambian contexts, including past use at homes and in schools, and by a variety of different users—teachers, parents, and students. Credibility for the teacher training component is low due to a lack of published evidence or evaluations. However, there is strong administrative support for the combined approach from teachers, MoGE officials, and the project staff.

Observability

For an intervention or innovation to be scaled, it should have observable results that show efficacy or impact. Observability of results is key to providing non-technical audiences proof that an intervention or innovation achieves its intended outcomes and thus will have positive impacts on beneficiaries.

Key Considerations:

1. Are the results visual and observable?
2. What is the relationship between (any) results and the intervention?
3. Is there any emotional appeal associated with the evidence?

EGRA results from the intervention and comparison groups provide observable results of the GG-TTS project. Overall, students who participated in the project experienced larger gains in early grade reading skills than their peers in the comparison group—namely, on the orientation to print, letter sound identification, nonword reading, ORF, and reading comprehension subtasks. Because GG targets and reinforces early grade reading skills in letter sound identification and nonword reading, gains on these two subtasks by intervention group students are particularly notable and suggest that there may be a relationship between use of GG and early grade reading skill acquisition.

Dosage data points to uneven implementation across schools and among students. Although this variability in implementation highlights some of the challenges for scalability, it was possible to analyze the different results associated with different levels of dosage. Indeed, students who played GG for at least the recommended 240 minutes throughout the implementation period had significantly higher gains on the letter sound identification subtask than their peers who played less than the recommended number of minutes. This provides encouraging evidence that when implemented with fidelity, the GG-TTS project has high potential for providing better reading gains. It should be noted that gains on other EGRA subtasks—including nonword reading—were not significantly different for students who played at least the minimum amount of GG and those who played less. Because of these inconclusive findings on the relationship between reading gains and the intervention, further research may be necessary to definitively establish a relationship between early reading gains and GG playing.

Additionally, there is insufficient evidence to show that teachers applied information from the teacher training website in their classroom teaching practices. Monitoring data and classroom observations indicated that behavior change in the classroom was inconsistent at best—although it should be noted that classroom observations were only conducted once and do not provide information on change over time. Similarly, the relationship between the adoption of teacher training content into classroom practices and EGRA gains was inconclusive. Given the implementation challenges faced by the GG-TTS project in rolling-out the teacher training website, it is unclear if there is a connection between the implementation of this component of the project and the results.

In EOP interviews, teachers and head teachers generally expressed positive feedback when asked about the results of the project. Most said that they believed GG helped their students learn letter sounds and to help them learn to read; although, one teacher stated that the GG knowledge did not transfer into practical reading skills. There did appear to be an emotional appeal for teachers seeing their students use technology and student-centered tools. Similarly, head teachers and MoGE district officials shared the general belief that the GG-TTS project has strong potential to support the overstretched school system in meeting the needs of the lowest performing students.

Observability Conclusion

Although results from the GG-TTS project indicate a statistically positive relationship between GG playing and early grade reading skills gains, there is not an observable connection between the teacher training website and project results. Further piloting and research should be conducted to better integrate the teacher training website into the project design and to improve fidelity of implementation on all GG-TTS components; this would allow for better identification of observable early reading gains as a result of the project.

Relevance

In order to be scalable, an intervention must be relevant to the context in which it is being implemented and it should effectively address a problem that is recognizable and considered important by stakeholders.

Key Considerations:

1. How significant is the problem that the intervention is trying to address?
2. Is the intervention addressing a policy priority for potential adopters?
3. Does the intervention address a need felt by the potential beneficiaries?

According to 2007 data from the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ), the vast majority of Zambian students in Grade 6 are not able to achieve beyond the first three levels of reading. According to these findings, 15.8 percent of Grade 6 students are at a pre-reading level, 28.3 percent are at an emergent reading level, and 28.6 percent are at a basic reading level.[54] This means that just over a quarter of Grade 6 students show any level of advanced reading skills.[55] Further, findings from a 2014 Grade 2 National Assessment Survey conducted in Zambia show that the majority of students were unable to read with either fluency or comprehension. In fact, only about 2 percent of the Grade 2 students read with comprehension in their local language.[56] These data indicate that Zambia faces a significant problem in early grade reading.

There is strong evidence that government stakeholders and the literacy community recognize the severity of the problem. A number of local initiatives and policies aimed at improving the literacy levels of primary school students have already been executed in Zambia. The most recent, the 2013 National Literacy Framework (NLF), emphasizes reading acquisition through phonics. Published by the Zambian Curriculum Development Centre, NLF was a departure from the reading methodology previously promoted in Zambia under the Primary Reading Program.[57] As a result of that study's success, there is a strong need for innovations or interventions that support primary teachers' ability to effectively teach phonics in accordance with the NLF strategy. Potential beneficiaries—particularly primary school teachers—need professional development on phonics instruction. Teachers expressed a desire for training in how to implement NLF strategies and cited their own struggles in supporting lowest performing students in overcrowded classrooms.[58]

Because the GG-TTS project included a teacher training component and a student-centered technology component, both of which focused on phonics instruction, it is a highly relevant intervention at both the policy and classroom levels. It should be noted that GG is currently only available in ciNyanja—the second most commonly spoken regional language. The GG-TTS project would have more relevance if it were available in other regional languages—such as Bemba, the most commonly spoken regional language, or Tonga, the third most commonly spoken regional language.

Relevance Conclusion

GG-TTS is addressing a persistent problem in the Zambian context. The project is highly relevant and has the potential to effectively support NLF and overall goal of improving pre-literacy skills of teachers and students in the Zambian education system. The project provides students with self-paced phonics support, and it provides teachers an online platform for professional development on a range of literacy topics as well as the ability to connect with other teachers. There is strong support and enthusiasm for both components from stakeholders based on its relevance to current Zambian education policy.

54 http://www.sacmeq.org/?q=sacmeq-members/zambia/reading-and-math-achievement-levels

55 Reading for meaning: 14.9%; interpretive reading: 6.0%; inferential reading: 3.7%; analytical reading: 2.2%; critical reading: 0.5%.

56 RTI International. (2016). Zambia Benchmarking Report. https://globalreadingnetwork.net/eddata/zambia-benchmarking-report

57 UNICEF. (2016). The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa, Zambia. https://www.unicef.org/esaro/UNICEF(2016)LanguageandLearning-Zambia.pdf.

58 UNICEF. (2016). The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa, Zambia. https://www.unicef.org/esaro/UNICEF(2016)LanguageandLearning-Zambia.pdf.

Relative Advantage

A project's relative advantage relates to whether the intervention offers an improvement over current and alternative solutions to the problem.

Key Considerations:

1. How adequate are the current solutions for the problem?
2. Is this intervention more effective than the current solution?
3. Is this intervention more effective than other innovative models established?

Current approaches to address the low level of early grade literacy in Zambia are primarily focused on in-person teacher training and professional development, as well as the creation of teaching and learning materials for in-class use. To support struggling red-level students, teachers are advised to provide remedial classes after regular classroom hours. These approaches are generally costly, and there is no rigorous evidence showing that these approaches are effective in supporting early grade reading gains. During EOP interviews, MoGE representatives expressed a desire to continue searching for innovative and effective solutions for the problem, especially in recognition of the inadequacy of the current solution.

The GG-TTS project components have the potential to be more effective solutions than the existing approaches in Zambia. ICT-based training, available to a wide range of teachers through ICT's accessibility, may reach more teachers than traditional training models for a lower cost. It also saves time and may prove less disruptive, as teachers do not need to take time from their classrooms to travel to attend in-person training sessions. Although a comprehensive cost comparison between in-person and ICT-based teacher training is not currently available, in the long term, internet-based trainings may prove more cost effective than in-person training. Evidence is currently insufficient to conclusively determine if GG-TTS's teacher training website is more effective than current teacher professional development approaches in Zambia; still, the MoGE appears interested in exploring how it or other similar ICT solutions can impact teachers' classroom practices.

GG provides a significant advantage over current solutions for supporting struggling readers in the classroom. In EOP interviews, teachers expressed concerns over class sizes and their ability to effectively implement the currently recommended solutions to support red-level students. Under the current approach, some teachers are expected to provide remedial lessons to over 50 percent of a classroom. With classroom sizes of upwards of 65 students, the current approach makes it nearly impossible for teachers to provide in-depth, individualized support to struggling readers during remedial lessons. Because GG is student-centered and allows players to advance at their own pace, this approach has strong potential both to encourage student learning and alleviate pressures on teachers to be the sole provider of remedial support.

Although many ICT-based learning innovations are in-development or being implemented in other countries throughout Africa, there are not any known models within Zambia currently being used to address early grade reading deficiencies in public primary schools. In EOP interviews, head teachers and stakeholders mentioned anecdotal cases of technologies being used by other schools or in pilot projects sponsored by the government or by donors; nevertheless, these solutions are not yet widespread nor is there sufficient evidence available on their efficacy. The greatest challenge of incorporating ICT-based literacy solutions in Zambia seems to be the lack of technology penetration generally in schools and in homes. This challenge may be a reason that anecdotal innovations have yet to be widely adopted within the Zambian school system.

Relative Advantage Conclusion

GG-TTS offers an innovative contribution to addressing low-levels of early grade reading found within the Zambian education system. The project provides a targeted approach for improving specific early reading skills: letter sound awareness; syllable identification; and decoding. Its approach is unique and provides advantages over current classroom approaches to teaching these skills in Zambia. However, this intervention does not address whole reading, reading fluency, or reading comprehension. It is best seen as one component within a more comprehensive literacy intervention. To fully understand the relative advantage of the GG-TTS model, development and implementation challenges should first be solved. With a more stable model, GG-TTS has the potential to provide relative advantage over current solutions.

Ease of Transfer and Adoption

Ease of transfer and adoption relates to whether the characteristics and components of the intervention lend themselves to being adopted by organizations other than the original implementer. This parameter of scalability looks at how complex or resource-heavy an intervention is, as well as if specific elements may be deemed inappropriate or unattractive to other implementers.

Key Considerations: [59]

1. How technically sophisticated are the intervention's components and activities?
2. How complex is the intervention?
3. What level of supervision and monitoring is needed?

The components of the GG-TTS project do not require high levels of technical literacy among their primary users. In EOP interviews and through monitoring surveys, most teachers and students reported ease in learning to use the smartphones and GG. Those teachers and students who expressed concerns over their ability to use the technologies were generally able to rely on support from their peers who were more technologically savvy. However, the development and troubleshooting of the technical components did require a more substantial level of knowledge. This included dedicated staff at the Agora Center and CAPOLSA as well as developers from BongoHive and, in certain cases, teachers who worked with GG-TTS project management staff to solve technical issues. It is unclear if, after further improvement of GG and the teacher training website, significant personnel would be required to ensure effective functioning of the technical components of the project.

GG-TTS activities were neither highly technical nor highly complex. Activities were, for the most part, self-guided by the teachers or students and did not require significant time investments. GG provides an easy and fun experience for students; the teacher training website was reportedly very easy to use, aside from some programming issues. The most significant challenge faced by teachers seemed to be finding time before, during, or after class for their students to play GG. Although this was not a complex challenge, a teacher's ability to find time significantly impacted students' ability to benefit from the project.

59 In the original Management Systems International tool, this section includes 11 questions. This analysis includes the questions deemed most relevant for the intervention model and context.

Further, while the technical components of GG-TTS were relatively easy to master and initial training sessions were short in duration, the project required intense supervision and monitoring from project staff—particularly to ensure dosage requirements were being met. The GG-TTS team regularly monitored dosage logs and followed-up with teachers who appeared to be lagging behind. In addition to weekly check-ins, the project team conducted periodic in-person visits to monitor and assess progress, support regular use of GG, and provide any needed technical assistance.

Ease of Transfer and Adoption Conclusion

The intervention requires low levels of technical expertise by both teachers and students, and, if existing technical challenges are solved, both GG and the teacher training website could easily be transferred and adopted by schools. The project did require significant investments in monitoring, technical trouble-shooting, and oversight. It is possible that further piloting and better incentives for teachers to effectively implement the required dosage would allow for a more easily transferrable model for more schools.

Testability

The testability parameter examines how easy it is for organizations to pilot the intervention on a small scale prior to full adoption. Testability assesses whether potential adopters would need to commit significant resources or time to test the model if they chose to pilot it in a new context.

Key Considerations:

1. Is the model able to be tested on a limited scale?

The GG-TTS project is easily testable on a limited scale, as no significant investments in training are required. The GG application in ciNyanja is stable and well-established; the teacher training website is already developed and only requires further refinement and improvement of content prior to administration to a different or larger cohort of teachers. If the project were to be tested in a new context, the greatest challenge would be in adopting GG-TTS to a different regional language in Zambia. As mentioned, ciNyanja is the second most widely spoken regional language in Zambia, and there is potential to expand to regions of the country in which Bemba or Tonga are the languages of instruction in early grades. It is currently unclear what level of time or financial investment would be required to develop GG and update the teacher training website for a new language. It is unclear the length of time and investment needed to adapt GG to any of the additional Zambian languages used in the classroom, and further, any research relating early reading gains to GG-TTS would need to be replicated for any new languages. Because of these considerations, piloting outside of areas in Zambia in which ciNyanja is the regional language would be costly and time intensive.

Testability Conclusion

The current GG-TTS model is easily replicable in parts of Zambia in which ciNyanja is the language of instruction. However, the project would be difficult to pilot in areas of the country that use other languages without a significant time and financial investments.

Sustainability of Funding

Sustainability of funding refers to how cost-effective the intervention is and whether there are funds, either through government or other organizations, available to scale the intervention.

Key Considerations:

1. Is the model more cost-effective than other solutions?
2. What kind of funding commitment is required to scale the model?
3. Is there any potential for internal revenue from the model (i.e. service fees)?

No comprehensive cost-effectiveness analysis was conducted on GG-TTS; instead, a cost analysis was conducted. Cost analysis is often a component of scalability assessments, as it helps decision makers and stakeholders understand the feasibility of replication given budgetary constraints. Because ACR GCD grantees implemented new approaches, they often allotted significant financial resources to develop new materials that could be used on a recurring basis. To better understand the funding requirements of the GG-TTS project, a cost analysis was conducted to present the total cost of the intervention and to understand the investments that would be needed for project replication or scale-up.

USAID guidance on conducting cost analyses on early grade reading projects suggests that the "ingredients method"[60] be used to calculate costs in the following categories:

- Management and associated technical costs
- Development costs
- Implementation costs

Project staff completed a costing template with guidance from World Vision and STS. Costs were outlined based on the activities from the project work plan, and each cost expenditure was classified based on the three categories above. Invoiced costs were used for analysis from the beginning of the project in fiscal year 2015 and projected costs were used for the period from January to March 2017, which covered dissemination events and project close-out.[61, 62]

Despite attempts to fully match costs invoiced to those in the costing analysis, there were approximately $34,145 included in the costing analysis that had not yet been invoiced. This is most likely due to the inclusion of approximated costs for project close-out in the costing analysis, but it may also be due to costs that were duplicated in the analysis.

60 RTI International (2015). *Measurement and Research Support to Education Strategy Goal 1: Early Grade Reading Costing Template and Guidance.* United States Agency for International Development (USAID). Washington, D.C. Accessed via: http://www.youblisher.com/p/1362487-Early-Grade-Reading-Costing-Template-and-Guidance/

61 An additional $10,000 had been projected for FY2017 Q2 expenses. To categorize costs for FY 2017 Q2, the projected expenses were categorized according to FY 2017 Q1 expenses, and the cost values based on projected activities.

62 Matching funds from Agora Center, in the amount of $27,247.15, were included in this analysis.

Table 8: **Cost Analysis**

Activity	Management	Development	Implementation
Activity 1.1 Start Up Meetings		$ -	$ 21,316.90
Activity 1.2 Materials Preparation		$ 5,103.85	$ -
Activity 1.3 Training		$ -	$ 37,974.25
Activity 2.1 Baseline EGRA		$ -	$ 15,860.70
Activity 2.2 Midline EGRA		$ -	$ 20,811.72
Activity 2.3 Endline EGRA		$ -	$ 32,237.87
Activity 3.1 Monitoring CG		$ -	$ 37,664.97
Activity 3.2 Monitoring the GG-TTS Website/supporting teachers		$ -	$ 4,624.95
Activity 3.3 Dissemination & Benchmarking		$ -	$ 2,355.00
Activity 4.1 International Literacy Day		$ -	$ 998.98
Total	**$ 66,187.12**	**$ 5,103.85**	**$ 173,845.33**
Proportion	**0.27**	**0.02**	**0.71**

The management category includes costs that are not directly related to implementation and are likely to vary widely based on who is overseeing the implementation of the intervention. Management costs for the project represent about 27 percent of the costs expended and include the management costs from Agora Center, CAPOLSA, and other indirect rates and fees.

Development includes the costs related to the development of the website, teacher materials, and material preparation and finalizations that would not need to be redeveloped in the scale-up of a project. The development costs represent the smallest proportion of expenditures at about two percent. These costs would not incur again in case of scale-up.

The implementation cost category is arguably the most relevant for stakeholders who are considering scaling up a project or intervention. This category includes all the recurrent activities and costs that would need to be expensed should the project be replicated, including material production and distribution, training, monitoring and evaluation, events and presentations, workshops, and human resources activities. Implementation costs represented about 71 percent of the total project cost, by far the largest proportion of the three categories. Within this cost category, the largest expenses were training and monitoring GG, which accounted for nearly half the implementation costs.

Projects sometimes benefit from in-kind services, institutional support, or preexisting relationships with stakeholders or governments that may provide the project with tangible benefits, although it may be difficult, if not impossible, to monetize the costs. Examples of this include local volunteers, support from a large non-governmental organization, or relationships with local governments that could ease logistics and procedures. Based on feedback from Agora Center and CAPOLSA staff, the major non-monetized costs to consider were: donated phone credit from Airtel for CAPOLSA staff; GG phones recycled from previous projects; and an undervaluing of CAPOLSA management's level of effort spent on the GG-TTS project.

Scalability of Funding Conclusion

Because GG was an established technology, development costs were relatively low for the project. Once the project has stabilized and observable results have been obtained, additional analysis should be conducted to assess the cost-effectiveness of the model.

X. Conclusions

Despite some issues with rollout and implementation, **the GG-TTS project had promising results that could reasonably lead to future projects.** The GG-TTS project built upon significant research on the use of GG in Zambia to improve early grade literacy acquisition. It also incorporated a teacher training website component to help teachers learn and utilize new methods for literacy instruction, specifically providing targeted support for their struggling readers. Students who participated in the GG-TTS project had significantly higher EGRA gains over their peers in the comparison group, except on the listening comprehension subtask. Students in the intervention group also scored significantly higher on the GG letter sound assessment than students in the comparison group. Although gains for the intervention group were higher on most assessments, only gains on the EGRA letter sound identification subtask were large enough to contribute to struggling readers' development of functional pre-literacy skills.

The project faced challenges during the rollout and implementation of the intervention. The two major challenges—and the ones most likely to have had an impact on student reading gains—were the fact that GG did not allow most students to advance beyond the letter sound module and the fact that the roll-out of the teacher training website was delayed. Although the GG-TTS project management team provided extensive monitoring and technical assistance to the schools and teachers, implementation of the project was variable: while four schools did not provide the recommended dosage to students, other schools were able to deliver at least double the recommended dosage.

Based on the scalability and assessment results, there is potential for the GG-TTS project to be replicated. The following are lessons that should be taken into account for any future interventions incorporating GG-TTS components.

Lessons Learned

GG-TTS, in its current state, is producing promising reading gains, but it is not contributing substantially to higher-order reading skills development.

The most notable gains were on letter sound identification, which was an expected outcome given the focus of the GG-TTS project. Overall, students in the intervention group had statistically significant gains across five of six subtasks on the EGRA: orientation to print, letter sound identification, nonword reading, ORF, and reading comprehension. On average, students in the intervention group were able to read an additional 7.4 CLSPM, 2.5 CNWPM, and 3.3 CWPM at endline—compared with 4.8 CLSPM, 1.3 CNWPM, and 2.0 CWPM in the comparison group. On the GG assessments, students in the intervention group also had statistically larger gains from baseline to endline than students in the comparison group. While these are statistically significant gains, they do not represent the magnitude needed to impact students' reading fluency. Also, due to implementation challenges, the gains cannot be delineated and distinctly attributed to either GG or the teacher practices. Further studies should explore the extent to which gains in reading outcomes result from students playing the GG modules versus changes in teachers' instruction as a result of their online learning experiences.

Creating a more stable approach to how GG should be implemented at schools could help ensure teachers deliver the intervention as intended.

Even with extensive monitoring—including several site visits and ongoing check-ins via text and phone calls from the project staff—four schools were not able to implement the minimum dosage of GG time to their struggling readers. The project gave teachers autonomy to determine how to schedule GG time outside the MoGE's classroom timetable; yet, several teachers reported challenges in finding a suitable time or feeling overloaded with other school responsibilities. This created inconsistency for the students and meant that the intervention was not rolled out with fidelity at each school.

For GG to be most effective, the minimum threshold dosage should be further investigated to determine how much playing time is needed to achieve optimal reading gains.

GG dosage varied greatly across schools. In the intervention group, 11 schools met the minimum threshold of 240 minutes of playing time. Four schools were below the threshold with the two lowest schools averaging only 74.0 and 95.2 minutes of playing time per student, respectively. "Above threshold" students identified 8.1 additional letter sounds at endline compared with only 5.6 additional letter sounds at endline for "below threshold" students. The students who received less than the recommended minimum dosage only identified 0.8 additional letter sounds more than their peers in the comparison group who had no access to any additional reading support. GG dosage did not significantly influence students' performance on any of the other EGRA subtasks.

Girls showed greater improvements than boys after participating in the GG-TTS intervention.

The average gains scores were statistically significantly larger for girls in the intervention group; however, it is unclear why this happened. While the functional gains were small, they were consistent

across all subtasks. Any further implementation of this type of intervention should explore if, and why, girls respond differently than boys.

Significant time and resources are needed to rigorously measure change in teacher practices for struggling readers as a result of the teacher training website.

The online teacher training modules were delivered near the end of the intervention, and only one classroom observation was conducted to monitor implementation of the new strategies in teaching. There is not enough evidence from this aspect of the intervention to present conclusive results. Through stakeholder interviews, it is clear that there is excitement regarding an online learning platform for teachers as it could support the professional development of teachers in rural locations. Further implementation and testing could be done to better demonstrate the impact of the website—its content and its structure as a delivery method—on teacher practices for struggling readers.

Technology-based projects like GG-TTS need to have sufficient time allocated for testing and updating any technology prior to the intervention rollout.

Although the ciNyanja version of GG had been used extensively in Zambia prior to the GG-TTS project, it was only through this implementation that the developers discovered programming challenges that impacted students' progression through the game. Since students were not able to play the more advanced modules of GG, it is unclear what impact better progression through the modules could have had on their reading gains. Literacy projects utilizing technology should devote significant resources to developing, testing, and monitoring the technology component to ensure higher quality learning experiences for the students and teachers.

Self-reporting from teachers is helpful, but other assessments are needed to better understand the value-add of ICT based-training for teachers.

Most of the teacher learning measures utilized self-reporting with limited external observations. These returned unclear results: for example, teachers in both the intervention and comparison schools reported using ICTs for educational purposes. Additional investment should be made in developing better ways to measure how ICT can change teacher learning and practices and to what extent this change impacts their students.

Low-tech solutions like GG and online training websites have the potential to help Zambian teachers address additional literacy challenges in their classroom.

This project utilized low-tech solutions to solve a pressing issue facing Zambian teachers. The GG-TTS components have the potential to be delivered either independently or be incorporated into a larger reading initiative sponsored by a donor or MoGE. Stakeholders expressed a desire to pair a GG-type game with another literacy intervention project focused on passage reading and comprehension. There is also interest in exploring ways to incorporate mobile learning into more general literacy teacher training since the model is readily able to utilize low-tech technology already present in Zambia.

Annexes

Annex A: **Baseline and Endline EGRA Instrument**

Enumerator Name	

Start Time

Date	
Time	

GG_ID

School Name	
Student ID	

EGRA ID

EGRA ID	

Verbal Consent

It is important to establish a playful and relaxed rapport with the child. The child should perceive the assessment almost as a game to be enjoyed rather than a severe situation. It is important to read the directions slowly and clearly. After you have finished, thank the child for their time and effort.

Uli bwanji. Dzina lango ndine.....ndipo ndikhala ku.....Ndingakonde kukuuza za moyo wanga. Good morning. My name is ___ and I live in ___. I'd like to tell you a little bit about myself.

[Number and ages of children; favourite sport, radio or television program, etc.]

1. **Kodi umakonda kucita ciani ngati siuli mu sukulu?** What do you like to do when you are not in school?

   [Wait for responses; if pupil is reluctant, ask question 2, but if they seem comfortable continue to verbal consent].

2. **Kodi ndi masewera otani ameme umakonda kusewera?** What games do you like to play?

READ THE FOLLOWING WORD-FOR-WORD

Ndifuna kukuuza cifukwa cake ndabwera kuno lero. Ndigwira nchito pa sukulu yama phunziro apamwamba ya mu Zambia (Universityr of Zambia). Ndipo tikufuna kumvetsetsa mmene ana amaphunzirira kuwerenga. Iwe wasankhidwa mwamwar. Let me tell you why I am here today. I work with the University of Zambia and we are trying to understand how children learn to read. You were picked by chance.

Ndifuna thandizo lado pa nkhaniyi. Koma suyenera kutengako mbali ngati sufuna. We would like your help in this. But you do not have to take part if you do not want to.

Ife tizachita sewero la kuwerenga. Ine ndizakufunsa kuwerenga malembo, mau ndi ka nthano kakafupi mokweza mau. Ndizakufunsanso kuzindikira ndi kuyankha mafunso ocepa. We are going to play a reading game. I am going to ask you to read letters, words, and a short story out loud.

Mwakugwiritsa nchito lamya iyi, ndizaiemba mayakho ako. I am going to use this "phone" to record your answers.

Zimene tizachita pano si mayeso ndipo sizidzakhudza maphunzilo ako pasukulu lino. This is NOT a test, and it will not affect your grade at school.

Nsizakufunsanso mafunso ena monga kumene umayeselera kuwerenga ndiponso ngati ukonda kuwerenga. I will ask you other questions about where you practice reading and whether you like it.

Kaciwirinso, sungatengeko mbali ngati sufuna kutero. Tikayamba kufunsa mafunso, ngati siufuna kuyankha funso ungakhale cete, zilibwino cabe. Once again, you do not have to participate if you do not wish to. Once we begin, if you would rather not answer a question, that's all right.

Kodi uli ndi mafunso alionse? Do you have any questions?

Kodi uvomela kutengako mbali musewero iyi? Would you like to participate?

Kodi wakonzeka kuti tiyambe? Are you ready to get started?

(If verbal consent is not given, thank the child and move on the next child.)

Check box if verbal consent is given. ☐

Demographics

Name

Age

☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ Other

Grade

☐ G2 ☐ Other

Sex

☐ Male ☐ Female

Orientation to Print

Show the child a story passage in the pupil stimuli packet. Read the instructions in the gray boxes below. Provide the child ten seconds to respond. Record the child's response before moving to the next instruction. If the child doesn't respond in the ten seconds, mark as no response and move on.

Sindifuna kuti uwerenge tsopano. Pa pepala iri, ungayambire kuti kuwerenga? Ndionetse ndi cala cako.
[I don't want you to read this now. On this page, where would you begin to read? Show me with your finger.]
([Child puts finger on the top row, left-most work])

☐ Correct ☐ Incorrect ☐ No response

Tsopano ndionetse mbali imene udzawerenga motsatira.
[Now show me in which direction you would read next.]
([Child moves finger from left to right.])

☐ Correct ☐ Incorrect ☐ No response

Ukafika kotsirizira kwa mzere, udzawerenga kuti motsatira?
[When you get to the end of the line, where would you read next?]
([Child moves finger to left-most word of second line])

☐ Correct ☐ Incorrect ☐ No response

Letter Sound Knowledge

Pano ndili ndi tsamba limene liri ndi malembo a alifabeti ya muchinyanja. Coonde ndizue MAMVEKERO a malembo a alifabeti amene ungather kuwerenga. Usanene maina ake. Koma mvekero zake. Here is a page full of letters of the Chinyanja alphabet. Please tell me the SOUNDS of as many letters of the alphabet as you can. Not their names, but their sounds.

[point to the letter A] **Mwacitsanzo, mvekero la limbo ili ndi /a/.** For example, the sounds of this letter is /a/.

[point to the letter P] **Tiye tiyese: ndiuze mvekero la limbo ili:** Let's practice: Let me the sound of this letter.

Correct: Cabwino, mvekero la limbo ili ndi /p/. Good, the sound of this letter is /p/.

Incorrect: Mvekero la lembo ili ndi /p/. The sound of this letter is /p/.

[point to the letter L] **Tsopano tiye tiyese lembo lina. Ndiuze mvekero la lembo ili.** Now let us try another one. Tell me the sound of this letter.

Correct: Cabwino, mvekero la lembo ili ndi /l/. Good, the sound of this letter is /l/.

Incorrect: Mvekero la lembo ili ndi /l/. The sound of this letter is /l/.

[point to first letter] **Ndikanena kuti "yamba", uyambire apa ndi kupitiriza mopingasa tsamba ili. Lata pa lembo lirilonse ndipo ndiuze mvekero la lembo limenelo mmau okweza. Uwerenge mwamsanga ndiponso modekha. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wafika pa lembo limene sudziwa, pitiriza kupita ku lembo lotsatira. Ika cala cako pa lembo loyamba. Wakonzeka? Yamba.** When I say "Begin," you start here and go across the page. Point to each letter and tell me the sound of that letter in a loud voice. Read as quickly and carefully as you can. I will remain silent and listen while you read. If you come to a letter you do not know, go on to the next letter. Put your finger on the first letter. Ready? Begin.

m	N	K	l	d	k	A	J	m	u
C	d	b	o	L	I	U	K	A	w
G	n	a	e	s	E	A	D	I	g
I	r	A	a	v	f	A	T	W	I
D	a	t	L	N	a	A	M	i	Y
t	u	z	N	i	I	N	k	e	O
u	Z	P	i	U	N	i	M	i	I
A	p	A	a	B	W	T	k	c	M
a	w	N	m	E	R	a	A	k	a
n	A	o	I	O	n	a	U	T	S

Time Remaining

Autostop? ☐

Letter Sound Feedback

Did the respondent use English letter names in the letter-sound test?

☐ More than 5 times ☐ 3-5 times ☐ 1-3 times ☐ Not at all ☐ Don't remember/don't know

Non-Word Decoding

Apa pali mau opangidwa mcinyanja. Ndifuna kuti uwerenge mau amene ungakwanitse kuwerenga, Uwerenge mau awa osati masipelo. Here are some made-up words in Chinyanja. I would like you to read as many as you can. Do not spell the words, but read them.

[point to the word "oli"] **Mwacitsanzo, liu lopangidwa ili ndi: "oli"** For example, this made-up word is: "oli".

[point to the word "koki"] **Tiye tiyese: conde werenga liu ili.** Let's practice. Please read this word.

Correct: wacita bwino, Liu ili ndi "koki" Good, this made-up word is "koki."

Incorrect: Lui lopangidwa ili ndi "koki" This made-up word is "koki."

[point to the word "cota"] **Tsopano tiye tiyese liu lina: conde werenga liu ili:** Now let us try another one. Please read this word.

Correct: wacita bwino, liu lopangidwa ili ndi "cota" Good, this made-up word is "cota."

Incorrect: Liu lopangidwa ili ndi "cota" This made-up word is "cota."

[point to first word] **Ndikanena kuti "yamba" uyambire apa ndipo uwerenge mopingasa patsamba ili. Lata liu lirilonse ndipo uliwerenge mokweza mau. Uwerenge mofulumira ndi mosamala mmene ungakwanitsire. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wapeza liu limene sudziwa, pita ku liu lotsatira. Ika cala cako pa liu loyamba. Wakonzeka? Yamba.** When I say "Begin," start here *[point to first word]* and read across the page *[point]*. Point to each word and read it in a loud voice. Read as quickly and carefully as you can. I will remain silent and listen while you read. If you come to a word you do not know, go on to the next word. Put your finger on the first word. Ready? Begin.

nipe	atapi	gelu	kelo	mdzimu
ninane	wondi	umbe	rizi	ninda
ledesi	fikiraku	tomo	ngalo	zirama
yu	ane	mwane	mukudi	dzimo
liraku	ia	anuli	wekusera	dzimoli
cofukwa	udi	kubu	anauma	mtisinaka
wera	eka	diko	amoi	kasuci
ateta	lia	nacho	komi	labo
menepa	ncheto	ndaako	nthua	balo
mtanyama	mtutu	ndokonda	mtingi	ko

Time Remaining

Autostop? ☐

Oral Passage Reading

Show the child the story in the pupil stimuli booklet. Say:

Apa pali ka nthano kakafupi. Ndifuna kuti uwerenge mokweza, mofulumira komanso mosamala. Ukatsiriza kuwerenga, ndizakufunsa mafunso onena za nkhani imene wawerenga. Ndikanena kuti "yamba," uwerenge bwino kwambiri mmene ungakwanisire. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wapeza liu limene sudziwa, pita kuliu lotsatira. Ika cala cako pa liu loyamba. Wakonzeka? Yamba. Here is a short story. I want you to read it aloud, quickly but carefully. When you finish, I will ask you some questions about what you have read. When I say "Begin," read the story as best you can. I will remain silent and listen while you read. If you come to a word you do not know, go on to the next word. Put your finger on the first word. Ready? Begin.

Amai	anapita	kumsika	m'masana	tsiku
lina.	Anasiya	mwana	ndi	mkulu
wake	Dolika.	Anzake	a	Dolika
anabwera	kudzamtenga	pamodzi	ndi	mwanayo.
Dolika	ndi	anazke	anaphuzitsa	mwana
kuyimba.	Anamuphunzita	nyimbo	ya	alifabeti.
Atabwerako	Kumsika	amai,	anapeza	mwana
ali	kuyimba.	Amai	anakondwera	kwambiri.

Time Remaining

Autostop? ☐

Reading Comprehension

Tsopano ndidzakufunsa mafunso ocepa onena za nthano imene wawerenga. Yesa kuyankha mafunso mmene ungakwanisire. Now I am going to ask you a few questions about the story you just read. Try to answer the questions as well as you can.

1. Ndani anapita kumsika? (Amai)

☐ Correct ☐ Incorrect ☐ No response

2. Mwana anatsala ndi ndani? (Dolika, kapena azake a Dolika, Dolika ndi Azake)

☐ Correct ☐ Incorrect ☐ No response

3. Kodi mwana anaphunzitsiwa kucita ciani? (Kuyimba, Kuyimba nyimbo ya alifabeti)

☐ Correct ☐ Incorrect ☐ No response

4. Kodi mwana anadziwa bwanji kuyimba nyimbo ya alifabeti? (Dolika ndi anzake anamphunzitsa, anzake a Dolika anamuphuzitsa, Dolika anamuphuzitsa)

☐ Correct ☐ Incorrect ☐ No response

5. N'cifukwa ciani amai anakondwera? (Mwana anali kuyimba)

☐ Correct ☐ Incorrect ☐ No response

Listening Comprehension

Do not allow the child to look at the passage or the questions. Say,

Ndidzakuwerengera ka nthano/nkhani mokweza KAMODZI ndipo pambuyo pake ndidzakufunsa mafunso. Conde umvetsere mosamalira ndipo uvankhe mafunso mmene ungakwanitsire. Wakonzeka? Tive Tivambe.

Patsiku Lolemba, Mangani anapita kusukulu.

Ananyamula mabuku ndi nyama m'chola cake.

Pamene anali kuyenda, anapeza galu wamkulu panjira.

Anafuna kuthawira pathengo koma anagwa pansi.

Yunifomu yake inada ndipo galu anatenga nyama yake.

Mangani anathawira kunyumba.

Pamene anafika kunyumba, m'bale wake anamubwereka yunifomu yake. Anakondwera.

1. Ndi tsiku liti pamene Mangani anapita kusulu? (Pa Lolemba)

☐ Correct ☐ Incorrect ☐ No response

2. Ananyamula ciani mu chola cake? (Mabuku, Nyama, Mabuku ndi nyama)

☐ Correct ☐ Incorrect ☐ No response

3. N'ciani cimene anapeza panjira? (Anapeza galu wamkulu)

☐ Correct ☐ Incorrect ☐ No response

4. Ndi cifukwa ciani Mangani anathawa galu? (Anaopa kuti galu angamulume, nyama, Anaopa, Galu wamkulu, kapena zonse izi (of all of these))

☐ Correct ☐ Incorrect ☐ No response

5. Ndi cifukwa ciani m'bale wake anamubwereka yunifomu Mangani? (Cifukwa yunifomu yake inada, Anagwa)

☐ Correct ☐ Incorrect ☐ No response

Student Survey

1. Kodi pa sukulu panu aphunzitsi amakukambitsani mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Iyai ☐ Sindidziwa kanthu/ ndilibe yankho

2. Pa sukulu panu kodi anzanu amakukambitsani mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Iyai ☐ Sindidziwa kanthu/ ndilibe yankho

3. Kodi pasukulu panu muli ndi mabuku ndi zowerenga zina zolembedwa mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Palibiletu ☐ Sindidziwa kanthu/ ndilibe yankho

4. Kodi pasukulu panu muli ndi mabuku ndi zowerenga zina zolembedwa mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Iyai ☐ Sindidziwa kanthu/ ndilibe yankho

5. Ngati muli kunyumba kwanu mulankhula kwa azing'ono ndi akulu anu mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Iyai ☐ Sindidziwa kanthu/ ndilibe yankho

6. Ngati muli kunyumba kwanu mulankhula ndi makolo anu mu CiNyanja?

☐ Inde ☐ Nthawi zina ☐ Iyai ☐ Sindidziwa kanthu/ ndilibe yankho

7. Kunyumba kwanu ziliko zowerenga za mu CiNyanja?

☐ Inde ☐ Ziliko zowerenga ☐ Kulibiletu ☐ Sindidziwa kanthu/ ndilibe yankho

8a. Panyumba panu muli ndi: Wailesi?

☐ Inde ☐ Iyai

8b. Lamya/foni yam'manja?

☐ Inde ☐ Iyai

8c. Magesi

☐ Inde ☐ Iyai

8d. Wailesi ya kanema?

☐ Inde ☐ Iyai

8e. Cimbuszi cha mkati mwanyumba?

☐ Inde ☐ Iyai

8f. Njinga yopalasa?

☐ Inde ☐ Iyai

8g. Njinga ya moto?

☐ Inde ☐ Iyai

8h. Galimoto laling'ono, chimbayambaya, galimoto wa mtundu wa 4x4, talakita kapena bothe la moto

☐ Inde ☐ Iyai

9. Kodi ngi mtunda wautali bwanji kuchokera komwe mukhala kuti mufike ku masitolo kuyenda ndimendo?

☐ Kuchepelelako mphindi makumi awiri ☐ Kupitilira mphindi mahumu awiri koma kosafika ola lathunthu ☐ Kupitilira ora ☐ Sindidziwa/ Ndilibe yankho

10. Usiku wadzulo muntenga maola angati muli kugwira nchito za panyumba?

☐ Palibe ☐ Angapo ☐ Ambiri-mbiri ☐ Hake etse letho

☐ Sindidziwa/ Ndilibe yankho

11. Kodi amai anu akhoza kuwerenga?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

12. Kodi abambo anu(atate) akhoza kwerenga?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

13. Kodi makolo anu anafika pati ndi maphunziro yawo?

- ☐ Ali ndi dipoloma
- ☐ Anatsiriza cabe maphunziro opitilira sekondale sukulu
- ☐ Anatha chadbe maphunziro a sekondale sukulu
- ☐ Analekezela cabe pa maphunziro a pulaimale sukulu
- ☐ Sindidziwa/ Ndidlibe yankho

14. Kodi pali wina wace aliyense panyumba omwe akhoza kumuwerengerani nthano iliyonse?

- ☐ Inde
- ☐ Iyai
- ☐ Sindidziwa/ Ndidlibe yankho

15. Kodi pali wina wace aliyense panyumba omwe ali ndi chidwi co ona nchito yanu yakusukulu?

- ☐ Inde
- ☐ Iyai
- ☐ Sindidziwa/ Ndidlibe yankho

16. Mulungu wathawu, ndi masiku angati pomwe munawerenga lilmodzi ndi wina wache ndiponso kumalo kwina kwacha osati muli kusuklu?

- ☐ Masiku asanu ndi awiri kapena mulungu wathunthu
- ☐ Masiku anai mpaka asanu ni limodzi
- ☐ Kucoketrsiku limodzi mpaka lacitatu
- ☐ Sindinawerenge ndipan'ono ponse ndi ena kunyumba mulungu watha
- ☐ Sindidziwa/ Ndidlibe yankho

17. Kodi muli ndi buku lomwe limuthandizirani kuti muphunzire kuwerenga mu CiNyanja?

- ☐ Inde
- ☐ Iyai
- ☐ Sindidziwa/ Ndidlibe yankho

18. Kodi zipangizo zowerengera za msikulu nozolebedwa mu CiNyanja?

- ☐ Inde
- ☐ Iyai
- ☐ Sindidziwa/ Ndidlibe yankho

19. Kodi muli ndi buku lomwe likutandizirani ku phunzira masamu?

- ☐ Inde
- ☐ Iyai
- ☐ Sindidziwa/ Ndidlibe yankho

20. Kodi aphunzitsi anu amakuthandizirani mkuphunzira kuwerenga pamulungu uliwonse?

- ☐ Nthawi zonse
- ☐ Nthawi zina
- ☐ Iyai ndipong'ono ponse
- ☐ Sindidziwa/ Ndidlibe yankho

21. Kodi aphunzitsi anu amakuthandizirani mkuphunzira kulemba pamulungu uliwonse?

- ☐ Nthawi zonse
- ☐ Nthawi zina
- ☐ Iyai ndipong'ono
- ☐ Sindidziwa/ Ndidlibe yankho

22. Pomwe muli kusukulu mumakhala ndi nthawi yakuwerenga kwaumwini mwakachetechete?

☐ Nthawi zonse ☐ Nthawi zina ☐ Iyai ndipong'ono ponse ☐ Sindidziwa/ Ndidlibe yankho

23. Pomwe muli pasukulu, kodi aphunzitsi anu amakufunsani pa zomwe mukuwerenga?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

24. Ndinthawi zingati pomwe aphunzitsi amuthandizani pomwe mukumayesa-yesa mobvutikira kuwerenga?

☐ Nthawi zonse ☐ Nthawi zina ☐ Iyai ndipong'ono ponse ☐ Niganiza kuti sindifukira thansizo

☐ Sindidziwa/ Ndidlibe yankho

25. Kodi aphunzitsi anu amayesetsa kukupangani kukhala muwerengi wabwino?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

26. Kodi mukonda kapena kuipitsidwa ndi kuwerenga?

☐ Akonda knwabiri kuwerenga ☐ Ndikonda zowerenga werenga ☐ Sinikonda zowerenga werenga ☐ Nizonda zowerenga werenga

☐ Sindidziwa/ Ndidlibe yankho

27. Mumazimva motani pomwe mukuphunzira kuwerenga pa sukulu?

☐ Nizimva wotsimikizira ☐ Ndinali wotekeseka ☐ Sindinakonde kuphunzira kuwerenga ☐ Sindidziwa/ Ndidlibe yankho

28. Kuwerenga ndikofunikira kwambiri paumoyo wanga wamsogolo.

☐ Ndibvomekeza zolimba ☐ Ndivomekeza ☐ Sindibvomekeza ☐ Ndikalibe cosankha

29. Kodi munaphunzirako ku sukulu ya ana pomwe musanayambe Kalasi yoyamba ya sukulu?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

30. Kodi ndi kalasi iti yomwe munalimo caka cathachi?

☐ Sikulu yakumwana ☐ Kalasi loyamba ☐ Kalasi laciwiri ☐ Sinili kusukulu

☐ Sindidziwa/ Ndidlibe yankho

31. Kodi ndi kangati pomwe simunabwere ku sukulu mwa kulova mwenzi wathawu?

☐ Kupitilira masiku asanu ☐ Pakati pa masiku atatu andi asanu ☐ Pakati pa triku limodzi and atatu ☐ Sininapezeke

32. Kodi ndi masiku angati pomwe munapezekako ku kalasi yophunzira kuwerenga mulungu wathawu?

☐ Masiku asanu ☐ Pakati pa masiku atatu ndi asanu ☐ Osakwanira masiku awiri ☐ Sindidziwa/ Ndidlibe yankho

33a. Kodi mumagwiritsira nchito zitsulo za makono ziri zonse kukuthandizirani kuphunzira?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

33b. Ngati mubvomera kuti munatero, kudi ndi chitsuro chiti comwe munagwirizira nchito pa kuphunzira?

☐ Iamya ☐ Komputa ☐ Ka komputa kochepera Ka thabuleti ☐ Zina

☐ Sindidziwa/ Ndidlibe yankho

34. Kodi munagwirritsirako nchito chiri chonse cha izi pa kuphunzira kuwerenga?

☐ Lamya ☐ Ka komputa kochepera Ka thabuleti ☐ Komputa ☐ Zina

☐ Sindidziwa/ Ndidlibe yankho

35. Ngati muli ndi lamya ya mmanja, kodi mugaigwiritsirako nchito pa kuphunzira kwanu za kuwerenga?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

36. Ngati munatero, ndi kangati pomwe munagwiritsira nchito citurulocho pa mulungu wapiti?

☐ Nthawi zonse ☐ Kawiri pa tsiku ☐ Kamodzi Patsiku ☐ Sindinatero NIkamodzi komwe

☐ Sindidziwa/ Ndidlibe yankho

37a. Kodi ndinu omasuka motani kugwiritsira nchito lamya yanu pa kugwiritisra chithandizo cakuwerenga cha Graphogame?

☐ Omasuka kwanbiri ☐ Omasukako ndithu ☐ Not comfortable ☐ Sindidziwa/ Ndidlibe yankho

37b. Kodi kutenga mbali mu pologalamu ino kwakuthandizirani kukhala omasuka ndi kuphunzira kuwerenga mogwiritsirira nchito chithandizo ca kuwerenga cha Graphogame?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

37c. Kodi mungakonde kuphunzira kuwerenga mogwiritsira nchito chitandizo ca Graphogame kapen kuphunzira cabe zakuwerenga mu kalasi?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

38. Kugwiritsira nchito cithandizo ca kuwerenga cha Graphogame candithandiza kupititsa pasogolo kuwerenga kwanga.

☐ Ndibvomera kwanbiri ☐ Ndibvomera ☐ Ndikana ☐ Ndikalibe cosankha

39. Ndifuna kupitilizabe kugwiritsira nchito Graphogame kuti ndiphunzire kuwerenga.

☐ Ndibvomera kwanbiri ☐ Ndibvomera ☐ Sindibvomera ☐ Ndikalibe cosankha

40. Zinthu zomwe mwakhoza kuwerenga ndi chitandizo cha Graphogame zinali:

☐ Zopusa ☐ Mwina zobvutako ☐ Zobvutako nthawi zonse ☐ Ndikalibe cosankha

41. Ndikonda Graphogame

☐ Ndibvomera kwanbiri ☐ Ndibvomera ☐ Ndikana ☐ Ndikalibe cosankha

42. Ndikonda nkhani zomwe nawerenga chaka chino.

☐ Ndibvomera kwanbiri ☐ Ndibvomera ☐ Ndikana ☐ Ndikalibe cosankha

43. Kugwiritsira nchito chithandizo cakuwerenga cha Graphogame chasintha maganizo anga pa kuwerenga.

☐ Zotulukapo zabwino ☐ Sizinatelo ☐ Mosiyanako

44. Kugwiritsira nchito chithandizo chakuwerenga cha Graphogame chapanga kuti ndi onjezere nthawi yanga ya kwerenga.

☐ Ndibvomera kwanbiri ☐ Ndibvomera ☐ Sindibvomera ☐ Ndikalibe cosankha

45. Kodi makolo anu adziwa kuti mugwiritsira nchito chithandizo cha kuwerenga cha Graphogame?

☐ Inde ☐ Iyai ☐ Sindidziwa/ Ndidlibe yankho

46. Kodi makolo anu akumva bwanji pakuona kuti inuyo mugwiritsira nchito chitandizo cha kuwerenga cha Graphogame?

☐ Okondwera kwanbiri ☐ Okondwera ☐ Okondwera kwambirir ☐ Ndikalibe kuganizirapo

47. Kodi pali chifukwa china cace cirichonse comwe simunafunire ku gwiritsira nchito chithandizo chakuwerenga cha Graphogame?

☐ Palibe chifukwa chiri Conse chingapelekedwe ☐ Kambana ana omwe amakusekani ☐ Chula aziphunzitsi Omwe Sali kufuna Zogwiritsira nchito Graphogame ☐ Parents not wanting them to play

☐ Other reason given ☐ Sindidziwa/ Ndidlibe yankho

48. Pamulungu wathawu kodi pali nthawi pomwe mumafuna kugwiritsira nchito chithandizo chakuwerenga cha Graphogame koma danga simunakhale nalo la kupeza Graphogame?

☐ Kamodzi pa mulungu ☐ Kawiri pa mulungu ☐ Katatu kapena mopitililapo pa mulungu ☐ Iyai

☐ Ha ke tsebe, ha no karabo

49. Pamulungu wathau ndinagwiritsira nchito chithandizo cha kiwerenga cha Graphogame?

☐ Kamodzi pa mulungu ☐ Kawiri pa mulungu ☐ Kawiri kapena mopitililapo pa mulungu ☐ Nthawi zonse

☐ Iyai ndipong'ono ponse

50. Kodi mudalira kugwiritsira nchito ukatswiri wa zithandizo zakuwerenga kapena mudalira pa ana asukulu anzanu kapena mabwenzi anu?

☐ Nthawi zambiri ☐ Nthawi zina ☐ Iyai ndipong'ono ponse

Annex B: **Midline EGRA Instrument**

Enumerator Name	
Date and Time	
Date	
Time	
School Location	
School	
ID	
ID	

Verbal Consent

It is important to establish a playful and relaxed rapport with the child. The child should perceive the assessment almost as a game to be enjoyed rather than a severe situation. It is important to read the directions slowly and clearly. After you have finished, thank the child for their time and effort.

Uli bwanji. Dzina lango ndine.....ndipo ndikhala ku.....Ndingakonde kukuuza za moyo wanga. Good morning. My name is ___ and I live in ___. I'd like to tell you a little bit about myself.

[Number and ages of children; favourite sport, radio or television program, etc.]

1. **Kodi umakonda kucita ciani ngati siuli mu sukulu?** What do you like to do when you are not in school?

   [Wait for responses; if pupil is reluctant, ask question 2, but if they seem comfortable continue to verbal consent].

2. **Kodi ndi masewera otani ameme umakonda kusewera?** What games do you like to play?

READ THE FOLLOWING WORD-FOR-WORD

Ndifuna kukuuza cifukwa cake ndabwera kuno lero. Ndigwira nchito pa sukulu yama phunziro apamwamba ya mu Zambia (Universityr of Zambia). Ndipo tikufuna kumvetsetsa mmene ana amaphunzirira kuwerenga. Iwe wasankhidwa mwamwar. Let me tell you why I am here today. I work with the University of Zambia, and we are trying to understand how children learn to read. You were picked by chance.

Ndifuna thandizo lado pa nkhaniyi. Koma suyenera kutengako mbali ngati sufuna. We would like your help in this. But you do not have to take part if you do not want to.

Ife tizachita sewero la kuwerenga. Ine ndizakufunsa kuwerenga malembo, mau ndi ka nthano kakafupi mokweza mau. Ndizakufunsanso kuzindikira ndi kuyankha mafunso ocepa. We are going to play a reading game. I am going to ask you to read letters, words, and a short story out loud.

Mwakugwiritsa nchito lamya iyi, ndizaiemba mayakho ako. I am going to use this "phone" to record your answers.

Zimene tizachita pano si mayeso ndipo sizidzakhudza maphunzilo ako pasukulu lino. This is NOT a test, and it will not affect your grade at school.

Nsizakufunsanso mafunso ena monga kumene umayeselera kuwerenga ndiponso ngati ukonda kuwerenga. I will ask you other questions about where you practice reading and whether you like it.

Kaciwirinso, sungatengeko mbali ngati sufuna kutero. Tikayamba kufunsa mafunso, ngati siufuna kuyankha funso ungakhale cete, zilibwino cabe. Once again, you do not have to participate if you do not wish to. Once we begin, if you would rather not answer a question, that's all right.

Kodi uli ndi mafunso alionse? Do you have any questions?

Kodi uvomela kutengako mbali musewero iyi? Would you like to participate?

Kodi wakonzeka kuti tiyambe? Are you ready to get started?

(If verbal consent is not given, thank the child and move on the next child.)

Check box if verbal consent is given. ☐

Demographics

Name

Age

☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12 ☐ Other

Grade

☐ G2 ☐ Other

Sex

☐ Male ☐ Female

Letter Sound Knowledge

Pano ndili ndi tsamba limene liri ndi malembo a alifabeti ya muchinyanja. Coonde ndizue MAMVEKERO a malembo a alifabeti amene ungather kuwerenga. Usanene maina ake. Koma mvekero zake. Here is a page full of letters of the Chinyanja alphabet. Please tell me the SOUNDS of as many letters of the alphabet as you can. Not their names, but their sounds.

[point to the letter A] **Mwacitsanzo, mvekero la limbo ili ndi /a/.** For example, the sounds of this letter is /a/.

[point to the letter P] **Tiye tiyese: ndiuze mvekero la limbo ili:** Let's practice: Let me the sound of this letter.

Correct: Cabwino, mvekero la limbo ili ndi /p/. Good, the sound of this letter is /p/.

Incorrect: Mvekero la lembo ili ndi /p/. The sound of this letter is /p/.

[point to the letter L] **Tsopano tiye tiyese lembo lina. Ndiuze mvekero la lembo ili.** Now let us try another one. Tell me the sound of this letter.

Correct: Cabwino, mvekero la lembo ili ndi /l/. Good, the sound of this letter is /l/.

Incorrect: Mvekero la lembo ili ndi /l/. The sound of this letter is /l/.

[point to first letter] **Ndikanena kuti "yamba", uyambire apa ndi kupitiriza mopingasa tsamba ili. Lata pa lembo lirilonse ndipo ndiuze mvekero la lembo limenelo mmau okweza. Uwerenge mwamsanga ndiponso modekha. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wafika pa lembo limene sudziwa, pitiriza kupita ku lembo lotsatira. Ika cala cako pa lembo loyamba. Wakonzeka? Yamba.** When I say "Begin," you start here and go across the page. Point to each letter and tell me the sound of that letter in a loud voice. Read as quickly and carefully as you can. I will remain silent and listen while you read. If you come to a letter you do not know, go on to the next letter. Put your finger on the first letter. Ready? Begin.

k	m	u	A	l	m	D	n	J	K
b	A	d	k	l	U	W	o	C	L
G	s	D	A	g	l	E	n	a	e
v	W	r	i	a	A	A	f	l	T
A	a	L	a	M	t	Y	i	D	N
i	N	z	l	t	e	U	O	k	N
P	Z	i	N	M	l	l	i	u	U
W	k	A	M	T	B	P	c	A	a
E	a	a	w	m	R	H	A	a	N
S	U	n	a	o	l	A	T	O	n

Time Remaining

Autostop? ☐

Non-Word Decoding

Apa pali mau opangidwa mcinyanja. Ndifuna kuti uwerenge mau amene ungakwanitse kuwerenga, Uwerenge mau awa osati masipelo. Here are some made-up words in Chinyanja. I would like you to read as many as you can. Do not spell the words, but read them.

[point to the word "oli"] **Mwacitsanzo, liu lopangidwa ili ndi: "oli"** For example, this made-up word is: "oli".

[point to the word "koki"] **Tiye tiyese: conde werenga liu ili.** Let's practice. Please read this word.

Correct: wacita bwino, Liu ili ndi "koki" Good, this made-up word is "koki."

Incorrect: Lui lopangidwa ili ndi "koki" This made-up word is "koki."

[point to the word "cota"] **Tsopano tiye tiyese liu lina: conde werenga liu ili:** Now let us try another one. Please read this word.

Correct: wacita bwino, liu lopangidwa ili ndi "cota" Good, this made-up word is "cota."

Incorrect: Liu lopangidwa ili ndi "cota" This made-up word is "cota."

[point to first word] **Ndikanena kuti "yamba" uyambire apa ndipo uwerenge mopingasa patsamba ili. Lata liu lirilonse ndipo uliwerenge mokweza mau. Uwerenge mofulumira ndi mosamala mmene ungakwanitsire. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wapeza liu limene sudziwa, pita ku liu lotsatira. Ika cala cako pa liu loyamba. Wakonzeka? Yamba.** When I say "Begin," start here *[point to first word]* and read across the page *[point]*. Point to each word and read it in a loud voice. Read as quickly and carefully as you can. I will remain silent and listen while you read. If you come to a word you do not know, go on to the next word. Put your finger on the first word. Ready? Begin.

gelu	mdzimu	nipe	kelo	atapi
umbe	ninda	ninane	rizi	wondi
tomo	zirama	ledesi	ngalo	fikiraku
mwane	dzimo	yu	mukudi	ane
annuli	dzimoli	liraku	wekusera	ia
kubu	mtisinaka	cofukwa	anauna	udi
diko	kasuci	wera	amoi	eka
nacho	labo	ateta	komi	lia
akonda	balo	menepa	anthu	ncheto
ndokonda	ko	mtanyama	mtingi	mntutu

Time Remaining

Autostop? ☐

Oral Passage Reading

Show the child the story in the pupil stimuli booklet. Say:

Apa pali ka nthano kakafupi. Ndifuna kuti uwerenge mokweza, mofulumira komanso mosamala. Ukatsiriza kuwerenga, ndizakufunsa mafunso onena za nkhani imene wawerenga. Ndikanena kuti "yamba," uwerenge bwino kwambiri mmene ungakwanisire. Ndizakhala cete ndi kumvelera iwe pamene uli kuwelenga. Ngati wapeza liu limene sudziwa, pita kuliu lotsatira. Ika cala cako pa liu loyamba. Wakonzeka? Yamba. Here is a short story. I want you to read it aloud, quickly but carefully. When you finish, I will ask you some questions about what you have read. When I say "Begin," read the story as best you can. I will remain silent and listen while you read. If you come to a word you do not know, go on to the next word. Put your finger on the first word. Ready? Begin.

Chikondi	anali	mwana	wamng'ono.	Tsiku
lina,	atacoka	kusukulu	ndiwo.	Alikuyenda
anakumana	ndi	mnzake	Beti.	Awiriwo
anayamba	kusewera	pamodzi.	Posewerapo	anataya
ndalama	ina.	Motero,	anagula	repu
cabe	osagula	kabichi.	Anasowa	cocita.
Pamene	amaganizira	cocita	anaona	abambo
ena	akubwera.	Iwowa	ananyamula	nkhuni
ndi	basiketi.	Mtsikanayo	anawathandiza	kunyamula
basiketi	ndipo	anamupatsa	ndalama.	Iye
anatenga	ndalama	ija	nagulira	kabichi
nabwerera	kunyumba.			

Time Remaining

Autostop? ☐

Reading Comprehension

Tsopano ndidzakufunsa mafunso ocepa onena za nthano imene wawerenga. Yesa kuyankha mafunso mmene ungakwanisire. Now I am going to ask you a few questions about the story you just read. Try to answer the questions as well as you can.

1. Kodi Chikondi anali mwana wotani? (wamng'ono)

☐ Correct ☐ Incorrect ☐ No response

2. Kodi mnzake wa Chikondi anali ndali? (Beti)

☐ Correct ☐ Incorrect ☐ No response

3. Dindiwo zanji zimene Chikondi anapita kukagula? (Repu and/or kabichi)

☐ Correct ☐ Incorrect ☐ No response

4. Kodi anapeza bwanji ndalama ina yogulira ndiwo zina? (anampatsa)

☐ Correct ☐ Incorrect ☐ No response

5. N'cifukwa ciani Chikondi anampatsa ndalamu abambo? (kumuwonga)

☐ Correct ☐ Incorrect ☐ No response

Listening Comprehension

Do not allow the child to look at the passage or the questions. Say,

Ndidzakuwerengera ka nthano/nkhani mokweza KAMODZI ndipo pambuyo pake ndidzakufunsa mafunso. Conde umvetsere mosamalira ndipo uvankhe mafunso mmene ungakwanitsire. Wakonzeka? Tive Tivambe.

Tsika loyamba sukulu ndinaona zodabwitse.

Inalnali nthwi yamvula ndipo pansi panali matika.

Ndina lowa mu kalasi, ndi kukhala padesiki.

Mnyamata wamkulu ana bwera pafupi ndiine nandilanda nthoci.

Ine ndina lira mwambiri ndi kupita kumudzi.

1. Kodi nthawi yamvula pansi pama khala ciani? (matika)

☐ Correct ☐ Incorrect ☐ No response

2. Ndani amane ana landa nthoci mwanayu? (mnyamata wamkulu)

☐ Correct ☐ Incorrect ☐ No response

3. Kodi mwanayu ana nyamula cakudya cotani popita kusukulu? (ntochi)

☐ Correct ☐ Incorrect ☐ No response

4. Niciani cimene mwanayu ana cita atalandidwa nthoci poyamba? (analira)

☐ Correct ☐ Incorrect ☐ No response

5. Ngati unalandidwa nthoci ukana cita ciani? (ndikanauza aphunzitsi)

☐ Correct ☐ Incorrect ☐ No response

Annex C: **Classroom Observation Checklist**

This Box Filled by Enumerator

Province	
District	
School Name	
Enumerator Name	
Name of Team Leader	

Instructions

1. Make prior arrangements with the teacher to attend their literacy class. Be sure to arrive for the lesson on time.
2. Re-introduce yourself to the teacher (see below) and confirm permission to conduct the interview with the teacher.
3. Ask the teacher to give you sitting space at the side or the end of the classroom where you will not be a distraction to the learners and the teacher.
4. As the lesson is progresses, please look out for aspects of the lesson highlighted in the checklist and tick the things you will notice happen during the lesson. It is important that you familiarize yourself with the questions in the checklist before attending the lesson observation.

Part 1: **Introduction and Teacher Identification**

Introduction, Confidentiality and Consent

Start time: _____ : _____ hr/min

Let me re-introduce myself. My name is _____ and I am working for the Centre for Promotion of Literacy in Sub-Saharan Africa based at the University of Zambia, Department of Psychology under the All Children Reading Project. Thank you for letting me observe your classroom.

I have a checklist which I will make reference to as I observe your lesson. The statements on the checklist mainly focus on the content of your literacy lesson.

Please note that the information collected during this session will be kept strictly confidential and will only be accessed by members of the research team. Please note that this information will be used for research purposes.

Please ensure that you conduct your literacy lesson as usual. Nothing should change because of the observation I will conduct. We greatly appreciate your cooperation.

Do you have any questions for me now? Please sign here: _____

Do you agree to participate in the survey? a. ☐ Yes b. ☐ No

If you agree please sign here: _____

Teacher Information

1. Name of teacher observed

Last

First

ID No

2. Age

3. Teacher cell phone (if applicable)

4. Sex

1. ☐ Male 2. ☐ Female

5. How long have you been teaching?

6. What grades do you currently teach? *(mark all that apply)*

1. ☐ 1 2. ☐ 2 3. ☐ 3 4. ☐ 4 5. ☐ 5 6. ☐ 6 7. ☐ 7 8. ☐ Other

7. What subjects do you currently teach? *(mark all that apply)*

1. ☐ All 2. ☐ Language/ literacy 3. ☐ Math 4. ☐ Science 4. ☐ Other

8. What is the highest level of education you have completed?

1. ☐ Primary school 2. ☐ Secondary school 3. ☐ Teacher's college 4. ☐ University/ bachelor's program

5. ☐ Master's program 6. ☐ Other

9. What teaching qualifications do you have? *(mark all that apply)*

1. ☐ Primary Teachers Certificate 2. ☐ Primary Teachers Diploma 3. ☐ BED Primary 4. ☐ Secondary Teachers Diploma

5. ☐ Bachelor's Degree 6. ☐ Untrained 7. ☐ Other

Part 2: **Teacher Classroom Practices**

Teaching Literacy in the Local Language – CiNyanja

10. Does the teacher use ciNyanja to welcome the class and introduce the lesson?

1. ☐ Yes 2. ☐ No 3. ☐ Both English and ciNyanja

11. Does the teacher use ciNyanja to teach the literacy lesson?

1. ☐ Yes 2. ☐ No 3. ☐ Both English and ciNyanja

12. Does the teacher make reference to Cinyanja letter sounds and syllables during the lesson?

1. ☐ Yes 2. ☐ No

13. Did the teacher use English letter names during the lesson?

1. ☐ Yes 2. ☐ No

14. How did the teacher present the letter sounds, syllables, or sentences used during the lesson?

1. ☐ On the board 2. ☐ On a paper chart 3. ☐ An electronic gadget 4. ☐ Flash cards

5. ☐ Books 6. ☐ just by saying and not using a visual aid 7. ☐ Not at all 8. ☐ Other (specify)

15. Did the teacher use repetition of key sounds, syllables and sentences during the lesson?

1. ☐ Yes 2. ☐ No

16. What language was used in the introduction and teaching of sounds, syllables and sentences?

1. ☐ ciNyanja only 2. ☐ English only 3. ☐ Both English and ciNyanja 4. ☐ Other (specify)

17. How many sounds, syllables, letters or sentences did the teacher focus on during the lesson?

1. ☐ One 2. ☐ Two 3. ☐ Three

18. How long did the reading lesson usually take?

1. ☐ One hour 2. ☐ Less than an hour 3. ☐ More than an hour

19. Which of the following did the teacher use during the lesson?

1. ☐ Songs 2. ☐ A story 3. ☐ Rhymes 4. ☐ Games (not GG)

5. ☐ Other (specify)

20. Did the teacher make the lesson fun and exciting for the learners?

1. ☐ Yes 2. ☐ No 3. ☐ Somewhat

21. Did the teacher encourage use of local language when they children responded in English?

1. ☐ Yes 2. ☐ No 3. ☐ Not applicable

22. Were the responses given by learners conducted individually or as a group?

1. ☐ Individual 2. ☐ Group 3. ☐ Both

Supporting Struggling Readers

23. Did the teacher make the lesson interactive?

1. ☐ Yes 2. ☐ No 3. ☐ Sometimes

24. Were some children at least given an opportunity to read during the lesson?

1. ☐ Yes 2. ☐ No

25. Did the teacher try to make sure that all/most of the learners understood the lesson?

1. ☐ Yes 2. ☐ No

26. Did the teacher identify those that did not understand the lesson?

1. ☐ Yes 2. ☐ No

27. Did the teacher try and help these learners?

1. ☐ Yes 2. ☐ No

28. How did the teacher try and help these learners during the lesson?

1. ☐ Encouraged and praised them
2. ☐ Gave them individualized attention
3. ☐ Ignored them
4. ☐ Gave them an easier question
5. ☐ Repeating the question
6. ☐ The teacher shouted at the learner

29. Did the teacher give the class any homework?

1. ☐ Yes 2. ☐ No

Story Reading and Telling

30. Did the teacher read a story to the learners?

1. ☐ Yes 2. ☐ No

31. Were the learners given an opportunity to read individually in silence during the lesson?

1. ☐ Yes 2. ☐ No

32. Were the children given an opportunity to read collectively during the lesson?

1. ☐ Yes 2. ☐ No

33. Were some children given the opportunity to read a text or story aloud to the class during the lesson?

1. ☐ Yes 2. ☐ No

34. Did the teacher conduct a spelling exercise during the lesson?

1. ☐ Yes 2. ☐ No

35. Did the teacher conduct a comprehension exercise during the lesson?

1. ☐ Yes 2. ☐ No

36. Did the teacher ask any questions about a story that was used during the lesson?

1. ☐ Yes 2. ☐ No

37. Were children given any books to use during the lesson?

1. ☐ Yes 2. ☐ No

38. Were children given any books to take home as reading books?

1. ☐ Yes 2. ☐ No

39. Were children given the opportunity to ask questions about the story during the lesson?

1. ☐ Yes 2. ☐ No

40. Were children encouraged to tell a story orally (a story from their heads)?

1. ☐ Yes 2. ☐ No

41. Were the learners praised and encouraged for any efforts they made during the lesson?

1. ☐ Yes 2. ☐ No

42. Did the teacher have flip charts of stories, either by the teacher or the learners put up in class (e.g. on the walls or written down) for the class to use?

1. ☐ Yes 2. ☐ No

Designing your own Literacy Games

43. Did the teacher use any of the games presented at the website during the literacy lesson? (Memory game for letters, Letter in the hand, Car chase for words, Syllable chart countdown, or Word Race)

1. ☐ Yes 2. ☐ No

44. Did you see the children playing the literacy games presented at the website during the break time?

1. ☐ Yes 2. ☐ No

45. Did the teacher use some new, self-made literacy game as part of the lesson? Describe the game you saw in the notes.

1. ☐ Yes 2. ☐ No

Singing as a Literacy Tool

46. Did the teacher sing the CAPOLSA songs with the class?

1. ☐ Yes 2. ☐ No

47. Did you hear the children singing the CAPOLSA songs during the school visit?

1. ☐ Yes 2. ☐ No

48. Did you see the teacher singing with the class and then using the words of the song during the literacy lesson?

1. ☐ Yes 2. ☐ No

Annex D: **ACR GCD Student Questionnaire Composites**

Language Consistency

The language consistency composite includes items related to students' use of ciNyanja at school and at home. Because the project sought to improve student reading skills in ciNyanja, the consistency and exposure of students to ciNyanja at school and at home were important to understanding how a student performed on the EGRA assessment.

The language consistency composite was comprised of six items for a maximum composite score of 6.0, with higher scores indicating more language consistency. The average composite score for the language consistency composite was 5.5 for the comparison group and 5.4 for the intervention group, indicating that all students were frequently exposed to ciNyanja.

Table D.1: **Language Consistency Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
At school, does your teacher talk to you in ciNyanja?	No	1	0.4	2	0.9
	Yes	215	92.7	199	90.9
	Sometimes	16	6.9	15	6.8
	Don't know / No response	0	0.0	3	1.4
At school, do your friends speak to you in ciNyanja?	No	8	3.4	1	0.5
	Yes	208	89.7	204	93.2
	Sometimes	15	6.5	11	5.0
	Don't know / No response	1	0.4	3	1.4
At school, are there reading materials in ciNyanja?	None	7	3.0	6	2.7
	Yes	212	91.4	203	92.7
	Sometimes	11	4.7	6	2.7
	Don't know / No response	2	0.9	4	1.8
At home, do you speak to your siblings in ciNyanja?	No	5	2.2	1	0.5
	Yes	219	94.4	210	95.9
	Sometimes	8	3.4	7	3.2
	Don't know / No response	0	0.0	1	0.5
At home, do you speak to the adults in your home in ciNyanja?	No	6	2.6	5	2.3
	Yes	216	93.1	202	92.2
	Sometimes	10	4.3	9	4.1
	Don't know / No response	0	0.0	3	1.4
At home, are there reading materials in ciNyanja?	No	70	30.5	73	33.3
	Yes	155	66.8	133	60.7
	Sometimes	5	2.2	10	4.6
	Don't know / No response	2	0.9	3	1.4

Socio-Economic Status

SES is a commonly reported composite used to describe an individual's or household's education, financial situation, and occupation, among other variables. Students were asked a series of questions that were used to estimate SES based on assets and characteristics of their home.

The SES composite was calculated using ten items for a maximum composite score of 10.0. Scores closer to ten indicate higher SES; scores closer to one indicate lower SES. Overall, students had relatively low SES composite scores: the average score for the comparison group was 4.2, and the average for the intervention group was 4.5. While nearly all students reported having a telephone/mobile phone in their house, only about 33 percent of students in the intervention group and about 23 percent in the comparison group reported having electricity. A very small proportion of students report having a toilet inside their house: about 11 percent in the intervention group and about 8 percent in the comparison group.

Table D.2: **SES Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
At your house, do you have a radio?	No	74	31.9	59	26.9
	Yes	158	68.1	160	73.1
At your house, do you have a telephone/mobile phone?	No	29	12.5	32	14.6
	Yes	203	87.5	187	85.4
At your house, do you have electricity?	No	155	66.8	168	76.7
	Yes	77	33.2	51	23.3
At your house, do you have a television?	No	131	56.5	140	63.9
	Yes	101	43.5	79	36.1
At your house, do you have a toilet inside the house?	No	207	89.2	202	92.2
	Yes	25	10.8	17	7.8
At your house, do you have a bicycle?	No	55	23.7	58	26.5
	Yes	177	76.3	161	73.5
At your house, do you have a motorcycle?	No	201	86.6	199	90.9
	Yes	31	13.4	20	9.1
At your house, do you have a car, truck, 4x4, or tractor?	No	198	85.3	199	90.9
	Yes	34	14.7	20	9.1

Table D.2: **SES Composite Frequencies** (continued)

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
How long does it take - on foot - to travel to a shopping area (or center) from your home?	Less than 20 munites	78	33.6	81	37.0
	More than 20 minutes but less than 1 hour	46	19.8	56	25.6
	More than 1 hour	86	37.1	56	25.6
	Don't know / No response	22	9.5	26	11.9
Last night, how much time did you spend on your household chores?	I don't do chores	14	6.0	19	8.7
	None	41	17.7	43	19.6
	Some	99	42.7	93	42.5
	A Lot	78	33.6	64	29.2
	Don't know / No response	0	0.0	0	0.0

Parental Literacy

The parental literacy composite included items related to the educational level and literacy skills of students' parents or guardians. The parental literacy composite was comprised of three items for a maximum composite score of 3.0. Average parental literacy composite scores for both intervention and comparison groups were high: 2.1 for the intervention group and 2.2 for the comparison group. This indicates that the adults in the students' homes have high levels of literacy.

Table D.3: **Parental Literacy Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
Can your mother read?	No	47	20.3	47	21.5
	Yes	185	79.7	172	78.5
	Don't know / No response	0	0.0	0	0.0
Can your father read?	No	24	10.3	26	11.9
	Yes	200	86.2	185	84.5
	Don't know / No response	8	3.4	8	3.7
What is the highest level of education your parents have achieved?	Diploma or above	10	4.3	7	3.2
	Post-Secondary	31	13.4	37	16.9
	Secondary	77	33.2	69	31.5
	Primary	47	20.3	42	19.2
	Don't know / No response	67	28.9	64	29.2

Parental Reading Support

The parental reading support composite was comprised of questions related to whether parents, guardians, or other adults in the home read with the students or provide them with support for reading. The composite included three items for a maximum composite score of 3.0. Relatively high averages were observed on the parental reading support composite. The average intervention group score was 2.2, and the average comparison group score was 2.1, indicating high levels of reading support outside of the classroom for all students.

Table D.4: **Parental Reading Support Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
Does someone at home read stories to you?	No	46	19.8	52	23.7
	Yes	186	80.2	165	75.3
	Don't know / No response	0	0.0	2	0.9
Does someone at home look at your school work?	No	45	19.4	32	14.6
	Yes	186	80.2	183	83.6
	Don't know / No response	1	0.4	4	1.8
Last week, how many days did you read with someone outside of school?	None	38	16.4	54	24.7
	All 7 days	59	25.4	41	18.7
	4-6 Days	35	15.1	43	19.6
	1-3 Days	72	31.0	46	21.0
	Don't know / No response	28	12.1	35	16.0

Reading Materials Access

The reading materials access composite included items about students' access to textbooks or reading materials in ciNyanja. The composite consisted of three items for a maximum composite score of 3.0. Average composite scores for reading materials access were 2.4 for both intervention and comparison groups, indicating that students have high levels of access to reading materials.

Table D.5: **Reading Materials Access Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
Do you have a textbook that helps you learn to read in ciNyanja?	No	78	33.6	67	30.6
	Yes	153	65.9	150	68.5
	Don't know / No response	1	0.4	2	0.9
Are the reading materials at school in ciNyanja?	No	21	9.1	16	7.3
	Yes	205	88.4	194	88.6
	Don't know / No response	6	2.6	9	4.1
Do you have a textbook that helps you learn math?	No	50	21.6	58	26.5
	Yes	182	78.4	159	72.6
	Don't know / No response	0	0.0	2	0.9

Teacher Reading Support

The teacher reading support composite included items on teachers' instruction and support provided for reading. The composite consisted of six items for a maximum composite score of 6.0. The average score for the teacher reading support composite was 4.7 for both the intervention and comparison groups. Students' responses indicate that teachers provide regular reading support in the classroom.

Table D.6: **Teacher Reading Support Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
How often does your teacher teach you to read each week?	Never	9	3.9	8	3.7
	Sometimes	83	35.8	72	32.9
	Every day	137	59.1	134	61.2
	Don't know / No response	3	1.3	5	2.3
How often does your teacher teach you to write in each week?	Never	14	6.0	23	10.5
	Sometimes	73	31.5	78	35.6
	Every day	144	62.1	117	53.4
	Don't know / No response	1	0.4	1	0.5
At school, do you get time to read silently by yourself?	Never	36	15.5	29	13.2
	Sometimes	104	44.8	90	41.1
	Every day	88	37.9	93	42.5
	Don't know / No response	4	1.7	7	3.2
At school, does your teacher ask you questions about what you are reading?	No	26	11.2	26	11.9
	Yes	205	88.4	190	86.8
	Don't know / No response	1	0.4	3	1.4
How often does your teacher help you when you are struggling with reading?	I don't think I need help.	0	0.0	1	0.5
	Never	10	4.3	11	5.0
	Sometimes	125	53.9	107	48.9
	Every time	94	40.5	94	42.9
	Don't know / No response	3	1.3	6	2.7
Does your teacher work with you to help you become a better reader?	No	14	6.0	9	4.1
	Yes	218	94.0	203	92.7
	Don't know / No response	0	0.0	7	3.2

Disposition to Reading

The disposition to reading composite included items related to students' attitudes toward reading in general. The composite consisted of three items for a maximum composite score of 3.0. The average composite score was 2.8 for both the intervention and comparison groups, indicating that all students had highly favorable attitudes toward reading.

Table D.7: **Disposition to Reading Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
Do you love, like, dislike, or hate reading?	Hate Reading	2	0.9	2	0.9
	Dislike Reading	6	2.6	9	4.1
	Like Reading	30	12.9	22	10.0
	Love Reading	189	81.5	177	80.8
	Don't know / No response	5	2.2	9	4.1
How do you feel when you are learning to read at school?	I don't like to read.	4	1.7	5	2.3
	I feel anxious.	22	9.5	25	11.4
	I feel confident.	198	85.3	173	79.0
	Don't know / No response	8	3.4	16	7.3
Reading is important to my future.	Undecided	1	0.4	8	3.7
	Disagree	2	0.9	2	0.9
	Agree	17	7.3	22	10.0
	Strongly Agree	212	91.4	187	85.4

Technology Use

The technology use composite was split in two subgroups. The first composite, called common items, contained eight items that were asked to students in both the intervention and comparison groups. Questions in technology use–common items related to students' current and previous technology use. A second technology use composite was constructed including the eight common items and three additional items related to technology use during the project. The second technology use composite with all items was reported for the intervention group only, as the three additional items were only asked to the intervention students.

The average score for both the intervention and comparison groups on the technology use–common items composite was 4.6 out of 8.0. This score indicates that usage of technology for learning did not vary across groups, and it also indicates that some—but not all—students had used technologies for learning in the past. The average composite score for the technology use–all items was 7.5 out of 11. This indicates that the students in the intervention group provided relatively positive feedback regarding their comfort with the technology and game, as well as preference for using technology to learn.

Table D.8: **Technology Use Composite Frequencies**

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
Do you use any technology to help you learn?	No	67	28.9	108	49.3
	Yes	157	67.7	106	48.4
	Don't know / No response	8	3.4	5	2.3
Have you used a phone to help you learn?	No	108	46.6	147	67.1
	Yes	124	53.4	72	32.9
	Don't know / No response	0	0.0	0	0.0
Have you used a computer to help you learn?	No	208	89.7	204	93.2
	Yes	24	10.3	15	6.8
	Don't know / No response	0	0.0	0	0.0
Have you used a tablet to help you learn?	No	229	98.7	217	99.1
	Yes	3	1.3	2	0.9
	Don't know / No response	0	0.0	0	0.0
You do not know if you have used technology to help you learn.	No	157	67.7	104	47.5
	Yes	75	32.3	115	52.5
	Don't know / No response	0	0.0	0	0.0
Have you used any of the following for learning to read?	Phone	156	67.2	77	35.2
	Tablet	4	1.7	1	0.5
	Computer	17	7.3	12	5.5
	Other	10	4.3	14	6.4
	Don't know / No response	45	19.4	115	52.5
If a mobile, have you used the phone to learn to read?	No	49	21.1	69	31.5
	Yes	152	65.5	65	29.7
	Don't know / No response	31	13.4	85	38.8
If yes, how often did you use the technology in the last week?	Every day	75	32.3	21	9.6
	Twice a week	38	16.4	26	11.9
	Once a week	39	16.8	20	9.1
	Not at all	13	5.6	24	11.0
	Don't know / No response	67	28.9	128	58.4

Table D.8: **Technology Use Composite Frequencies** (continued)

Composite Questions	Response Options	Group			
		Intervention		Comparison	
		n	%	*n*	%
How comfortable do you feel using the phone to play GraphoGame?	Very comfortable	201	86.6	-	-
	Somewhat comfortable	22	9.5	-	-
	Not comfortable	7	3.0	-	-
	Don't know / No response	2	0.9	-	-
Did participation in this program increase your comfort with learning to read through using GraphoGame?	No	3	1.3	-	-
	Yes	225	97.0	-	-
	Don't know / No response	4	1.7	-	-
Do you prefer learning to read using GraphoGame rather than just the classroom?	No	18	7.8	-	-
	Yes	213	91.8	-	-
	Don't know / No response	1	0.4	-	-

Engagement in Program

The engagement in program composite included a series of seven statements that were read to students in the intervention group. Students were asked to respond to the statements on a 4-point scale of agreement or difficulty, depending on the statement. The seven statements were used to create the engagement in program composite with the maximum score of 7.0. The average on the composite was 6.5, indicating that students in the intervention group were highly engaged in the project.

Table D.9: **Engagement on Program Composite Frequencies**

Composite Questions	Response Options	Intervention	
		n	%
Using the GraphoGame improved my reading.	Undecided	4	1.7
	Disagree	5	2.2
	Agree	27	11.6
	Strongly Agree	196	84.5
You want to continue using GraphoGame to learn to read.	Undecided	0	0.0
	Disagree	3	1.3
	Agree	18	7.8
	Strongly Agree	211	90.9
The things you've read this year with the help of GraphoGame were easy/difficult.	Undecided	4	1.7
	Always Hard	9	3.9
	Sometimes Hard	46	19.8
	Easy	173	74.6

Table D.9: **Engagement on Program Composite Frequencies** (continued)

Composite Questions	Response Options	Intervention	
		n	%
You like GraphoGame.	Undecided	0	0.0
	Disagree	2	0.9
	Agree	13	5.6
	Strongly Agree	217	93.5
You like the stories you've read this year.	Undecided	2	0.9
	Disagree	2	0.9
	Agree	41	17.7
	Strongly Agree	187	80.6
Using GraphoGame changed your attitude toward reading.	Negative Impact	7	3.0
	No Impact	49	21.1
	Positive Impact	176	75.9
Using GraphoGame increased your reading time overall.	Undecided	3	1.3
	Disagree	5	2.2
	Agree	36	15.5
	Strongly Agree	188	81.0
Do your parents know that you play GraphoGame?	No	8	3.4
	Yes	224	96.6
	Don't know / No response	0	0.0
Did not play due to bullying.	Undecided	218	94.0
	Always Hard	14	6.0
	Easy	0	0.0
Did not play due to teacher not liking it.	No	224	96.6
	Yes	8	3.4
	Don't know / No response	0	0.0
Did not play due to parents not liking it.	No	232	100.0
	Yes	0	0.0
	Don't know / No response	0	0.0

Annex E: **Descriptive Statistics**

Table E.1: **EGRA Results by Subtask and Group**[1]

EGRA Subtasks	Intervention[2]						Comparison[3]					
	Baseline		Endline		Gain		Baseline		Endline		Gain	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Orientation to print (correct out of three)	1.4	0.1	2.5	0.1	1.1*	0.1	1.6	0.1	2.3	0.1	0.7*	0.1
Letter sound identification (CLSPM)	3.6	0.3	11.1	0.5	7.4*	0.5	3.9	0.3	8.7	0.5	4.8*	0.4
Nonword reading (CNWPM)	0.3	0.1	3.0	0.4	2.5*	0.3	0.5	0.2	1.8	0.3	1.3*	0.3
ORF (CWPM)	0.4	0.1	4.1	0.5	3.3*	0.4	0.7	0.2	2.8	0.5	2.0*	0.4
Reading comprehension (correct out of five)	0.0	0.0	0.4	0.0	0.3*	0.0	0.1	0.0	0.2	0.0	0.2*	0.0
Listening comprehension (correct out of five)	2.9	0.1	3.2	0.1	0.4	0.1	2.8	0.1	3.0	0.1	0.3	0.1

* Statistically significant differences between baseline and endline at $p<0.05$.

1 Because of the differences in *n*, gain scores in this table may not be equal to the endline minus baseline mean scores.

2 Intervention Group – Baseline $n=222$, Endline $n=232$, Gain $n=222$

3 Comparison Group – Baseline $n=211$, Endline $n=219$, Gain $n=211$

Table E.2: **EGRA Results by Subtask, Group and Gender**

EGRA Subtasks	Period	Intervention[1]				Comparison[2]			
		Boys		Girls		Boys		Girls	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE
Orientation to print (correct out of three)	Baseline	1.4	0.1	1.4	0.1	1.6	0.1	1.6	0.1
	Endline	2.5	0.1	2.5	0.1	2.3	0.1	2.3	0.1
	Gain**	1.1	0.2	1.2	0.1	0.7	0.1	0.7	0.1
Letter sound identification (CLSPM)	Baseline	3.2	0.5	3.8	0.3	4.1	0.4	3.7	0.4
	Endline	11.7	0.9	10.8	0.6	8.7	0.7	8.6	0.7
	Gain**	8.1	0.8	7.0	0.5	4.6	0.6	4.9	0.6
Nonword reading (CNWPM)	Baseline	0.4	0.2	0.2	0.1	0.4	0.2	0.5	0.2
	Endline	3.9	0.7	2.5	0.4	2.1	0.4	1.7	0.5
	Gain*	2.9	0.6	2.2	0.4	1.7	0.4	1.1	0.3
ORF (CWPM)	Baseline	0.5	0.3	0.3	0.1	0.6	0.2	0.8	0.3
	Endline	4.9	0.9	3.6	0.6	3.1	0.7	2.5	0.6
	Gain*	3.6	0.7	3.1	0.5	2.6	0.6	1.5	0.5
Reading comprehension (correct out of five)	Baseline	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0
	Endline	0.4	0.1	0.3	0.1	0.3	0.1	0.2	0.1
	Gain*	0.3	0.1	0.3	0.1	0.2	0.1	0.1	0.0
Listening comprehension (correct out of five)	Baseline	2.9	0.1	2.8	0.1	2.7	0.1	2.8	0.1
	Endline	3.2	0.1	3.2	0.1	3.2	0.1	3.0	0.1
	Gain*	0.3	0.2	0.4	0.1	0.4	0.1	0.2	0.1

* Statistically significant difference in gain score from baseline to endline for girls at $p<0.05$.

** Statistically significant difference in gain score from baseline to endline for boys and girls at $p<0.05$.

1 Intervention Group – Boys: Baseline *n*=82, Endline *n*=88, Gain *n*=82; Intervention Group – Girls: Baseline *n*=140, Endline *n*=144, Gain *n*=140

2 Comparison Group – Boys: Baseline *n*=92, Endline *n*=96, Gain *n*=92; Comparison Group – Girls: Baseline *n*=119, Endline *n*=123, Gain *n*=119

Table E.3: **Percent of Students Receiving Zero Scores by Group and Gender at Baseline and Endline**

EGRA Subtasks	Period	Intervention[1]						Comparison[2]					
		Boys		Girls		Total		Boys		Girls		Total	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Orientation to print (correct out of three)	Baseline	27	32.9	49	35.0	76	34.2	23	25.0	33	27.7	56	26.5
	Endline	7	8.0	11	7.6	18	7.8	9	9.4	18	14.6	27	12.3
	Change	20	24.9	38	27.4	58	26.4	14	15.6	15	13.1	29	14.2
Letter sound identification (CLSPM)	Baseline	33	40.2	47	33.6	80	36.0	25	27.2	44	37.0	69	32.7
	Endline	14	15.9	19	13.2*	33	14.2*	13	13.5	28	22.8	41	18.7
	Change	19	24.3	28	20.4	47	21.8	12	13.7	16	14.2	28	14.0
Nonword reading (CNWPM)	Baseline	76	92.7	129	92.1	205	92.3	82	89.1	109	91.6	191	90.5
	Endline	55	62.5	103	71.5*	158	68.1*	70	72.9	103	83.7	173	79.0
	Change	25	30.2	29	20.6	54	24.2	15	16.2	9	7.9	24	11.5
ORF (CWPM)	Baseline	78	95.1	122	87.1	200	90.1	78	84.8	101	84.9	179	84.8
	Endline	52	59.1	95	66.0*	147	63.4*	68	70.8	96	78.0	164	74.9
	Change	26	36.0	27	21.1	53	26.7	10	14.0	5	6.9	15	9.9
Reading comprehension (correct out of five)	Baseline	79	96.3	136	97.1	215	96.8	87	94.6	114	95.8	201	95.3
	Endline	65	73.9	109	75.7*	174	75.0*	78	81.3	109	88.6	187	85.4
	Change	14	22.4	27	21.4	41	21.8	9	13.3	5	7.2	14	9.9
Listening comprehension (correct out of five)	Baseline	0	0.0	1	0.7	1	0.5	1	1.1	0	0.0	1	0.5
	Endline	0	0.0	0	0.0	0	0.0	0	0.0	2	1.6	2	0.9
	Change	0	0.0	1	0.7	1	0.5	1	1.1	-2	-1.6	-1	-0.4

* Indicates the percent of students receiving zero scores in the intervention group was significantly smaller than the percent of students receiving zero scores in the comparison group $p<0.05$.

1 Intervention Group – Boys: Baseline *n*=82, Endline *n*=88, Gain *n*=82; Intervention Group – Girls: Baseline *n*=140, Endline *n*=144, Gain *n*=140

2 Comparison Group – Boys: Baseline *n*=92, Endline *n*=96, Gain *n*=92; Comparison Group – Girls: Baseline *n*=119, Endline *n*=123, Gain *n*=119

Table E.4: **Percent of Students Receiving Zero Scores by Group and Gender at Baseline, Midline, and Endline**

EGRA Subtasks	Period	Intervention[1]						Comparison[2]					
		Boys		Girls		Total		Boys		Girls		Total	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Orientation to print (correct out of three)	Baseline	33	40.2	47	33.6	80	36.0	25	27.2	44	37.0	69	32.7
	Midline	11	13.3	7	5.3*	18	8.4*	11	13.8	18	18.2	29	16.2
	Endline	14	15.9	19	13.2*	33	14.2*	13	13.5	28	22.8	41	18.7
Letter sound identification (CLSPM)	Baseline	76	92.7	129	92.1	205	92.3	82	89.1	109	91.6	191	90.5
	Midline	60	72.3	99	75.6*	159	74.3*	63	78.8	86	86.9	149	83.2
	Endline	55	62.5	103	71.5*	158	68.1*	70	72.9	103	83.7	173	79.0
Nonword reading (CNWPM)	Baseline	78	95.1	122	87.1	200	90.1	78	84.8*	101	84.9	179	84.8
	Midline	56	67.5*	97	74.1	153	71.5*	66	82.5	82	82.8	148	82.7
	Endline	52	59.1	95	66.0*	147	63.4*	68	70.8	96	78.1	164	74.9
ORF (CWPM)	Baseline	79	96.3	136	97.1	215	96.9	87	94.6	114	95.8	201	95.3
	Midline	82	98.8	128	97.7	210	98.1	77	96.3	95	96.0	172	96.1
	Endline	65	73.9	109	75.7*	174	75.0*	78	81.3	109	88.6	187	85.4
Reading comprehension (correct out of five)	Baseline	0	0.0	1	0.7	1	0.5	1	1.1	0	0.0	1	0.5
	Midline	0	0.0	10	7.6	10	4.7	3	3.8	5	5.1	8	4.5
	Endline	0	0.0	0	0.0	0	0.0	0	0.0	2	1.6	2	0.9
Listening comprehension (correct out of five)	Baseline	0	0.0	1	0.7	1	0.5	1	1.1	0	0.0	1	0.5
	Midline	0	0.0	0	0.0	0	0.0	0	0.0	2	1.6	2	0.9
	Endline	0	0.0	1	0.7	1	0.5	1	1.1	-2	-1.6	-1	-0.4

* Indicates the percent of students receiving zero scores in the intervention group was significantly smaller than the percent of students receiving zero scores in the comparison group $p<0.05$.

1 Intervention Group – Boys: Baseline *n*=82, Midline *n*=83, Endline *n*=88; Intervention Group – Girls: Baseline *n*=140, Midline *n*=131, Endline *n*=144

2 Comparison Group – Boys: Baseline *n*=92, Midline *n*=80, Endline *n*=96, Gain *n*=92; Comparison Group – Girls: Baseline *n*=119, Midline *n*=99, Endline *n*=123

Table E.5: **Reading Comprehension Attempts by Group and Gender**

Period	Attempts	Intervention						Comparison					
		Boys		Girls		All Students		Boys		Girls		All Students	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Baseline	0	73	89.0	117	83.6	190	85.6	73	79.3	94	79.0	167	79.1
	1	6	7.3	19	13.6	25	11.3	15	16.3	19	16.0	34	16.1
	2	3	3.7	4	2.9	7	3.2	4	4.3	4	3.4	8	3.8
	3	0	0.0	0	0.0	0	0.0	0	0.0	1	0.8	1	0.5
	4	0	0.0	0	0.0	0	0.0	0	0.0	1	0.8	1	0.5
Endline	0	49	55.7	91	63.2	140	60.3	65	67.7	91	74.0	156	71.2
	1	22	25.0	26	18.1	48	20.7	20	20.8	17	13.8	37	16.9
	2	13	14.8	26	18.1	39	16.8	8	8.3	9	7.3	17	7.8
	3	2	2.3	1	0.7	3	1.3	2	2.1	4	3.3	6	2.7
	4	2	2.3	0	0.0	2	0.9	1	1.0	1	0.8	2	0.9
	5	0	0.0	0	0.0	0	0.0	0	0.0	1	0.8	1	0.5

Table E.6: **Intervention Group Correlations for Composite Scores and Subtasks**

Subtask	Language Consistency	SES	Parental Literacy	Parental Reading Support	Reading Materials Access	Teacher Reading Support	Disposition to Reading	Technology Use	Engagement in Program
Letter sound identification	0.096	0.066	0.091	0.048	0.072	0.103	0.199**	0.103	0.140*
Nonword reading	-0.039	-0.095	0.135	-0.040	0.086	-0.062	0.140*	-0.125	0.101

* Statistically significant differences at p<0.05

** Statistically significant differences at p<0.01

Table E.7: **Comparison Group Correlations for Composite Scores and Subtasks**

Subtask	Language Consistency	SES	Parental Literacy	Parental Reading Support	Reading Materials Access	Teacher Reading Support	Disposition to Reading	Technology Use
Letter sound identification	0.069	0.074	0.070	0.141	0.036	0.158*	0.114	0.079
Nonword reading	0.100	-0.019	0.064	0.185*	0.058	0.130	0.111	-0.080

* Statistically significant differences at p<0.05

Table E.8: **Simple Regression Analysis for Composites on Letter Sound Identification**

Variable	B	SE B	β
Group (Intervention/Comparison)	2.70	0.69	0.21**
Disposition to Reading	1.79	0.82	0.12*
Teacher Support for Reading	0.45	0.34	0.07
Parent Support for reading	0.49	0.47	0.06
R2		0.07	
F		6.60	

* Statistically significant differences at p<0.05

** Statistically significant differences at p<0.01

Table E.9: **Simple Regression Analysis for Composites on Nonword Reading**

Variable	B	SE B	β
Group (Intervention/Comparison)	1.17	0.46	0.14*
Disposition to Reading	0.99	0.54	0.10
Parent Support for reading	0.15	0.31	0.03
R2		0.03	
F		3.78	

* Statistically significant differences at p<0.05

Table E.10: **Average Composite Score by GG Dosage Threshold**

Composite	Below Threshold[1] (< 240 minutes)		Above Threshold[2] (240 minutes or more)		All Students[3]	
	Mean	SE	Mean	SE	Mean	SE
Language consistency	5.4	0.1	5.4	0.1	5.4	0.1
SES	4.3	0.2	4.5	0.1	4.5	0.1
Parental literacy	2.1	0.1	2.1	0.1	2.1	0.1
Parental reading support	2.2	0.1	2.2	0.1	2.2	0.1
Reading materials access	2.3	0.1	2.4	0.1	2.4	0.1
Teacher reading support*	4.3	0.1	4.9*	0.1	4.7	0.1
Disposition to reading	2.8	0.1	2.9	0.0	2.8	0.0
Technology use	7.2	0.2	7.7	0.1	7.5	0.1
Engagement in program*	6.3	0.1	6.6*	0.1	6.5	0.1

* Statistically significant differences at p<0.05

1 Below threshold *n*=71

2 Above threshold *n*=158

3 All students *n*=232

Table E.11: **EGRA Results by Subtask at Midline**

EGRA Subtasks	Intervention[1]						Comparison[2]					
	Boys		Girls		All Students		Boys		Girls		All Students	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Letter sound identification (CLSPM)	8.5	0.8	10.9	0.6	10.0	0.5	8.6	8.8	8.9	0.8	9.0	0.7
Nonword reading (CNWPM)	2.2	0.5	1.9	0.4	2.0	0.3	1.5	8.4	1.4	0.4	1.4	0.3
ORF (CWPM)	2.6	0.6	1.9	0.4	2.2	0.3	1.8	8.6	1.8	0.6	1.8	0.4
Reading comprehension (correct out of five)	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.1	0.4
Listening comprehension (correct out of five)	2.7	0.1	2.5	0.1	2.6	0.09	2.7	8.1	2.7	0.1	2.7	0.1

1 Intervention Group – Boys *n*=83, Girls *n*=131, All Students *N*=214

2 Comparison Group – Boys *n*=80, Girls *n*=99, All Students *N*=179

Annex F: **GG Descriptive Statistics**

Table F.1: **GraphoGame Playing Time Descriptive Statistics by School**

School Name	*n*	Time of GG play (in seconds)				Time of GG play (in minutes)				Days of GG play			
		Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
School "A"	8	32141.3	2824.9	28867.7	37724.9	535.7	47.1	481.1	628.7	46.5	2.9	43.0	51.0
School "B"	15	21257.3	3864.2	14387.1	26389	354.3	64.4	239.8	439.8	33.7	5.4	24.0	40.0
School "C"	7	37042.4	3153.5	33789.6	40636.3	617.4	52.6	563.2	677.3	31.6	1.9	29.0	34.0
School "D"	13	19059	1989.6	15368.7	23171.8	317.7	33.2	256.1	386.2	28.4	2.2	25.0	32.0
School "E"	30	20056.4	2838.7	14038	24586.7	334.3	47.3	234	409.8	27.4	3	21.0	33.0
School "F"	20	25475.7	6467.3	16464.8	34897.9	424.6	107.8	274.4	581.6	26.4	5.6	16.0	33.0
School "G"	17	15401.8	5572.3	252.9	22015.2	256.7	92.9	4.2	366.9	25	8	1.0	32.0
School "H"	7	17141.3	1999.5	14863.2	20018.2	285.7	33.3	247.7	333.6	24	2.2	22.0	27.0
School "I"	7	12904.8	2052.7	9307.4	15904.2	215.1	34.2	155.1	265.1	23	4.6	16.0	31.0
School "J"	25	16162.6	4703.6	7626.4	28932	269.4	78.4	127.1	482.2	22.5	4.3	14.0	30.0
School "K"	15	19520.8	3507.8	11767.6	24605.5	325.3	58.5	196.1	410.1	22	3.3	16.0	29.0
School "L"	8	13696.6	6999.9	612.9	21407.5	228.3	116.7	10.2	356.8	20.9	9.7	2.0	30.0
School "M"	27	15931.3	3985.6	9910.5	28164.2	265.5	66.4	165.2	469.4	17	4.2	10.0	27.0
School "N"	8	4440.2	3778.8	276.6	10359.6	74.0	63.0	4.6	172.7	10.9	7.6	2.0	20.0
School "O"	25	5710.7	2561.4	853.7	10261	95.2	42.7	14.2	171	10.5	3.4	4.0	15.0
All schools	232	17480.0	7876.1	252.9	40636.3	291.3	131.3	4.2	677.3	23.5	9.1	1.0	51.0

Table F.2: **GG Assessment Results by Group**[1]

GG Subtask	Period	Intervention[2]						Comparison[3]					
		Boys		Girls		All Students		Boys		Girls		All Students	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Letter sound	Baseline	3.8	0.3	3.8	0.2	3.8	0.2	3.6	0.4	3.9	0.3	3.8	0.2
	Midline	11.4	0.6	12.0	0.5	11.8	0.4	5.9	0.5	5.5	0.4	5.7	0.3
	Endline	13.5	0.6	13.4	0.5	13.4	0.4	6.4	0.5	6.1	0.4	6.2	0.3
Word recognition	Baseline	2.2	0.3	1.9	0.2	2.0	0.2	2.5	0.3	2.5	0.3	2.5	0.2
	Midline	7.8	0.9	7.4	0.7	7.5	0.5	4.3	0.7	3.7	0.5	4.0	0.4
	Endline	9.9	0.9	8.3	0.7	8.9	0.6	5.4	0.7	4.3	0.5	4.8	0.4

1 Because of the differences in *n*, gain scores in this table may not be equal to the endline minus baseline mean scores.

2 Intervention Group – Boys: Baseline *n*=82, Midline *n*=83, Endline *n*=88; Intervention Group – Girls: Baseline *n*=140, Midline *n*=131, Endline *n*=144

3 Comparison Group – Boys: Baseline *n*=92, Midline *n*=80, Endline *n*=96, Gain *n*=92; Comparison Group – Girls: Baseline *n*=119, Midline *n*=99, Endline *n*=123

Annex G: **EGRA Reliability Results**

Table G.1: **Reliability Results for Midline EGRA**

Subtask	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Letter sound identification (percent correct)	0.449	0.476
Nonword reading (percent correct)	0.600	0.404
ORF (percent correct)	0.599	0.409
Reading comprehension (percent correct)	0.426	0.521
Listening comprehension (percent correct)	0.234	0.845
	EGRA Coefficient Alpha	**0.556**

Table G.2: **Reliability Results for Endline EGRA**

Subtask	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Orientation to print (percent correct)	0.296	0.722
Letter sound identification (percent correct)	0.567	0.604
Nonword reading (percent correct)	0.640	0.568
ORF (percent correct)	0.585	0.528
Reading comprehension (percent correct)	0.571	0.556
Listening comprehension (percent correct)	0.241	0.659
	EGRA Coefficient Alpha	**0.645**

Annex H: **Additional Results**

Table H.1: **EGRA Results by Classroom Observation Composite Score**

Subtask	Below Average (> 19.3)	Above Average (19.3 or above)
Letter sound identification (CLSPM)	7.1	6.9
Nonword reading (CNWPM)	2.1	3.3
ORF (CWPM)	2.4	4.8

Table H.2: **Teacher Questionnaire Results**

Questions	Response Options	Group							
		Intervention				Comparison			
		Baseline		Endline		Baseline		Endline	
		n	%	*n*	%	*n*	%	*n*	%
How well are the current teaching methods addressing the reading problems of learners in your school?	Very well	5	29.4	10	62.5	7	36.8	4	26.7
	Relatively well	11	64.7	5	31.3	10	52.6	10	66.7
	Not sure	1	5.9	1	6.3	2	10.5	0	0.0
	Not at all	0	0.0	0	0.0	0	0.0	1	6.7
How much time is allocated to the teaching of literacy in your school time table per week? Indicate number of hours.	Hours	18	4.9	16	5.1	19	5.0	15	4.6
Have you ever used the internet before?	Yes	13	72.2	15	93.8	14	73.7	11	73.3
	No	5	27.8	1	6.3	5	26.3	4	26.7
Have you ever used an ICT related technology for educational purposes, e.g. preparing lessons for teaching?	Yes	9	50.0	12	75.0	7	36.8	11	73.3
	No	9	50.0	4	25.0	12	63.2	4	26.7
Has your school organized any training for the teachers in the use of computers?	Yes	6	33.3	8	50.0	6	33.3	3	20.0
	No	12	66.7	8	50.0	12	66.7	11	73.3
	No response	0	0.0	0	0	0	0.0	1	6.7