USAID IMPACT EVALUATION OF THE *MAKHALIDWE ATHU* PROJECT (ZAMBIA)

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

ENDLINE REPORT (FINAL)

USAID IMPACT EVALUATION OF THE *MAKHALIDWE ATHU* PROJECT (ZAMBIA)

ENDLINE REPORT (FINAL)

EXECUTIVE SUMMARY

EVALUATION PURPOSE

Using mobile devices as a channel to deliver educational materials in low income countries constitutes a promising area for education policy. However, the evidence on the effectiveness of such types of programs, let alone their cost-effectiveness, is still scarce. Recognizing this, USAID on behalf of the All Children Reading: A Grand Challenge for Development Partners (USAID, World Vision, and the Australian Government), engaged NORC under the Reading and Access contract to evaluate the impact of the Makhalidwe Athu project (MA), a pilot intervention aimed at improving the reading skills of young children in Zambia's Eastern province, using cell phones as a means to provide reading materials. In this report, we describe the main impact evaluation results, and discuss recommendations in terms of policy implementation and future research.

PROJECT BACKGROUND

The MA project was a nine-month pilot intervention aimed at improving the reading skills of 1,200 students in 2nd and 3rd grade in the Chipata and Lundazi districts of Zambia's Eastern province. The project, funded by the All Children Reading: A Grand Challenge for Development Partners (USAID, World Vision and the Australian Government), and designed and implemented by Creative Associates, provided reading materials in ChiNyanja (the predominant local language) and supported reading activities through SMS text messaging.
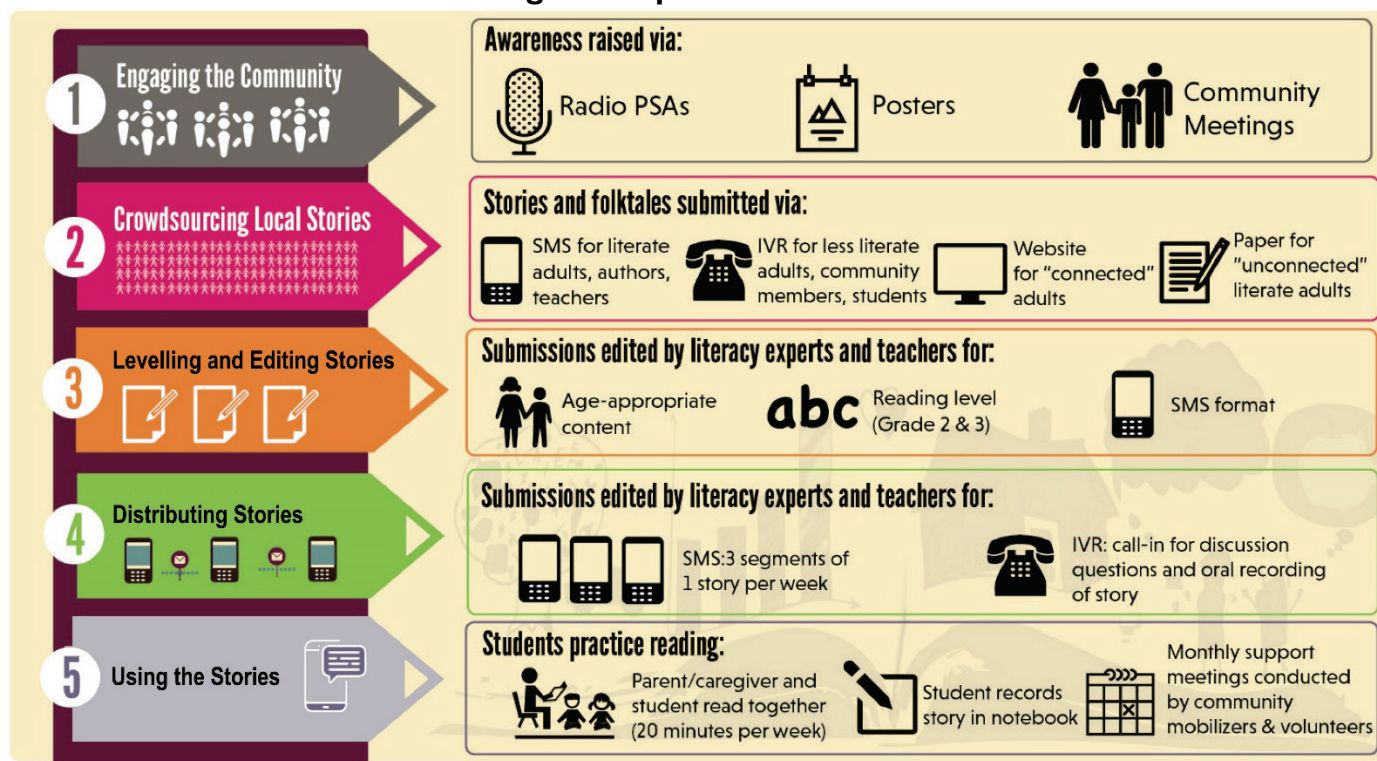
The MA project developed and provided short stories for 2nd and 3rd graders, using basic household cell phones. **Over a 9-month period, participant households received three SMS on their mobile phones each week. These three messages comprised a short story (e.g. 160 characters each) for children to read with their families, as well as a question about the story.** In addition, participants could call in for a pre-paid recorded voice message (IVR), which included comprehension questions, and a recording of the story itself. The underlying theory of change of the program was that by providing reading materials via caregiver phones that were culturally relevant and entertaining for children, they were going to read more and become better readers.

Creative Associates also conducted monthly meetings with participant parents/caregivers to guide them on how to read and listen to the stories with their children. These meetings lasted about three hours each and covered issues with the SMS transmissions and how to troubleshoot, as well as how to work with children on the stories, which included demonstrations with both parents and children. Creative also provided children with a notebook and encouraged them to record the stories.[1]

---

[1] Parents were asked in the monthly meetings to encourage their children to transcribe the stories, and the children themselves were encouraged as well in the monthly meetings.

Program Implementation[2]

Engaging the Community
1

Awareness raised via:
Radio PSAs
Posters
Community Meetings

Crowdsourcing Local Stories
2

Stories and folktales submitted via:
SMS for literate adults, authors, teachers
IVR for less literate adults, community members, students
Website for "connected" adults
Paper for "unconnected" literate adults

Levelling and Editing Stories
3

Submissions edited by literacy experts and teachers for:
Age-appropriate content
abc Reading level (Grade 2 & 3)
SMS format

Distributing Stories
4

Submissions edited by literacy experts and teachers for:
SMS: 3 segments of 1 story per week
IVR: call-in for discussion questions and oral recording of story

Using the Stories
5

Students practice reading:
Parent/caregiver and student read together (20 minutes per week)
Student records story in notebook
Monthly support meetings conducted by community mobilizers & volunteers

According to the survey data, program take-up was high. Ninety-three percent of households in the treatment group attended at least one MA meeting, and 81.4 percent showed SMS to the MA child at least once a week.

EVALUATION QUESTIONS, DESIGN, METHODS AND LIMITATIONS

The purpose of the evaluation was to measure the impact of MA on the reading habits and skills of participating students. It specifically sought to address the following research questions:

1. Did the reading skills of Grade 2 and 3 students receiving the treatment improve as a result of the MA intervention? What are the magnitudes of these improvements, and in which reading domains have skills improved?

2. How and to what extent have student attitudes toward reading changed as a result of MA? Are students enjoying reading at home? Are they more likely to participate in out-of-school reading activities in the community as a result of MA?

3. Do parents spend more time supporting their children's reading activities as a result of MA? How much time are they spending on reading activities as a result of MA?

4. Do students spend more time reading at home on their own as a result of MA? How much time are they spending on reading activities?

5. Are there any spillover effects of MA? Are other children in the household participating in the MA/SMS reading activities?

6. What is the cost-effectiveness of the program?

---

[2] Graphic prepared by Creative Associates International.

To evaluate the impact of MA, we conducted an experiment where we randomized schools into treatment and control groups, and surveyed samples of students in each school, as well as their caregivers.

NORC and the Institute of Economic and Social Research at the University of Zambia (INESOR) collected all the data used in the study. Baseline data was collected between November 2015 and January 2016. Three instruments were fielded at baseline: a caregiver questionnaire, a student questionnaire, and a reading assessment. The reading assessment was a version of the Early Grade Reading Assessment (EGRA). In January 2017, we surveyed the same caregivers and students as at baseline, in order to construct a panel. The same instruments were used at endline. Caregiver and student endline surveys were modified to incorporate program uptake questions. In addition, in June 2016 a survey on a subsample of treatment caregivers was fielded to document program uptake.

In total 1,942 caregivers were surveyed at baseline and endline, 965 in the treatment group and 977 in the control group. These 1,942 caregivers correspond to 2,091 children (there are more students than caregivers because 149 siblings were also interviewed). Of these 2,091 children for which there is caregiver data, 2,054 were interviewed for the EGRA and student survey at endline, 1,021 in the treatment group and 1,033 in the control group.

Endline Sample of Caregivers and Students

	Treatment	Control
Caregivers	965	977
Student	1,021	1,033

Source: Own calculations using MA baseline and endline data.

FINDINGS AND CONCLUSIONS

The MA project had a positive impact on three of the five EGRA subtasks evaluated. The table below shows the mean score for each EGRA subtask at baseline and endline, by treatment status. To estimate the impact of MA for each EGRA subtask, we compare the change between baseline and endline in the treatment group to the change between baseline and endline for the control group. These difference-in-difference estimates and effect sizes are also displayed in the table. **We found positive and significant impacts for non-word reading, oral reading fluency (ORF) and reading comprehension.**

EGRA scores at baseline and endline by experimental group

	Baseline			Endline			Diff-in-Diff	
	Con-trol (1)	Treat-ment (2)	Diff (3=2-1)	Con-trol (4)	Treat-ment (5)	Diff (6=5-4)	DID (7=6-3)	Eff Size[a] (8)
Letter Sounds (total correct)	6.9	8.6	1.7	9.5	10.9	1.4	-0.3	-0.03
Non Words (total correct)	2.2	3.2	1	4.5	7.0	2.5***	1.5***	0.20
Oral Reading Fluency	3.5	4.9	1.4	7.2	12.0	4.8***	3.4***	0.27
Reading comp. (total correct)	0.3	0.4	0.1	0.7	1.1	0.4***	0.3***	0.23
Listening comp. (total correct)	2.8	2.9	0.1	3.0	3.1	0.1	0	-0.05

[a]The effect size is calculated as the DID over the pooled standard deviation at endline.
Source: Own calculations using MA baseline and endline data. All observations recorded at baseline and endline are included, even those for which the panel is not balanced. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

No effects were found for letter sound identification or listening comprehension. Perhaps it is not very surprising that no effects were found for letter sound identification as MA was not specifically providing materials to teach letter sounds. On the other hand, the fact that no impacts were found for listening comprehension is a little surprising given that the voice messages that were available could presumably had the potential to enhance children's listening comprehension.

The table below shows effect sizes for the different EGRA subtasks. For non-word reading, ORF, and reading comprehension, where we found significant impact, these effect sizes are, respectively 0.2, 0.27 and 0.23.

To put these effects in perspective, note that the treatment group showed double the improvement compared to the control group for the same period. In effect, the control group increased ORF by 3.6 wpm between baseline and endline, while the increase in the treatment group was 7.1 wpm.

Gains in EGRA scores between baseline and endline by treatment group
Mean and 95 percent confidence interval
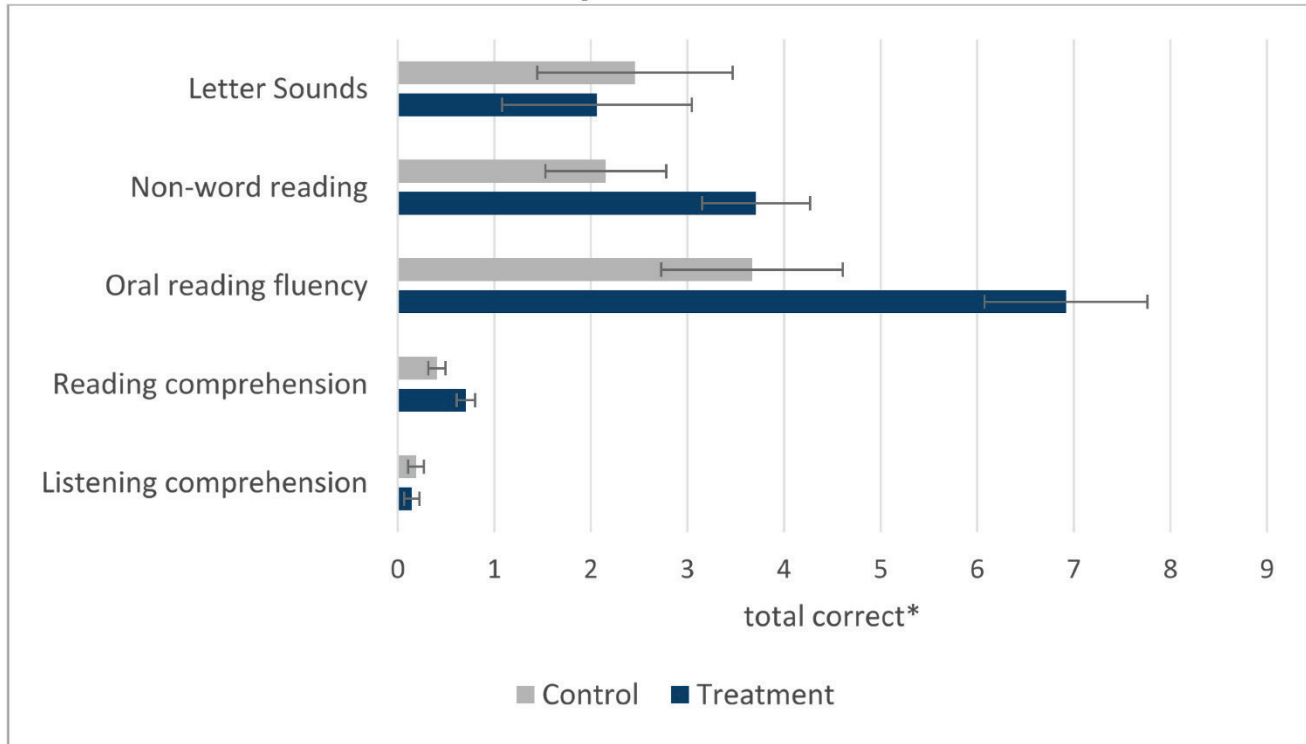
*For Oral reading fluency the units are wpm

In terms of treatment impacts by subgroups we found that students in 3rd grade benefitted more from the program than students in 2nd grade. We also found that program impact estimates are higher for girls than for boys, but in this case the differences are not substantial.

In addition to the analysis on the means, we also analyze the fraction of students that reached certain meaningful reading thresholds. We follow performance standards adopted by the government (Ministry of Education, Science, Vocational Training and Early Education, 2014). For 2nd grade these standards define 'below minimum' as reading fewer than 25 wpm, 'minimum' as reading between 25 and 39 wpm, 'desirable' as reading between 40 and 59 wpm, and 'outstanding' as reading 60 wpm or more. For 3rd grade 'below minimum' is reading fewer than 40 wpm, 'minimum' is defined as reading between 40 and 59 wpm, 'desirable' as reading between 60 and 69 wpm, and 'outstanding' as reading 70 wpm or more.

The figure below shows the fractions of students in treatment and control groups by these reading levels for students in grades 2 and 3, separately. Given the large fractions of students that cannot read one single word, the graphs also display the fractions of zero-readers. Panel A shows results for students that were in 2nd grade in 2016. The fraction of 2nd graders in the control group not able to read a single word declined from 80 percent at baseline to 64 percent at endline. This decline was counterbalanced mostly by increases in the fractions of 'below minimum' and 'minimum' readers between baseline and endline. A similar dynamic can be observed for the treatment group although the changes were more pronounced as a result of the program. **In effect while the decline in the fraction of students that were not able to read a single word in the control group was 16 percentage points, the decline for the treatment group was 21 percentage points.** The

decline in the number of students that cannot read a single word in the treatment group was mostly counterbalanced by increases in 'below minimum' and 'minimum' readers.

Panel B shows results for 3rd graders. Students in 3rd grade also observed major decreases in the fractions of zero-readers between baseline at endline. **In the control group the decline in the fraction of zero-readers was 14 percentage points, while for the treatment group was 25 percentage points.** These declines were compensated, in both treatment and control groups, mostly by increases in the fractions of 'below minimum' students.

Students in treatment and control groups by reading level – Oral Reading Fluency
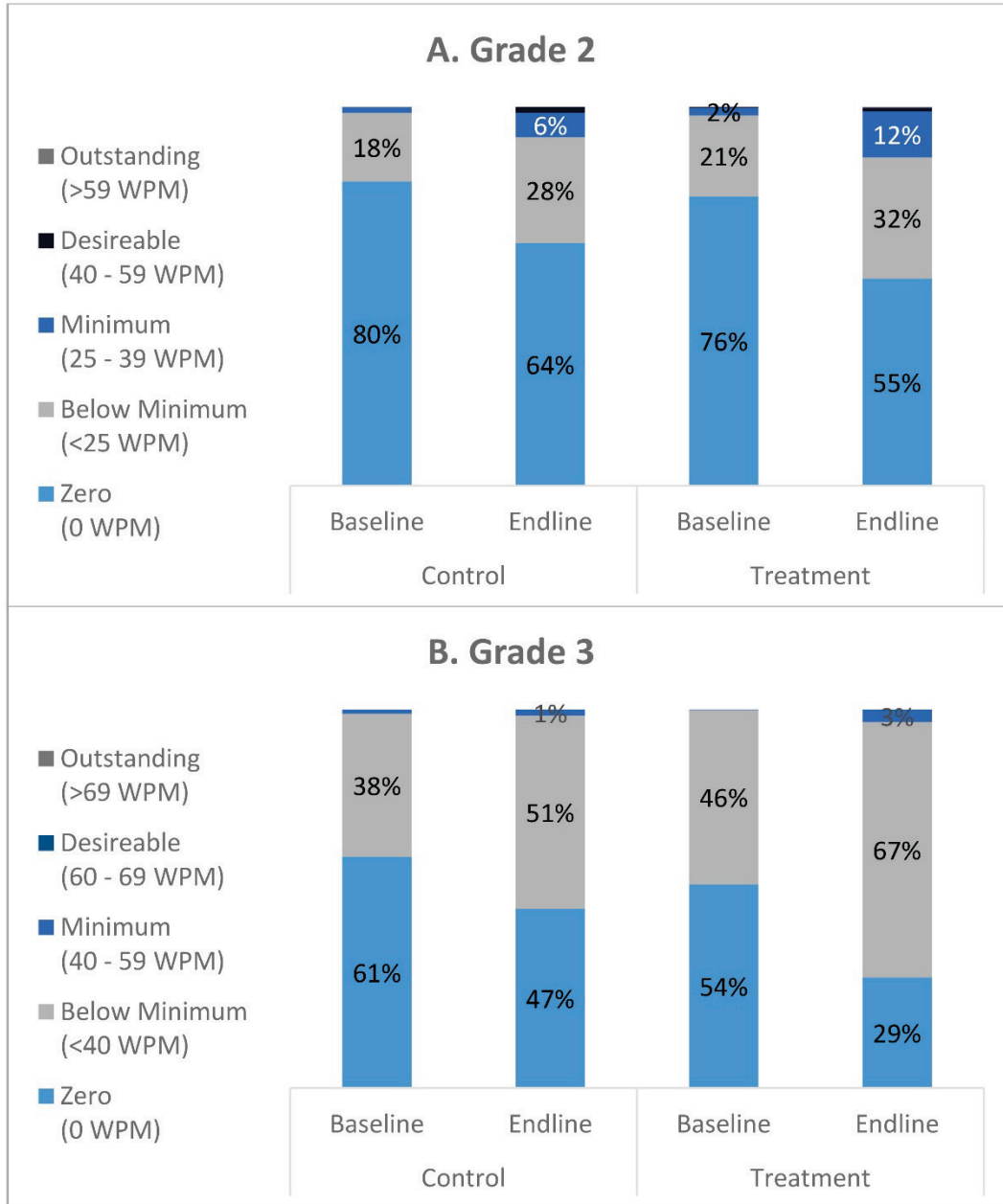
Note: Fraction labels of 1 percent or less are dropped from the graph for clarity.
Source: Own calculations using MA baseline and endline data. Performance levels are from Ministry of Education, Science, Vocational Training and Early Education (2014).

We did not find any evidence that the program changed children's attitudes towards reading, measured as whether they enjoy reading in different contexts. Note that most children in both treatment and control groups reported that they like reading, so the program had little room to improve this outcome. We also did not find evidence that the program changed the likelihood that children participate in out-of-school reading activities. In this case it would seem that the driving factor is the actual lack of out-of-school activities available in the communities.

To analyze the impact of the program on whether children read more with their parents, we asked both caregivers and children about reading habits. According to the caregiver responses, the program had a positive impact on the frequency with which children read with their parents; however, when we looked at the children's responses on how often they read with their parents, no statistically significant change was observed; therefore, **the results we found for time spent reading with a parent are inconclusive**.

We found positive impacts on the frequency with which children read on their own at home. This result was observed in both caregivers' and children's responses.

We also found that other children in participating households join the MA children when reading the SMS stories, suggesting that the program may have spillover effects on reading habits and reading skills of other children in the household.

If the program is scaled-up at the national level, we estimate that the final costs per student will be US 20.1 for students in Lusaka or Eastern Province, and US 21.6 for students in other parts of the country. The difference between these two estimates is explained by the fact that while for ChiNyanja (the main local language in Lusaka or Eastern Province) a bank of stories has already been constructed as a result of the MA pilot, for the rest of the country (where local languages other than ChiNyanja are more predominant), new stories will need to be developed.

Limitations. Threats to the external validity of the documented results are two-fold. First, the student sample frame was constructed from the group of children that were in school on the day the Advanced Teams visited schools. Twenty-seven percent of enrolled students were not in school during those visits. The estimated effects cannot be extrapolated to students that were not in school the day Advanced Teams were there constructing the student sampling frames.

Second, as is the case with any program fielded in a small area such as the districts of Chipata and Lundazi (relative to the whole country), the documented results cannot be automatically extrapolated to other parts of the country, or to the group of schools in these two districts that were not eligible for participation in MA. Further research should address how similar these two districts are to the rest of the country in terms of the conditions relevant to the program and the outcomes it affects, including households' socioeconomic characteristics, school enrollment, cell phone ownership and service quality, and spoken language.

Another important consideration is that the documented impact results correspond to offering the full set of services that were provided in the context of the program, namely the SMS messages combined with monthly meetings, the radio campaign, and the notebooks that were provided to MA children. As such, when computing the costs of the program we include the costs of this entire package of interventions.

RECOMMENDATIONS

The main purpose of this evaluation was to test an innovative approach to delivering reading materials. Given the positive findings on the project's impact, consideration should be given to scaling up the program.

However, any scale-up decision must take into consideration the answers to the following questions:

(1) **How feasible is it to replicate the MA program in its entirety?** MA consisted not only of the delivery of SMS, which alone would be relatively easy to replicate in areas of the country with similar cultural and socioeconomic background (and cell phone coverage); it also consisted of monthly meetings to raise awareness and guide parents/caregivers on what their roles and responsibilities were; awareness-raising through media; and a supporting voice-message system. These activities added to the cost of the program. The cost effectiveness analysis documented in this report should be used in this context when considering scaling up the program.

(2) **Are the gains in reading sustainable?** The endline data were collected just after the implementer stopped sending messages and conducting the monthly meetings with caregivers, so it would be useful to document whether these gains in EGRA scores and reading habits are maintained over time even in the absence of the program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ACR	All Children Reading
AT	Advanced Team
DEBS	District Education Board Secretary
EMIS	Education Management Information System
EGRA	Early Grade Reading Assessment
IE	Impact Evaluation
INESOR	Institute of Economic and Social Research
IRB	Institutional Review Board
IRR	Inter-rater reliability
ITT	Intent-to-Treat
MA	*Makhalidwe Athu*
MOGE	Ministry of General Education
NORC	NORC at the University of Chicago
ORF	Oral Reading Fluency
RCT	Randomized Control Trial
RQ	Research Question
RTS	Read To Success project
SMS	Short Message Service
STS	School-to-School International
TOT	Treatment-on-the-Treated
USAID	United States Agency for International Development
VAM	Value-Added Model
wpm	Words per minute

ACKNOWLEDGEMENTS

AUTHORS

This report was produced by Principal Investigator Dr. Alejandro Ome, Senior Evaluation Advisor Dr. Alicia Menendez, Project Manager Varuni Dayaratna, Research Analysts Russell Owen, Carlos Fierros and Aaron Wilson.

I.   EVALUATION PURPOSE

In the past few years, the use of information communication technology (ICT) for education has been gaining popularity. Using mobile devices like cell phones and tablets as a channel to deliver educational materials in low income countries constitutes a promising area for education policy. The appeal of these devices is that they can be used to transmit (or carry) large amounts of information at a relatively low cost. However, the evidence on the effectiveness of such types of programs, let alone their cost-effectiveness, is still scarce (Wagner, et al., 2014). Recognizing this, under the Reading and Access contract, USAID, on behalf of ACR GCD partners, engaged NORC to evaluate the impact of the Makhalidwe Athu project (MA), a pilot intervention aimed at improving the reading skills of young children in Zambia's Eastern province, using cell phones as an outlet to provide reading materials. In this report, we describe the impact evaluation results of the program, and discuss recommendations in terms of policy implementation and future research.

The purpose of this evaluation was to measure the impact of the MA project on the reading habits and skills of students treated by the program, as well as to estimate the cost-effectiveness of the intervention. This report discusses the impact of the program on students' attitudes towards reading, time students spend reading alone at home, time students spend reading supported by other family members, and students' reading test scores. It also discusses cost scenarios that could be considered for future replications of the program, and the associated gains in terms of reading skills.

The MA project was conceived as pilot program and this evaluation was designed to inform the benefits and costs of the program. Hence, the ultimate purpose of this evaluation is to provide policymakers, in particular the Zambian government, and aid organizations, the inputs required to make an informed decision on replicating or scaling up this program. More broadly, this study is also intended to contribute to the literature on the use of ICT in education policy.

II. PROJECT BACKGROUND

The *Makhalidwe Athu* (MA) project was a nine-month intervention aimed at improving the reading skills of 1,200 students in 2nd and 3rd grade in the Chipata and Lundazi districts of Zambia's Eastern province. The project, funded by the All Children Reading Partners (USAID, World Vision and the Australian Government), and designed and implemented by Creative Associates, developed reading materials in ChiNyanja (the predominant local language) and supported reading activities through SMS text messaging.

Between April and December 2016, participant households received three text messages on their mobile phones each week, transmitted on Mondays, Wednesdays and Fridays. These three messages comprised a short story (e.g. 160 characters each) for children to read with their families, as well as a question about the story.

Participants could also call in for a pre-paid recorded voice message (IVR), which included comprehension questions, and a recording of the story itself. The underlying theory of change

of the program was that by providing reading materials that were culturally relevant and entertaining for children, they were likely to read more and become better readers.

Stories were "crowdsourced" from the local communities that learned about the program through local media and community meetings. Creative's local reading specialist conducted workshops with MOGE staff and teachers to "level" the stories for Grade 2 and 3 readers and to fit the SMS format, making them both culturally relevant and in line with the school curriculum. A total of 267 stories were levelled, including 216 crowdsourced stories and 51 stories developed under another project conducted by Creative in Zambia (Read to Succeed).

This was five times the target of 52 stories the project had, and reflects the high level of engagement MA raised in the community (Creative Associates, 2017). Moreover, parents and teachers of treatment schools were major contributors to this crowdsourcing effort (Creative Associates, 2017), indicating that the program was able to engage key actors in the human capital development of the children targeted by the program.

Figure I.   Program Implementation[3]
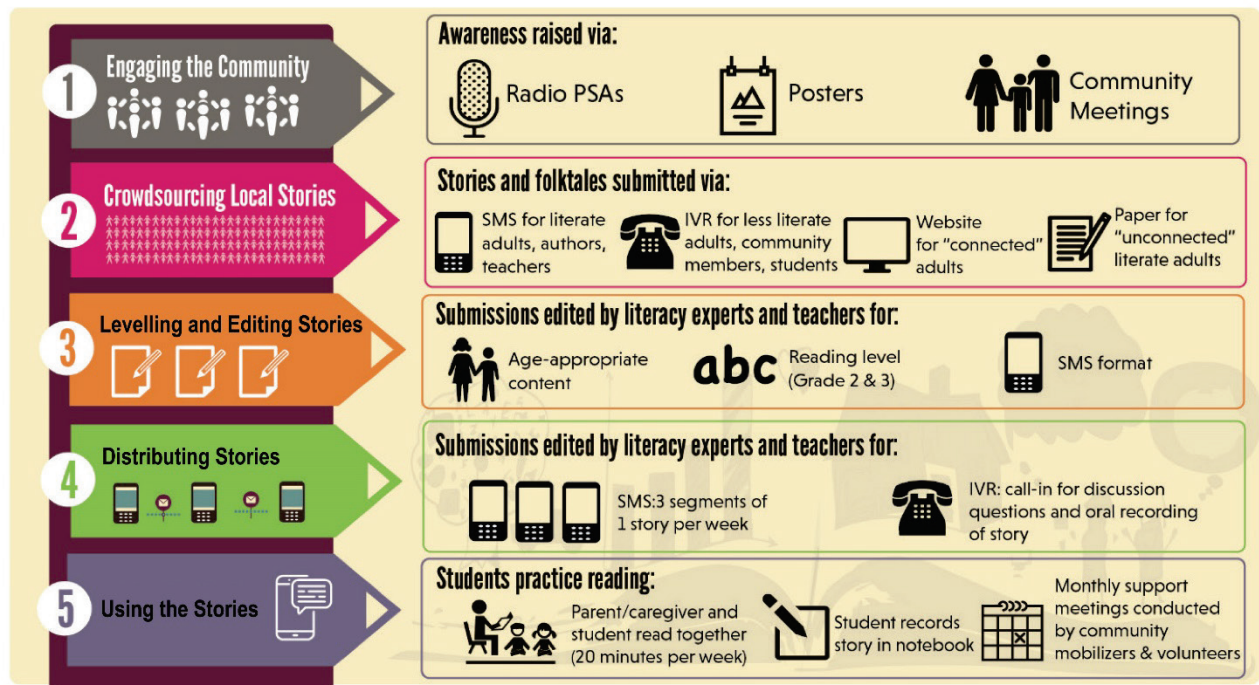
Following an initial orientation meeting with caregivers about the program and how to access the SMS stories and IVR, community mobilizers, aided by community volunteers, met monthly with caregivers to discuss and address any issues that had arisen in the context of the program. These meetings lasted about three hours and covered issues with the SMS transmissions and how to troubleshoot, as well as how to work with children on the stories, which included

---

[3] Graphic prepared by Creative Associates International.

demonstrations with both parents and children. Creative also provided children with a notebook and encouraged them to record the stories.[4]

Although unplanned and not part of the MA ACR program design, the local radio station BreezeFM independently started to broadcast a "variety show" midway through the program on reading, featuring the story of the week, as well as call-in questions and presentations by local experts.

To our knowledge the program was implemented without major difficulties.[5] An uptake survey conducted on a subsample of caregivers in the treatment group in June 2016 showed that participation in the program was high. Ninety-five percent of respondents said that they had attended at least one MA meeting. Most households reported they were receiving SMS (92 percent) and of these, all caregivers said they read SMS with the child at least once a week. Also, 68 percent of respondents that received text messages said they listened to the voice recording with their child.

The endline data reflects a high level of program uptake as well. Ninety-three percent of caregivers in the treatment group reported they had attended at least one MA meeting or training session; these respondents attended an average of 7 meetings over the life of the program.[6] This can be considered in itself an important result of the project, as school officials in these communities in general struggle to convince parents to attend school-related meetings (Creative Associates, 2017). Eighty-four percent of caregivers in the treatment group said they received stories via SMS. Eighty-one percent of respondents in the treatment group showed SMS to the child at least once a week. Seventy percent of respondents that received text messages said they listened to the voice recording with their child.

Moreover, informal observations by program's staff indicate that MA was well received by the community. For example, parents from not selected school communities called BreezeFM and expressed their interest in having their children enrolled in the program. Also, during the program close out event, attended by both the DEBS from Chipata and Lundazi, teachers and head teachers, PTA members, parents, students, and other members of the community, DEBS expressed their gratitude to the project staff and indicated that, given the success they had observed of the project, they will introduced the MA model of short stories accompanied by comprehension questions to be used in class in all schools in their districts (Creative Associates, 2017).

---

[4] Parents were asked in the monthly meetings to encourage their children to transcribe the stories, and the children themselves were encouraged as well in the monthly meetings.

[5] The program was originally scheduled to start in January 2016 but it was delayed to give to room for baseline data collection. Technical difficulties with the IVR also delayed program onset.

[6] Caregiver statistics are calculated at the child level; hence caregivers of more than one children in the survey are overrepresented. We could present results for caregivers at the caregiver level but we choose to do it this way as it reflects the results in terms of the number of interactions caregiver-child.

III.   EVALUATION METHODS & LIMITATIONS

III.1. EVALUATION QUESTIONS

The impact evaluation of MA seeks to address the following research questions:

1. Have the reading skills of Grade 2 and 3 students receiving the treatment improved as a result of the MA intervention? What are the magnitudes of these improvements, and in which reading domains have skills improved?

2. How and to what extent have student attitudes toward reading changed as a result of MA? Are students enjoying reading at home? Are they more likely to participate in out of school reading activities in the community as a result of MA?

3. Do parents spend more time supporting their children's reading activities as a result of MA? How much time are they spending on reading activities as a result of MA?

4. Do students spend more time reading at home on their own as a result of MA? How much time are they spending on reading activities?

5. Are there any spillover effects of MA? Are other children in the household participating in the MA/SMS reading activities?

6. What is the cost-effectiveness of the program?[7]

In addition, the evaluation also analyzes program participation rates (e.g. fraction of caregivers that attended meetings, fraction of households that read the SMS).

III.2. METHODOLOGY OVERVIEW

To evaluate the impact of MA, we conducted an experiment where students were assigned randomly to treatment and control groups. More specifically, we randomized schools into treatment and control groups, and surveyed samples of students in each school, as well as their caregivers. When successful, randomization ensures that characteristics of the treatment and control groups are, on average, observationally equivalent, with the only difference being their participation in the intervention. Therefore, any measured difference in outcomes between the groups over time can be attributed to the program.

Baseline data was collected between November 2015 and January 2016. For the endline, fielded in January 2017, we surveyed the same students and their caregivers again in order to construct a panel. Three instruments were fielded at baseline and endline (All questionnaires can be found in Annex A): a caregiver questionnaire, a student questionnaire, and a reading assessment.

The endline caregiver and student surveys were slightly modified relative to their baseline versions to incorporate program uptake questions. As mentioned in the previous section, in June 2016 a survey on a subsample of caregivers in the treatment group was fielded to document program uptake.

---

[7] This question was introduced during the endline phase of the evaluation; it was not part of the original evaluation questions discussed in the Evaluation Design Report.

The reading assessment was a short version of the Early Grade Reading Assessment (EGRA) translated into ChiNyanja. This instrument covers letter sound identification, oral reading (non-words and real words), reading comprehension and listening comprehension.

To estimate the impact of the MA package of interventions on reading skills and other outcomes (research questions 1 to 4) we use randomization as the identification strategy. The statistical analysis models the outcomes of interest as a function of being in the treatment group (i.e. receiving the MA intervention), the baseline value of the outcome of interest, and other sociodemographic variables as controls.[8]

Research question 5 looks at spillover effects of the program. Specifically, we document if children other than the targeted children in the household participated in MA reading activities.

Finally, we conduct a cost effectiveness analysis to answer research question 6.

Sample and Data Collection[9]

During the evaluation design phase, NORC estimated that a sample of 2,400 students, specifically 30 students surveyed in 40 treatment schools and 30 students in 40 control schools, will allow estimation of a standardized effect size of 0.19.[10]

To construct the sample of children and caregivers, the sampling strategy had two phases, first we sampled schools, and then we sampled students.[11] To sample schools we randomly selected 80 schools from the list of eligible 293 public schools in Chipata and Lundazi.[12] Therefore, the results of this evaluation are representative of the population of eligible students attending these 293 schools.

---

[8] The regression model we used to estimate the causal impact of the program on reading skills can be described by: $A_{s1} = \alpha + \beta D_s + \gamma A_{s0} + \mathbf{x_d'}\boldsymbol{\delta} + u_{s1}$; where, As1 and As0 measure achievement (EGRA score) for student s at endline and baseline, respectively; Ds is a dummy variable indicating treatment status; xs is a vector of characteristics of the student and of his or her household; us1 is an error term and α, β, γ and δ are the parameters to be estimated. The coefficient of interest is β, which measures the contribution of the intervention to student achievement. This model is usually referred as a Value-Added model (VAM), and is widely used in the education literature. The estimated effects by this model might be different from the effect of actually taking up the program. However, given that take-up of the program was relatively high, the effect of merely offering the services and the effect of taking up the program should be relatively similar. In Annex B we discuss the methodology and results of an approach to measure the treatment effect on those that engage in the MA activities.

[9] A more detailed description of the data collection activities is available in Annex C.

[10] Assuming an alpha of 5 percent, power level of 80 percent, an intracluster correlation coefficient of 0.1, and endline attrition of 20 percent.

[11] While the program does not work through schools but directly with students and their caregivers, the sampling frame for data collection was created through schools.

[12] From the original EMIS list of 350 school we dropped 10 due to missing enrollment data. We also dropped 13 schools due to missing EMIS number. In addition, schools that had no cell phone coverage or were difficult to access were also dropped from the sample frame, following a requirement from the program implementer. The DEBS of Chipata and Lundazi provided data sets with the information on which schools had cell phone coverage and which were easily accessible. We lost 2 schools that did not merge with DEBS data, in addition to 19 schools that did not have cell phone coverage and 5 schools that were hard to access. Finally, we dropped 8 schools because they either had less than 25 students in 1st grade or less than 25 students in 2nd grade.

Forty of the 293 schools were randomly assigned to the treatment group and 40 to the control group.[13]

In October 2015, NORC sent Advanced Teams (AT) of enumerators to survey all students in grades 1 and 2 across all 80 sampled schools to determine whether anyone in their home had a cell phone.[14] In total, the ATs attempted interviewing 8,681 students, out of which 4,910 (56.6 percent) reported having a cell phone at home; 2,354 students (27.1 percent) were absent during the AT visit, so the cell phone ownership rate of 56.6 percent can be considered a lower bound of the true cell phone ownership rate.

After the list of students that reported having a cell phone at home was constructed, enumerators randomly selected a sample of students in each school (15 from each grade and 10 replacements from each grade—for a total of 50 students per school). These students and their caregivers constituted the sample for the evaluation.

The baseline caregiver data collection occurred between mid-November and early December 2015. The caregiver questionnaire included a series of screener questions, intended to determine the respondent's eligibility to participate in the study. These screening criteria were agreed upon by NORC, Creative, USAID and World Vision, and included questions about the respondents' interest in participating in the intervention and the respondents' (actual) access to a working cell phone number. The same screening process was conducted in treatment and control groups to guarantee balance between the two groups across these dimensions. Respondents did not know to which group they were going to be assigned. In the case that multiple children from the same household were sampled, household-level questions were only asked once, and only child-specific questions were asked twice. We completed 2,222 household interviews, corresponding to a total of 2,397 students. During the caregiver survey, the replacement rate, calculated as the number of replacement households interviewed as a fraction of the total number of interviewed households, was 14.4 percent. The main reason why households could not be interviewed was because members of the household were not home or available to interview after repeated visits (a total of 149 households). Only 68 households were screened out because they did not have a working cell phone.

In January 2016, enumerators visited schools to conduct the EGRA and the student survey, assessing 2,260 students. All instruments were reviewed and approved by a local IRB (ERES Converge) as well as by NORC's IRB.

In June 2016, a randomly selected subsample of caregivers from the treatment group was surveyed to measure program uptake; that is, the extent to which parents were receiving the messages and sharing the content with their children. Data on other implementation aspects were also collected. The uptake report can be found in Annex D.

---

[13] The random selection of 80 schools used distance to DEBS and schools size for stratification. These variables were also used to match schools before randomization.
[14] Because the sample was constructed at the end of the school year (October 2015), we interviewed students in grades 1 and 2 so most of them would transition to grades 2 and 3, respectively, in 2016, when the program was treating them.

Endline data was collected between mid-January and early February of 2017. Instruments administered for the endline data collection were the caregiver questionnaire, student questionnaire, and the EGRA. For purposes of comparability, the same EGRA tool used for the baseline was repeated at endline data collection.

In total 1,942 caregivers were surveyed at baseline and endline, 965 in the treatment group and 977 in the control group. These 1,942 caregivers correspond to 2,091 children (there are more students than caregivers because 149 siblings were also interviewed). Table 1 shows final fieldwork results for the caregiver data at the children level. Out of these 2,091 children for whom there is caregiver data, 2,054 were interviewed for the EGRA and student survey at endline, 1,021 in the treatment group and 1,033 in the control group.[15]

Table 1. Endline data collection results

	N	Percent
Completed	2,091	87.2
Not locatable	231	9.6
Refused	28	1.2
Child transferred/dropped	4	0.2
Other	43	1.8
Total	2,397	100

Source: Own calculations using MA endline data

Along these lines, completion rate was 87 percent for caregivers and 86 percent for students.

Table 2 presents indicators, expected outcomes and data collection instruments associated with each research question.

---

[15] Not all the 2,054 children that were interviewed for EGRA at endline have EGRA baseline data. The sample of 'panel' students with EGRA data both at baseline and endline is 1,973, 991 in the treatment group and 982 in the control group.

Table 2. Evaluation Design Matrix

Research Question	Indicator / measurement question	Expected Outcomes	Instruments
1. Have the reading skills of Grade 2 and 3 students receiving the treatment improved as a result of MA? What are the magnitudes of these improvements, and in which reading domains have skills improved?	Reading assessment scores (EGRA)	Higher scores	Baseline and Endline Early Grade Reading Assessments (EGRA)
2. How and to what extent have student attitudes toward reading changed as a result of MA? Are students enjoying reading at home? Are they more likely to participate in out-of-school reading activities in the community?	Does [child] like to read? Does [child] like to listen to stories? Does [child] participate in reading activities outside home after school? How often?	Increase in reported motivation to read, listen to stories and participation in after school reading activities	Student questionnaire Caregiver questionnaire
3. Do parents spend more time supporting their children's reading activities as a result of MA? How much time are they spending on reading activities?	How often do parents read with [child] and for how long?	Increase in frequency/duration of time spent by parents reading with child	Caregiver questionnaire
4. Do students spend more time reading at home on their own as a result of MA? How much time are they spending on reading activities as a result of MA?	How often does [child] read on his/her own at home and for how long?	Increased frequency/duration of time spent reading independently	Student questionnaire Caregiver questionnaire
5. Are there any spillover effects of MA? Are other children in the household participating in the MA/SMS reading activities?	Other than [child], have any of his/her siblings participated in reading MA/SMS?	Siblings participate in reading MA/SMS activities	Caregiver questionnaire
6. What is the cost-effectiveness of the program?	Cost per child treated, cost per wpm gained due to the program	N/A	Costs provided by implementing partner

Limitations

The results discussed in this report correspond to the effects the MA program had on the type of individuals that were part of the study. Threats to the external validity of the documented results are two-fold.

First, the student sample frame was constructed from the group of children that were in school on the day the Advanced Teams visited schools. Twenty-seven percent of enrolled students

were not in school during those visits. If absenteeism is random then the documented program impacts can be extrapolated to all children in eligible schools even if they were absent when Advanced Teams visited schools. More likely, however, is that the students that were not in school that day miss school more often than the average child, and hence, are probably different in other aspects that could affect their reading skills. Along these lines, the estimated effects cannot be extrapolated to students that were not in school the day Advanced Teams were there constructing the student sampling frames. A similar problem arises when we consider attrition from the endline survey. However, we do not believe this is a major limitation for this analysis because: i) completion rates at endline were high (87 percent for caregivers and 86 percent for children), and ii) we ran regressions on the EGRA subtasks using Inverse Probability Weights to control for attrition and the results (now shown) showed negligible differences with respect to the results documented in the report.[16]

Second, as is the case with any pilot or program fielded in a small area such as the districts of Chipata and Lundazi (relative to the whole country), the results cannot be automatically extrapolated to the rest of the country (or to the group of schools in these two districts that were not eligible for the MA program). Further research should address how similar these two districts are to the rest of the country in terms of conditions relevant to the program functioning and the outcomes it affects, including household socioeconomic characteristics, school enrollment, cell phone ownership and service quality, and spoken language.

Another important consideration worth highlighting is that documented impacts correspond to offering the full set of services that were provided in the context of the program, not simply the SMS stories. In other words, "treatment" includes the SMS messaging, as well as the monthly meetings, the radio campaign, and the notebooks that were provided to MA children. As we discuss further in the recommendations section, the fact that we estimate the impact of a bundle of services rather than only the effect of SMS messaging has implications for scalability.

Finally, the analysis showed that collecting reliable data that involves time spent reading can be challenging. For example, as we show below, when we evaluated the impact of MA on the frequency at which parents read with children we found contradicting results between the caregivers' and children's responses, suggesting that at least one of these sources did not provide reliable data.

---

[16] Inverse Probability Weights are a common technique to control for attrition and other missing data problems. It consists of modeling the probability that a respondent is not surveyed, and then using the predicted probabilities to overweight the observed data that was more likely to be not observed.

IV. FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

IV.1. FINDINGS AND CONCLUSIONS

> ***RQ1. Have the reading skills of Grade 2 and 3 students receiving the treatment improved as a result of the MA intervention? What are the magnitudes of these improvements, and in which reading domains have skills improved?***

To answer this question, we evaluated the impact of MA on five EGRA subtasks: number of letter sounds correctly identified, non-word reading, oral reading fluency (ORF), reading comprehension and listening comprehension[17]. We found that **the MA project had a positive and significant impact on three of these subtasks - namely, non-word reading, Oral reading Fluency (ORF) and reading comprehension. No effects were found for letter sound identification or listening comprehension.**

For this research question we first discuss difference-in-difference estimates of the program impact, and then the results from value-added models.[18]

Difference-in-differences estimates

Table 3 shows the mean score for each EGRA subtask at baseline and endline, by treatment status. To estimate the impact of MA for each EGRA subtask, we compare the change between baseline and endline of the treatment group to the change between baseline and endline for the controls. These difference-in-difference estimates, along with effect sizes are also displayed in the table. Below, we list specific findings:

- For letter sound identification, we found no effects. This is perhaps not very surprising as MA was not specifically providing materials to teach letter sounds.

- For non-word reading, students in the treatment group read 1.5 more words than those in the control group due to the program; this is equivalent to an effect size of 0.2, which is a reasonable effect for a reading intervention.

---

[17] For letter sound identification students were asked to make the sound of a list of 100 letters in one minute. For Non-word reading students needed to read a list of 50 made-up words. For the Oral reading subtask, students were asked to read a short passage out loud that had 41 words in one minute. Students were also asked 5 comprehension questions on this passage. The number of questions each child was asked varied depending on how much of the text they were able to read. Students that were not able to read anything were automatically assigned a zero for the subtask and were not asked any reading comprehension questions. Finally, for the listening comprehension section, students were asked five comprehension questions about a text the interviewer read for them. As all other instruments fielded in study, EGRA was conducted in ChiNyanga.

[18] The non-parametric Difference-in-differences estimator is the difference in the changes over time for treatment and control groups on the outcome of interest. Because no covariates are included as controls then the estimator does not make any assumptions in terms of functional form (that is, how covariates affect the outcome of interest), hence this estimator can be described as non-parametric.

- For Oral Reading Fluency, we found that students in the treatment group increased the number of words correctly read per minute by 3.4 due to the program, equivalent to an effect size 0.27.

- For reading comprehension, we found that, as a result of the program, students in treatment schools correctly answered 0.3 more questions than their control school counterparts; this gain is equivalent to an effect size of 0.23.

> MA had a positive and significant impact on reading skills, as measured by non-word reading, ORF and reading comprehension. Estimated effect sizes were 0.2 for non-word reading, 0.27 for ORF and 0.23 for reading comprehension

- No impact was found for listening comprehension. Notably, students do not seem to have made any progress over time in this subtask, answering approximately 3 questions correctly both at baseline and endline, regardless of whether they were in the treatment or the control group.

Table 3. EGRA scores at baseline and endline by experimental group

	Baseline			Endline			Diff-in-Diff	
	Con-trol (1)	Treat-ment (2)	Diff (3=2-1)	Con-trol (4)	Treat-ment (5)	Diff (6=5-4)	DID (7=6-3)	Eff Size[a] (8)
Letter Sounds (total correct)	6.9	8.6	1.7	9.5	10.9	1.4	-0.3	-0.03
Non Words (total correct)	2.2	3.2	1	4.5	7.0	2.5***	1.5***	0.20
Oral Reading Fluency	3.5	4.9	1.4	7.2	12.0	4.8***	3.4***	0.27
Reading comp. (total correct)	0.3	0.4	0.1	0.7	1.1	0.4***	0.3***	0.23
Listening comp. (total correct)	2.8	2.9	0.1	3.0	3.1	0.1	0	-0.05

[a]The effect size is calculated as the DID over the pooled standard deviation at endline.
Source: Own calculations using MA baseline and endline data. All observations recorded at baseline and endline are included, even those for which the panel is not balanced. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

To put these effects in perspective, the treatment group showed twice as much improvement as the control group over the same time period. In effect, gains in ORF for the control group amounted to 3.6 wpm between baseline and endline, while the equivalent gains in the treatment group were 7.1 wpm.

That being said, despite improvements in three reading subtasks, absolute reading levels are still low. For example, as shown in Table 3 the ORF score at endline for the treatment group is only 12 wpm, while the "minimum" performance standard adopted by the MOGE for 2nd grade is 25 wpm.[19] In other words, while this intervention had a positive impact on a subset of reading skills, much work is still needed for these children to achieve accepted benchmarks.

---

[19] Performance levels are described as "outstanding," "desirable," "minimum," and "below minimum." (Ministry of Education, Science, Vocational Training and Early Education, 2014).

It is also worth mentioning the relative imbalance in test scores between treatment and control groups observed at baseline. Despite the fact that treatment was randomized, students in the treatment schools outperformed students in control schools across all five reading skills at baseline. While no difference is significant at 5 percent, differences in Letter Sound Identification and Non-Word Reading are significant at the 10 percent level. The main risk of having unbalanced data at baseline is that differences at endline may not be attributable to the program impact as other factors could be affecting the estimated results. As we show below, when we discuss the value-added models, including several covariates in the analysis in the regressions do not seem to affect the estimated impacts with respect to the difference-in-differences estimates, suggesting that controlling for baseline values of the outcomes of interest, as both difference-in-differences and value-added models do, is enough to control for differences at baseline that may affect the outcomes of interest.

Improvements in reading levels

In this section we analyze the fraction of students that reached certain meaningful reading thresholds. We follow performance standards adopted by the government (Ministry of Education, Science, Vocational Training and Early Education, 2014). For 2nd grade these standards define 'below minimum' as reading fewer than 25 wpm, 'minimum' as reading between 25 and 39 wpm, 'desirable' as reading between 40 and 59 wpm, and 'outstanding' as reading 60 wpm or more. For 3rd grade 'below minimum' is reading fewer than 40 wpm, 'minimum' as reading between 40 and 59 wpm, 'desirable' as reading between 60 and 69 wpm, and 'outstanding' as reading 70 wpm or more.

Figure II shows the fractions of students in treatment and control groups by these reading levels for students in grade 2 in panel A and grade 3 in panel B. Given the large fractions of students that cannot read one single word, the graphs also display the fractions of zero-readers. Panel A shows results for students that were in 2nd grade in 2016. The fraction of 2nd graders in the control group not able to read a single word declined from 80 percent at baseline to 64 percent at endline. This decline was counterbalanced mostly by increases in the fractions of 'below minimum' and 'minimum' readers between baseline and endline. A similar dynamic can be observed for the treatment group although the changes were more pronounced as a result of the program. **In effect while the decline in the fraction of students that were not able to read a single word in the control group was 16 percentage points, the decline for the treatment group was 21 percentage points.** The decline in the number of students that cannot read a single word in the treatment group was mostly counterbalanced by increases in 'below minimum' and 'minimum' readers.

Panel B shows results for 3rd graders. Students in 3rd grade also observed major decreases in the fractions of zero-readers between baseline at endline. In the control group the decline in the fraction of zero-readers was 14 percentage points, while for the treatment group was 25 percentage points. These declines were compensated, in both treatment and control groups, mostly by increases in the fractions of 'below minimum' students between baseline and endline.

Similar results for non-word reading, reading comprehension and listening comprehension are presented in Annex E.

Figure II. Students in treatment and control groups by reading level – Oral Reading Fluency
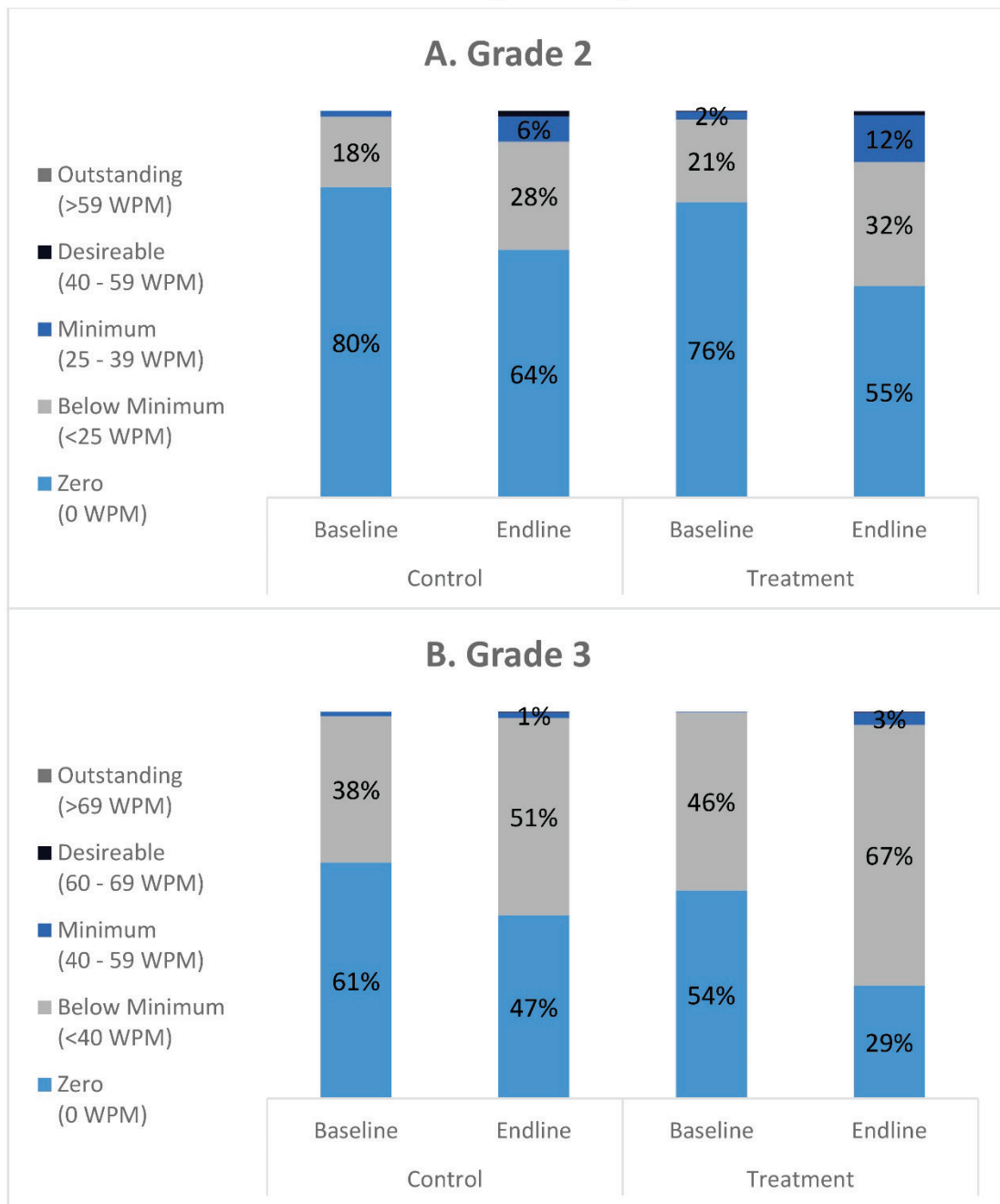
Note: Fraction labels of 1 percent or less are dropped from the graph for clarity.
Source: Own calculations using MA baseline and endline data. Performance levels are from Ministry of Education, Science, Vocational Training and Early Education (2014).

Value-added models

In addition to the difference-in-difference method shown above, we also calculate the impact of MA using Value-Added models. The key feature of these models is that we control for baseline test scores, as well as a series of other variables. Specifically, at the student level we control for age, gender and grade level, and at the household level we control for household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity. Descriptive statistics on these baseline sociodemographic characteristics are available in Annex F. The main improvement of this approach, compared to the non-parametric difference-in-difference just discussed, is that it controls for a range of sociodemographic characteristics, thereby improving the precision of the estimates.

The results obtained through the Value-Added models do not differ much to the non-parametric difference-in-difference estimates, which is not surprising given that treatment status was randomized. Similar to the findings under the difference-in-difference approach, we found no effects for the number of correctly pronounced letters or listening comprehension, but found positive and significant impacts for the other three evaluated subtasks.

Differential impacts

We explored whether the program had different impacts across four dimensions of interest. First, we estimate separate models for students that were able to read at least one word at baseline and students that were not. Given that the program was providing reading materials, it is reasonable to expect that it had a greater impact on students that were already able to read at least one word at baseline, than on students that were not able to read a single word at baseline. We also analyze if the program had different impacts depending on whether or not there were any reading materials at home at baseline. Considering that there were no reading materials for children in almost half the surveyed households, it can be expected that the program had greater impacts in households where no reading materials are available, than in households were children have some reading resources. We also estimate separate models by grade, and for boys and girls.

Out of these four dimensions of interest we only found significant differences in treatment effects when we analyze differential impacts by grade. Specifically, students in grade 3 progressed more than students in 2nd grade thanks to the program. These results are discussed in detail in Annex G.

Figure III shows the impact of MA on each EGRA subtask for grades 2 and 3, in terms of effect sizes. Overall, students in grade 3 made more progress than students in grade 2. The differences in the treatment effects by grade are statistically significant except for listening comprehension.

Figure III. Impact of MA on Reading Skills by Grade
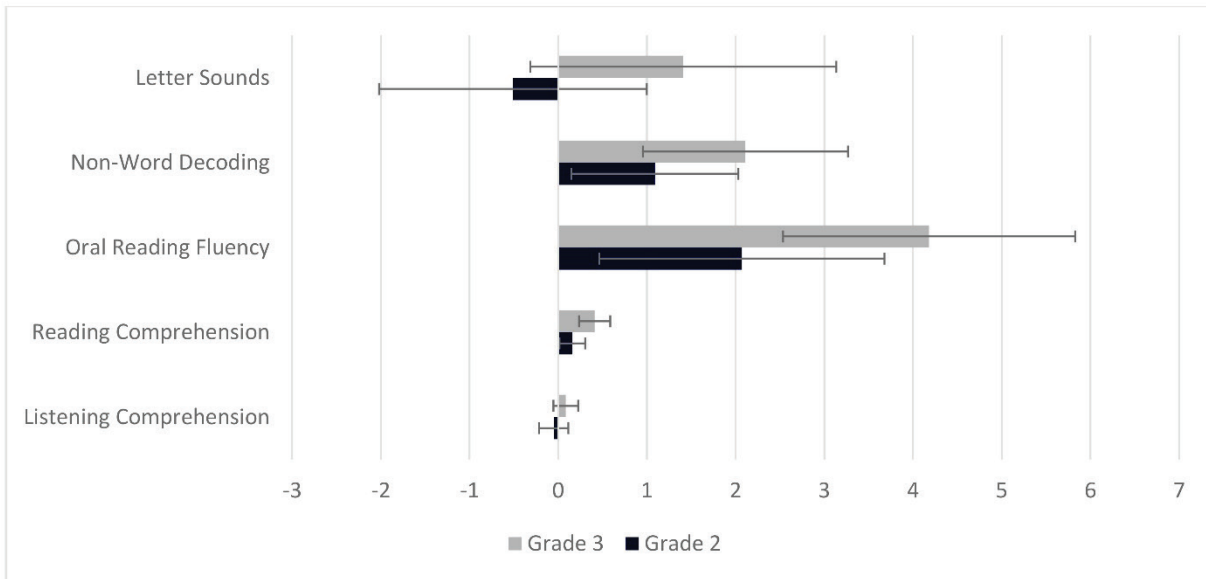
*For Oral reading fluency the units are wpm
Source: Own calculations using MA baseline and endline data.

The estimates presented above correspond to the effect of merely offering the program services. When we consider the effect of actually engaging in the program activities, discussed in Annex B, the impacts are larger. However, it is important to keep in mind that the policy-relevant effects are the ones discussed above; after all, whether participants engage more or less in program activities is, to some extent, outside the control of the program implementer (or the policymaker); hence, the effects that are relevant for programmatic and policy decisions are those related to offering the services.

In summary, the main conclusions with respect to RQ1 are as follows:

▪ The program had a positive and significant impact on reading skills, as measured by non-word reading, ORF and reading comprehension. Estimated effect sizes were 0.2 for non-word reading, 0.27 for ORF and 0.23 for reading comprehension.

▪ No effects were found for letter sound identification or listening comprehension. Failing to find impacts on letter sound identification is perhaps not surprising as the program was not addressing this skill specifically; more puzzling are the results for listening comprehension, as it would seem that reading the SMS (especially with others in the family) should positively affect this skill.

▪ More students in the treatment group reached meaningful reading thresholds than in the control group.

▪ Endline reading levels, even for students in the treatment group, are still low relative to Zambia's benchmarks. Reading levels were low at baseline and continued to remain low despite the improvements due to MA.

▪ Students in 3rd grade benefitted more from the program than students in 2nd grade.

To analyze children's attitudes towards reading, we asked them questions on whether they like reading or practicing reading in different contexts, and provided them with three response options: Like, Indifferent, and Do not like. Children in both treatment and control groups were surveyed at baseline and endline. Descriptive statistics on these and other data are available in Annex F. With these data, we estimated the probability that children would increase their preference for reading in a given context as a result of MA (for details of this analysis, see Annex H).

We did not find any evidence that the program changed children's attitudes towards reading. Children's preferences for listening to stories, reading or practicing reading, reading at home on his/her own, or reading with someone else did not change as a result of MA. It is important to highlight that most children in both treatment and control groups reported that they like reading. In effect, at endline 94 percent children in the treatment group said that like reading or practice reading, while the figure for the control group was 93 percent. This implies that the program had little room to improve reading attitudes as measured by these questions. Moreover, 98 percent of the children in the treatment group said they find the MA stories enjoyable or very enjoyable.

To evaluate the effect of the program on the likelihood that children participate in out-of-school reading activities, we questioned caregivers about the availability of such activities in their communities, as well as their children's participation in them. These activities include reading clubs, reading activities at the community center and visiting the school library, among other options.

> The program did not have a statistically significant effect on whether children like reading at home on their own or with a relative.

We found that MA did not have an effect on children's participation in out-of-school reading activities either. In this case it would seem that the driving factor is the actual lack of out-of-school activities available in the communities. In effect, both at baseline and endline only 12 percent of caregivers reported having access to out-of-school reading activities for the children; hence a program like MA probably had little chance to affect the likelihood that children increased their participation in this type of activities, considering that program did not generate these activities (see details in Annex H).

To address this question, we collected data on how often caregivers read with their children according to the respondent of the caregiver survey (who was a parent 80 percent of the cases at endline), and according to the child. As was the case with children, caregivers in both treatment and control groups were surveyed at baseline and endline.

In the caregiver survey, we asked whether parents read with the MA child four days a week or more, two or three days a week, one day a week, once or twice a month, less than once a month or never. We also asked the caregiver how long each reading session lasted in minutes.

Using parents' responses our estimates (presented in detail in Annex H) indicate that the parent that spends the most time reading with the child in the treatment group reads with the child 0.5 days more per week – or one more day every two-weeks - than their counterparts in the control group as a result of MA. This result is statistically significant.

We also estimated the impact of the program on the number of minutes that each reading session lasts but found the program did not have a significant impact on session duration.

We also asked children about the time they spent reading at home with parents. In this case, rather than giving the child the options provided to the caregiver, we opted for a simplified version of the question and only asked which days in the previous week someone read with him or her at home. The impact estimates using children's answers indicate no change in the number of days per week that children spend reading with parents at home due to MA (see details in Annex H). To keep the student survey simple, we did not ask how long sessions lasted.

It is rather puzzling that the effect of MA on joint reading as reported by the caregiver indicates a clear positive impact, while the effect on the same underlying construct as reported from the child's perspective indicates that the changes were negligible. Recognizing that it is possible some of these children may be too young to accurately report how often they read with someone at home, we ran regressions on the number of days children read with someone at home, as reported by the child, restricting the sample to older children (9 years old, which is the median age at baseline). However, again, we did not find any impact of the program, which suggests that the lack of results is not driven by some of these children being too young to assess correctly the number of days they read with someone at home. Of course, it is possible they could be all too young to report with any accuracy.

Another possible explanation is that caregivers in the treatment group tend to overstate how frequently they read with the child, given that they participated in the MA program and they may feel there is an expectation that they behave in a particular way.

Ideally, we would like to determine which respondent provided more reliable information: the child or the caregiver. To this end, we conducted a series of exercises to assess whether the caregivers' data is more reliable than the children's or vice-versa (see Annex H for details). The main exercise we conducted tries to evaluate how stable these two groups' responses are over time (between baseline and endline) regarding the frequency with which parents read with children. Our analysis indicates that caregiver responses seem more reliable than children's, suggesting that we should trust the (positive and statistically significant) results we found using caregivers' data, more than the (no change) results we found using children's responses.

However, given the contradicting results that we obtained from the caregivers' and children's responses for RQ3, there is no conclusive evidence that the program had a positive impact on the frequency at which children read with someone at home. Further research could address

more carefully the source of the discrepancy between caregivers' and children's results on this outcome.

RQ4. Do students spend more time reading at home on their own as a result of MA? How much time are they spending on reading activities?

To analyze the impact of the program on the frequency with which MA children read at home on their own, we collected data from both the child and the caregiver. Children and caregivers in both the treatment and control groups were surveyed at baseline and endline. In the caregiver survey, we asked if the child read four days a week or more, two or three days a week, one day a week, once or twice a month, less than once a month or never. In other words, we used the same response structure that we used to record the frequency at which household members read with the child. We also asked the caregiver how long each reading session lasted in minutes. In the student survey, we asked how many days they read at home on their own last week. Therefore, in this case too, we used the same question structure that we used in the student survey to record reading with a household member.

According to caregivers, children in the treatment group read on their own at home 0.36 days more per week than children in the control group. This translates into children reading 12 more minutes per week due to the program and it is a statistically significant impact. No significant effect is observed for the number of minutes each session lasts.

When we look at the effect of the program on the number of days children read alone according to the children, the result indicates that children in the treatment group read 0.23 more days per week due to the program.

> The program had a positive impact on the amount of time children spend reading at home on their own

While the findings on the number of days children read alone derived from the child and caregiver data are different, they are both positive and significant, and the actual difference between the two coefficients is small. These results suggest that the program had a positive impact in the number of days students read at home on their own.

Note that for this research question we did not find contradicting results between the responses provided by the caregiver and the responses provided by the child. It is possible that children were able to provide more reliable data when they were asked about reading alone than when they were asked about reading with someone at home; in effect, as we show in Annex H, children's responses on how often they read alone between baseline and endline are more correlated than when the outcome is reading with someone at home.

We can only speculate as to why children's responses would be more reliable for reading alone than for reading with someone else. Albeit the questions follow the same structure, perhaps it is easier for a child to remember how often s/he reads on her/his own, than remembering how often anyone at the household reads with them, which involves remembering interactions with probably several people.

To answer this research question we analyzed whether other children in treatment households participated in MA activities. Table 4 shows the number of MA children that read SMS with their sisters, brothers, cousins and other children from outside the household.

The results show that siblings, cousins and children from outside the households read the SMS stories with the MA child. We found that 138 children in the treatment group read the SMS stories with their sisters; this is equivalent to 25 percent of the MA children that read the SMS messages *and* have sisters. Brothers seem more enthusiastic than sisters about reading the SMS with their MA siblings; 246 MA children read the SMS stories with their brothers, which is equivalent to 43 percent of children that read the SMS messages and have brothers. Sixteen MA children read the SMS with their cousins, equivalent to 12 percent of children that read the SMS messages and have cousins (or more specifically, live with cousins). Finally, 112 MA children read the SMS with children from outside the household, equivalent to 12 of children that read the SMS stories.

It is interesting that MA children are more likely to read SMS with brothers than with sisters. Further research could address possible explanations for this.

These results indicate that MA children were joined by other children in the household when reading the MA stories. These children were not only family members living with the MA children but also children from outside the household. This indicates that the program could have had spillover effects on reading habits of children other than the MA children.

> MA stories are being read by children other than the MA children. We found that brothers of MA students are participating more actively in these reading activities than sisters, and that cousins and even children outside the household participate in the activities

Table 4. Spillover effects: Participation of other children in reading SMS

MA children that read with sisters MA stories (count)	138
As fraction of MA children with sisters	20%
As fraction of MA children with sisters that open MA messages	25%
MA children that read with brothers MA stories (count)	246
As fraction of MA children with brothers	35%
As fraction of MA children with brothers that open MA messages	43%
MA children that read with cousins MA stories (count)	16
As fraction of MA children with cousins	10%
As fraction of MA children with cousins that open MA messages	12%
MA children that read with children from outside the household (count)	112
As fraction of MA children that open MA messages	13%

Source: Own calculations using MA baseline and endline data.

RQ6. What is the cost-effectiveness of the program?

In this section we discuss the costs of the project. Specifically, we estimate the costs that would be incurred if the project were scaled up at the national level. For this analysis we consider the population of interest, which is children in grades 2 and 3 in Zambia, as only children in these grades were treated during the MA pilot. We divide this population between children living in Lusaka and the Eastern province, where ChiNyanja is the language predominantly spoken, and the rest of the country where six other local languages are spoken. The reason for this division is that the costs for these two groups will be rather different, at least initially. In effect, while for ChiNyanja a bank of stories has already been constructed as a result of the MA pilot, for the rest of the country new stories will need to be developed, which will entail costs similar to the ones observed for the MA pilot.

Moreover, we restrict the population of interest to the fraction that we estimate has access to a cell phone at home. According to the results discussed in section III.2, 56.6 percent can be considered a lower bound of the true cell phone ownership rate. By using this lower bound we are also excluding from the analysis the students that were absent the day advanced teams visited schools.

In Annex I we discuss in detail how the cost estimates are constructed. **The final costs per student for the MA program as it was implemented in its pilot phase are US 20.1 for students in Lusaka or Eastern Province, and US 21.6 for students in other parts of the country.**

There are several considerations worth highlighting about these estimated costs.

- MA was piloted in predominantly rural areas. Implementation realities and associated costs, among other things, could change when the intervention is launched in urban areas. For example, cell phone ownership in urban areas could be higher. This would imply that program's fixed costs per student could be lower, improving the cost-effectiveness of the program.

- It is possible that the lack of reading materials is less of a problem in urban than in rural areas. This could decrease the level of interest children and caregivers have in the program because reading resources are not in such need.

- We assume that cell phone coverage is acceptably reliable across the country, which may not be the case in some areas.

- For scale up, if stories are translated rather than developed from scratch, then the development costs could be reduced. Moreover, even if stories are developed from scratch for each official language, costs after the first year would fall because no new stories would need to be developed, at least not as many as at the beginning, as the same stories could be used in subsequent years of the project.

Table 5 shows gains for each of the five analyzed EGRA subtasks per USD 10 spent. For each cost scenario, gains in scores and in terms of effects sizes are displayed. The program gains in terms of ORF are between 1.6 and 1.7 wpm per USD 10 invested. Note that only the effects

for non-word reading, ORF and reading comprehension are statistically significant, as discussed above.

Table 5. Program gains in EGRA scores per USD 10 spent

	Lusaka and Eastern province		Rest of the country	
	Scores	Effect sizes	Scores	Effect sizes
Letter Sounds (tot correct)	-0.13	-0.01	-0.12	-0.01
Non Words (tot correct)	0.77	0.10	0.71	0.09
Oral Reading Fluency	1.68	0.13	1.56	0.12
Reading comp. (tot correct)	0.15	0.12	0.14	0.11
Listening comp. (tot correct)	-0.02	-0.02	-0.02	-0.02

Source: Own calculations using impact evaluation results and data from Creative and INESOR.

IV.2. RECOMMENDATIONS

The main purpose of this evaluation was to test an innovative approach to delivering reading materials. Given the positive findings on the project's impact, consideration should be given to scaling up the program.

However, any scale-up decision must take into consideration the answers to the following questions:

(1) How feasible is it to replicate the MA program in its entirety? MA consisted not only of the delivery of SMS, which alone would be relatively easy to replicate in areas of the country with similar cultural and socioeconomic background (and cell phone coverage); it also consisted of numerous meetings to raise awareness and guide parents on what their roles and responsibilities were; awareness-raising through media; and a supporting voice-message system. The cost effectiveness analysis documented in this report should be used when considering scaling up the program.

(2) Are the gains in reading sustainable? The endline data were collected just after the implementer stopped sending messages and conducting the monthly meetings with caregivers, so it would be useful to document whether these gains in EGRA scores and reading habits are maintained over time even in the absence of the program.

REFERENCES

Creative Associates (2017). Makhalidwe Athu - All Children Reading: A Grand Challenge for Development. ACR Final Project Report August 2015-April 2017.

Creative Associates (2015). Makhalidwe Wathu. All Children reading: A grand challenge for development. A concept paper for project expansion.

Ministry of Education, Science, Vocational Training and Early Education (2014). Reading performance level descriptors for grades 1 – 4.

NORC (2016a). USAID Impact Evaluation of the Makhalidwe Athu Project (Zambia) - Baseline Report.

NORC (2016b). USAID Impact Evaluation of the Makhalidwe Athu Project (Zambia) – Uptake Survey Report.

RTI (2015). Proposing Benchmarks and Targets for Early Grade Reading and Mathematics in Zambia. Retrieved May 24, 2017 from: https://ierc-publicfiles.s3.amazonaws.com/public/resources/Zambia%20Benchmarking%20Activity_25Sep2015_Clean-Final.pdf

Wagner, Daniel, et al. (2014) Mobiles for Reading: A Landscape Research Review. USAID: Washington, DC.

ANNEX A. ENDLINE INSTRUMENTS

A1. SCRIPT FOR PHONE CALL

USAID Impact Evaluation of the *Makhalidwe Athu* Project (Zambia)

SCRIPT FOR PHONE CALL TO SCHEDULE CAREGIVER'S INTERVIEW
Objective: Schedule interview, ask for consent to survey the child.

1. **Good afternoon, my name is [Enumerator's Name] and I'm calling on behalf of Inesor, a research organization with the University of Zambia. May I speak with [Baseline Respondent's Name]?**

01 Yes→Skip to Q6
02 No

2. **Do you live in the same home as [Child's Name] { and [Sibling's Name] }?**

01 Yes
02 No [*THANK PARTICIPANT FOR THEIR TIME, END CALL, VERIFY NUMBER, REPORT TO SUPERVISOR*]

3. **Can you tell me when can I call to speak with [Baseline Respondent's Name]?**

01 Yes [*THANK PARTICIPANT FOR THEIR TIME, END CALL, SCHEDULE NEW CALL*]
02 No

4. **Can I speak with [Child's Name]'s { and [Sibling's Name]'s } parent or guardian?**

01 Yes →Skip to Q6
02 No

5. **Can you tell me when can I call to speak with [Child's Name]'s { and [Sibling's Name]'s } parent or guardian?**

01 Yes [*THANK PARTICIPANT FOR THEIR TIME, END CALL, SCHEDULE NEW CALL*]
02 No [*END CALL, THANK PARTICIPANT FOR THEIR TIME*]

6. **According to our records, this household was surveyed in November or December 2015, in the context of Makhalidwe Athu, a program designed to improve children's reading skills by sending short stories via SMS. We would like to schedule an interview with you to conduct a follow-up survey, which will take 35 minutes. Can we schedule a meeting at your home or at the school this [Date (Or Dates) Enumerators Visit School]?**

01 Yes, at home *[GET DIRECTIONS and schedule time]*
02 Yes, at the school *[Confirm school on file and schedule time]*
02 No

7. <u>**LOOP CHILD then SIBLING**</u>**: We are visiting [CHILD'S NAME]'S school on [Date (Or Dates) Enumerators Visit School] to survey 30 students. We will ask [CHILD'S NAME] to take a short, 20 minute reading assessment and questionnaire on reading habits at home. Prior to asking [CHILD'S NAME] to participate, we will explain the purpose of the questionnaire and only proceed if s/he wants to participate. The results of this assessment will have no effect on [CHILD'S NAME]'s grades at school. Do you authorize [CHILD'S NAME] to participate in the assessment and answer the questionnaire?**

    01 Yes
    02 No
[*END CALL, THANK PARTICIPANT FOR THEIR TIME*]

A2. PARENT/CAREGIVER HOUSEHOLD SURVEY INSTRUMENT

ENDLINE PARENT QUESTIONNAIRE FOR *MAKHALIDWE ATHU* PROJECT IN ENGLISH

PRESENTED TO:

Eres Converge

PRESENTED BY:

NORC at the University of Chicago
Institute of Economic and Social Research (INESOR), University
of Zambia

BASELINE QUESTIONNAIRE - *Makhalidwe Athu Project*

COVER SHEET INFORMATION
Enter enumerator name [enu_name**]: _____**

Date [enu_date]**: DD:** |__|__| **MM**: |__|__| **YYYY**: |__|__|__|__|

School ID [school_code]**:** |__|__|__|__|__|__|

Questionnaire No [quex_id]**:** |__|__|__|__|__|__|__|

GPS location of household [gps]**:**

LATITUDE(N/S) |__| – **DEGREES:**|__|__|__| **MINUTES:** |__|__| **SECONDS** |__|__|·|__|__|__|
LONGITUDE (E/W) |__| – **DEGREES**: |__|__|__| **MINUTES:** |__|__| **SECONDS** |__|__|·|__|__|__|

SECTION A: INTRODUCTION AND CONSENT [section_a_intro]
Good day. My name is [ENUMERATOR NAME] and I represent INESOR, a research organization with the University of Zambia. According to our records, this household was surveyed in November or December 2015, in the context of Makhalidwe Athu, a program designed to improve children's reading skills by sending short stories via SMS.

0A: 01 Interview Scheduled on Phone
      02 Interview not scheduled on the Phone [SKIP to 1E]

1A. Is this the home of [RESPONDENT]? [ ]
 	
        01   Yes
        02   No → [*ASK FOR DIRECTIONS TO HOUSE, AND END INTERVIEW*]
	|__|__|

1B. May I speak to [RESPONDENT]?
 	
        01   Yes
        02 No → **SKIP TO 1D**	

1C. §Is this the home of [CHILD'S NAME]? [child_home]
 	
            01 Yes → **SKIP TO 2**
            02 No → [*END INTERVIEW*]	

1D. IF RESPONDENT IS NOT HOME ASK WHEN INDIVIDUAL WILL BE AT HOME, AND RECORD ANSWER. REVISIT. IF PERSON REFUSES, THEN END INTERVIEW	

1E. Is this the home of [CHILD'S NAME]? [child_home]
 	
        01   Yes
        02   No → [*ASK FOR DIRECTIONS TO HOUSE, AND END INTERVIEW*]
	|__|__|

1F. ASK THIS QUESTION <u>ONLY</u> IF THERE IS A SIBLING. Is this also the home **of [SIBLING'S NAME]**? [child_home_sib]
        01   Yes
        02   No [CUT SIBLING QUESTIONS; ASK FOR WHEREABOUTS OF SIBLING]
	|__|__|

1G. May I speak to [BASELINE RESPONDENT'S NAME]'s?
 	
        03   Yes → **SKIP TO 2**
        04   No
	|__|__|

ENUMERATOR RESERVE CODES: 999=REFUSED; 888=DON'T KNOW; 777=NOT APPLICABLE   **27**

1H. May I please talk to the parent or guardian of [CHILD'S NAME] ? [ask_parent]
    01 Yes
    02 No→ **SKIP TO 1J**	

1I. What is your name? [name_resp] _____→ **SKIP TO 2**	
1J. IF PARENT OR GUARDIAN IS NOT HOME ASK WHEN INDIVIDUAL WILL BE AT HOME, AND RECORD ANSWER. REVISIT. IF PERSON REFUSES, THEN END INTERVIEW	

2. We are conducting the follow-up survey of Makhalidwe Athu, a study approved by the Ministry of Education. The survey will take approximately 30 minutes and we will ask you questions mostly about your child's reading practices.
Whatever information you will provide will be strictly confidential and not be shown to any other persons. Participation in the survey is completely voluntary. If we should come to any question that you do not want to answer, just let me know and I will go on to the next question. You are also free to stop the interview at any time or withdraw altogether. [introduction]
Should have any queries about the survey when I am gone you can contact the following:
 [HAND CONTACT INFORMATION TO RESPONDENT]
4. Do you wish to participate in this survey? May we start now? [consent]
_IF YES, CHECK HERE IF RESPONDENT CONSENTS TO INTERVIEW |__|. ASK RESPONDENT TO SIGN CONSENT FORM. IF NO, END INTERVIEW._

4. <u>**LOOP FOR CHILD AND SIBLING**</u>: What is your relationship with [CHILD'S NAME]? [ENUMERATOR: DO NOT READ OPTIONS] [relationship] 01 MOTHER 02 FATHER 03 GRANDMOTHER 04 GRANDFATHER 05 AUNT 06 UNCLE 07 SISTER 08 BROTHER 09 COUSIN 10 OTHER: Specify _____	|__|__|
SKIP TO Q6 IF INTERVIEW WAS SCHEDULED (CONSENT ALREADY AQUIRED)	

Thank you. In this context of this study, we will also visit [CHILD'S NAME]'s school and ask [CHILD'S NAME] to take a short, 20 minute reading assessment and questionnaire on reading habits at home. Prior to asking [CHILD'S NAME] to participate, we will explain the purpose of the questionnaire and only proceed if s/he wants to participate. We will also not interview [CHILD'S NAME] if you do not want us to. The results of this assessment will have no effect on [CHILD'S NAME]'s marks at school.

5. <u>**LOOP FOR CHILD AND SIBLING**</u>: Do you authorize [CHILD'S NAME] to **participate in the assessment and answer the questionnaire?** [assessment_auth] 03 Yes 04 No → [*END INTERVIEW, THANK PARTICIPANT FOR THEIR TIME*]	|__|__|

SECTION B: HOME LITERACY ENVIRONMENT

6. *<u>**LOOP FOR CHILD AND SIBLING**</u> What is the grade of [NAME OF CHILD]? [child_grade]	|__|__| GRADE

First, I would like to ask you some basic questions about your household and reading practices with your children. You are a household member if: (i) You have lived under this "roof" or within the same compound/homestead/stand at least 15 days during the last 12 months OR you arrived here in the last 15 days and this is now your usual residence; (ii) when you are together you share food from a common source with other household members; and (iii) you contribute to or share in a common resource pool.

I am going to ask you some questions about the members of this household. IF SIBLINGS ARE SAMPLED, TELL PARENT THE FOLLOWING: I am first going to ask about [NAME OF CHILD 1] and then I will ask about [NAME OF CHILD 2]. If you do not feel comfortable providing the name of the household member, provide us with the initials. *ENUMERATORS: IF THEY DO NOT WANT TO GIVE NAMES, PLEASE ASK FOR INITIALS OR SOME OTHER WAY FOR US TO REFER TO THEM. COLLECT ALL NAMES FIRST IN 7A AND THEN ASK 7G-7H FOR EACH PERSON BEFORE MOVING ON TO THE NEXT HOUSEHOLD MEMBER. IF THEY DO NOT WANT TO PROVIDE THE INITIALS, TYPE IN RELATIONSHIP TO CHILD. E.G. FATHER, MOTHER, UNCLE."* [roster_msg]

[LOOP 7A TO 10 FOR CHILD THEN SIBLING]

	7A [roster_name]	7G* [roster_readfreq]	7Ga* [roster_readfreq_a]	7H* [roster_readlen]
	Please tell me the name of the parents of [NAME OF CHILD] and yourself if you are not one of the parents, starting with yourself. **Please note that no names will be entered in our reports in order to protect your privacy.**	**In a typical week, how often does [NAME] _read_ with [NAME OF CHILD]? Please include time reading books, notebooks or messages in cell phones, except for any MA material** 01 Four Days A Week Or More02 02 Two Or Three Days A Week 03 One Day A Week 04 Once Or Twice A Month 05 Less Than Once A Month 06 Never→ *[SKIP TO NEXT PARENT]*	**In a typical day, how many times does [NAME] _read_ with [NAME OF CHILD] in the same day?**	**Each time [NAME] _reads_ with [NAME OF CHILD], on average how long does he/she spend reading to [NAME OF CHILD]?** MINUTES
01	_____	|_|_|	|_|_|	|_|_|
02	_____	|_|_|	|_|_|	|_|_|
03	_____	|_|_|	|_|_|	|_|_|

8. * In a typical week, how often does [NAME OF CHILD] read with all <u>other</u> household members? 01 Four Days A Week Or More 02 Two Or Three Days A Week 03 One Day A Week 04 Once Or Twice A Month 05 Less Than Once A Month 06 Never→ *[SKIP TO Q11]*	|_|_|
9. * In a typical day, how many times does [NAME OF CHILD] read with all other household members in the same day?	|_|_|
10. * Each time other household members _read_ with [NAME OF CHILD], on average how long do they spend reading to [NAME OF CHILD]?	|_| MINUTES

[LOOP 11 THROUGH 20 FOR CHILD THEN SIBLING]
 	
11. * Does [CHILD'S NAME] have a reading materials (e.g. reader's book, reading cards, or reading passages) from school? [readersbook]
    01 Yes
    02 No→ *[SKIP TO Q0]*
	|_|_|

12. * Does [CHILD'S NAME] ever bring his/her reader's book home from school? [bringhomebook]
    01 Yes
    02 No→ *[SKIP TO Q0]*
	|_|_|

13. * If yes, on average, how many days a week did [CHILD'S NAME] bring it home in the last month of the last school term? [bringhomebook_days] *ENTER A NUMBER 0-5*: |_| days/week	|_| DAYS/WEEK
14. *How often does [CHILD'S NAME] read on his/her own at home? [readaloneathome] 01 Four Days A Week Or More 02 Two Or Three Days A Week 03 One Day A Week 04 Once Or Twice A Month 05 Less Than Once A Month 06 Never → *[SKIP TO Q16]*	|_|_|
15. *When [CHILD'S NAME] reads alone at home, how long does he/she usually read? [readaloneathome_mins] |_|_| MINUTES	|_|_|

16. *Is there a program or place where [CHILD'S NAME] can participate in reading activities in your community? [readingactivities]
 	
    01 Yes
    02 No → *[SKIP to Q21; DON'T KNOW SKIP TO Q21]*
	|_|_|

17. *Does [CHILD'S NAME] participate in any reading activities outside home after school, like reading clubs, visiting the school library, etc.? [readingactivitiespart]
    01 Yes
    02 No → *[SKIP to Q21]*
	|_|_|

18. *In which of these outside home reading activities does [CHILD'S NAME] participate? *[READ ALOUD OPTIONS; CHECK ALL THE APPLY]*
[readingactivities_type]
    01 Reading Clubs
    02 Reading activities at the community center
    03 Reading through church activities
    04 Visit the school library
    05 Reading with friends
    06 Reading with a mentor/older pupil in the community
    07 Story-telling competitions
    08 Reading competitions/Read-a-thons
    09 Writing clubs
    10 Other:_____
	|_|_

_	_
_	_
_	_
_|_|

19. *In a typical week, how often did [CHILD'S NAME] participate in any of these outside home activities (if [CHILD'S NAME] participated in more than one of these activities, consider the number of times he participated in total)? [readingactivities_often]
 	
    01 Four Days Or More A Week
    02 Two Or Three Days A Week
    03 Once A Week
    04 Once Or Twice A Month
    05 Less Than Once A Month
 	
20. *How long does [CHILD'S NAME] spend on these reading activities each time he/she participates, on average? [readingactivities_long]
 	
    |_|_| MINUTES
	|_|_|

I would now like to ask you for your opinion on your child's reading practices and progress.

[LOOP 21 THROUGH 26 FOR CHILD THEN SIBLING]

21. *By the end of what grade do you expect [CHILD'S NAME] should be able to read? [expectread] |_|_| GRADE	|_|_|
22. *Do you think [CHILD'S NAME] can read as well as a child his/her age is supposed to? [readatage] 01 Yes 02 No	|_|_|
23. *Do you feel confident you can help [CHILD'S NAME] to learn how to read? [confident] 01 Yes 02 No	|_|_|
24. *Would you say that helping [CHILD'S NAME] learn how to read is … [helpingread_difficulty] 01 **Very difficult** 02 **Difficult** 03 **Neutral** → *[SKIP TO SECTION C]* 04 **Easy** → *[SKIP TO SECTION C]* 05 **Very easy** → *[SKIP TO SECTION C]*	|_|_|
25. *Why is it difficult? *[CHECK ALL THAT APPLY]* [difficulty_why] 01 CANNOT READ WELL ENOUGH TO TEACH CHILD 02 DO NOT KNOW HOW TO HELP CHILD READ 03 DO NOT HAVE ACCESS TO APPROPRIATE READING MATERIALS 04 DO NOT HAVE ENOUGH TIME 05 DO NOT THINK THAT HELPING CHILD READ OUTSIDE OF SCHOOL WILL MAKE A DIFFERENCE 06 NO INTEREST FROM CHILD 07 CHILD DOES NOT HAVE ENOUGH TIME (E.G. HE/SHE HAS TO WORK) 08 OTHER: _____	|_|_| |_|_| |_|_| |_|_| |_|_|
26. Do you think that the responsibility of teaching a child to learn how to read is [responsibility]: 01 **Primarily a responsibility of the teacher** 02 **Equally a responsibility of the parent and the teacher** 03 **Primarily a responsibility of the parent** 04 **Other: _____**	|_|_|
27. Do you think it is important for children to read outside of school? [readoutside_imp] 01 Yes 02 No	|_|_|

SECTION C: PARTICIPATION IN *MAKHALIDWE ATHU*

28. Have you or any household member attended any meeting or training session about Makhalidwe Athu? [attend] 01 Yes 02 No → **[SKIP TO Q34]**	|__|__|
29. How many meetings have you or anyone in your household attended about Makhalidwe Athu? [num_attend] **[IF ANSWER to Q29 IS 0, SKIP TO Q34]**	|__|
30. Who attended these meetings? [mtg_attend] 01 I attended them **02** Another household member attended them **[SKIP TO Q33]**	|__|__|
31. What did you learn during these meetings? [SELECT ALL THAT APPLY: ENUMERATOR DO NOT READ OPTIONS] [month_learn] 01 WHAT THE PROGRAM WILL DO 02 WHAT IS EXPECTED OF ME/THE HOUSEHOLD AND STUDENT 03 HOW MANY TIMES A WEEK THE HOUSEHOLD WILL RECEIVE THE SMS STORIES 04 HOW TO ACCESS SMS STORIES ON MY PHONE 05 WHAT TO DO IF WE CANNOT RECEIVE SMS STORIES 06 HOW TO BEEP FOR A VOICE RECORDING OF THE STORIES 07 HOW TO HELP MY CHILD READ THE SMS STORIES AT HOME 08 HOW MUCH TIME I SHOULD DEDICATE EVERY WEEK TO HELPING MY CHILD READ THE STORIES AT HOME 09 HOW TO HELP MY CHILD WRITE THE SMS STORY IN A NOTEBOOK 10 WHO TO CONTACT IF I HAVE QUESTIONS 11 HOW AND WHERE TO SUBMIT STORIES THAT I CREATED OR MY IDEAS FOR STORIES 12 THAT BREEZEFM BROADCASTS A MA READING PROGRAM EVERY SATURDAY 13 HOW I MIGHT RECEIVE A SHORT SURVEY ON MY PHONE ASKING ME ABOUT THE MA PROGRAM 14 NOTHING 15 I CANNOT REMEMBER 16 OTHER: _____	|__|__| |__|__| |__|__| |__|__|
32. How prepared did you feel after this meeting to open and read the SMS stories with your child? [intro_prep] 01 Very prepared 02 Somewhat prepared 03 Not at all prepared	|__|__|

33. Which household member(s) attended these meetings? [ENUMERATOR: DO NOT READ OPTIONS, SELECT ALL THAT APPLY] [mtg_attend_who]
      01 CHILD'S MOTHER
      02 CHILD'S FATHER
      03 CHILD'S GRANDMOTHER
      04 CHILD'S GRANDFATHER
      05 CHILD'S AUNT
      06 CHILD'S UNCLE
      07 CHILD'S SISTER
      08 CHILD'S BROTHER
      09 CHILD'S COUSIN
      10 OTHER: Specify _____
 	
[SKIP TO Q35]
	|__|__|
	|__|__|
	|__|__|
	|__|__|

34. Why did you or your household not attend these meetings? [SELECT ALL THAT APPLY: ENUMERATOR DO NOT READ RESPONSES] [why_no_attend]
      01. DID NOT KNOW MEETINGS OCCURRED
      02. DID NOT THINK THEY WOULD BE USEFUL
      03. ALREADY KNEW INFORMATION
      04. DID NOT HAVE TIME
      05. OTHER: _____
	|__|__|
	|__|__|
	|__|__|
	|__|__|

35. Did your phone break or was your cell phone service interrupted for any reason since the start of the last school year? [phone_break]
      01 Yes
      02 No
	|__|__|

36. Did you change your number since the start of the last school year? [phone_change]
      01 Yes
      02 No
	|__|__|

37. In general, how often would you say you have a clear cell phone reception at home? [reception]
      01 Always
      02 Usually
      03 About half the time
      04 Seldom
05 Never
	|__|__|

38. Have you or someone in your household received SMS stories and messages from Mahkalidwe Athu? [receive]
      01 Yes
 	
      No → **[SKIP TO Q73]**
	|__|__|

39. How many times in a week do <u>you or someone else in your household</u> receive SMS stories and other text messages from Makhalidwe Athu? [text] [DAYS/WEEK, 0-7] [receive_sms] **[IF 0, SKIP TO Q73]**	|__|

40. How often are <u>you or someone else in your household</u> able to open these SMS stories and text messages when you receive them? [access]
      01 We can open every text message
      02 We can only open 2 out of 3 text messages
      03 We can only open 1 out of 3 text messages
      04 We can never open them **[SKIP TO Q52]**
	|__|__|

PROGRAM: ASK THIS QUESTION <u>ONLY</u> IF SIBLINGS ARE BEING SURVEYED
 	
41. **In general, are SMS stories shown to [CHILD'S NAME] and [SIBLING'S NAME] at the same time?**
      01 Yes **[ASK QUESTIONS Q42 TO Q47 ONLY ONCE]**
      02 No **[ASK QUESTIONS Q42 TO Q47 BY CHILD]**
      03 Stories are never shown to either children	

42. **^How often do <u>you or someone else in your household</u> show the SMS stories to [CHILD'S NAME]? [show_text]**
  01 We show every SMS message (3 times/week)
  02 We show 2 SMS messages (2 times/week)
  03 We show 1 SMS message (1 time/week)
  04 We show less than once a week/at least once a month
  05 Never **[SKIP TO Q47]**
	|__|__|

43. **^When you and [CHILD'S NAME] read the SMS stories at home, who else participates? [SELECT ALL THAT APPLY: ENUMERATOR: DO NOT READ OUT RESPONSES] [participants]**
  01 MOTHER OF CHILD
  02 FATHER OF CHILD
  03 GRANDMOTHER OF CHILD
  04 GRANDFATHER OF CHILD
  05 AUNT OF CHILD
  06 UNCLE OF CHILD
  07 SISTER OF CHILD
  08 BROTHER OF CHILD
  09 COUSIN OF CHILD
  10 CHILDREN FROM OUTSIDE THE HOUSEHOLD
  11 ADULTS FROM OUTSIDE THE HOUSEHOLD
  12 OTHER: Specify _____
  13 NO ONE
	|__|__|
	|__|__|
	|__|__|
	|__|__|

44. **^In the lasdt week that you received SMS stories, what days of the week <u>did you or someone else</u> help [CHILD'S NAME] read the SMS stories? [SELECT ALL THAT APPLY] [days_read]**
  01 Monday
  02 Tuesday
  03 Wednesday
  04 Thursday
  05 Friday
  06 Saturday
  07 Sunday
  08 None of the above **[SKIP TO Q47]**
	|__|__|

45. **^In general, when you or someone else helps read the SMS story messages with [CHILD'S NAME], how does it happen? [how_read]**
  01 You read and [NAME OF CHILD] listens most of the time
  02 You read half the time and [NAME OF CHILD] reads the other half
  03 You listen and [NAME OF CHILD] reads most of the time
  04 You are not present when child reads SMS messages
	|__|__|

46. **^On average, how long <u>did you or someone else</u> spend helping [CHILD'S NAME] read one SMS story message? [MINUTES] [min_read]** **SKIP TO Q48 IF > 0**	|__|__|

47. ^What prevented <u>you or someone else</u> from reading with [CHILD'S NAME] last week? [SELECT ALL THAT APPLY] [prev_days] ENUMERATOR: DO NOT READ RESPONSES
    01 DID NOT HAVE ENOUGH TIME IN DAY/HAD OTHER TASKS TO DO
    02 DID NOT KNOW HOW TO USE SMS OR VOICE RECORDING
    03 DID NOT RECEIVE/COULD NOT ACCESS THE SMS STORY OR VOICE RECORDING
    04 DO NOT FEEL CONFIDENT READING WITH CHILD
    05 DID NOT UNDERSTAND THE STORY WELL ENOUGH
    06 CHILD DID NOT HAVE ENOUGH TIME (E.G. SHE HAS TO WORK)
    07 NO INTEREST FROM CHILD
    08 TOO DIFFICULT FOR CHILD
    09 OTHER: _____
 	
[ASK IF Q44 is NONE OF THE ABOVE OR LESS THAN 3 DAYS ARE SELECTED]
	|__|__|
	|__|__|
	|__|__|
	|__|__|

48. * <u>LOOP FOR CHILD THEN SIBLING</u>: Does [CHILD'S NAME] use the notebook you were provided to record SMS stories? [notebook]
    01. Yes
    02. No → **SKIP TO 52**
    03. Your household did not receive a notebook → **SKIP TO 52**
	|__|__|

49. * Who writes down the stories in the notebook? [SELECT ALL THAT APPLY, ENUMERATOR DO NOT READ RESPONSES BUT PROBE 'WHO ELSE?']
 	
01 I DO
    02 **[CHILD'S NAME]**
    03 [SIBLING'S NAME]
 	
03 OTHER CHILDREN IN HOUSEHOLD
04 OTHER ADULTS IN HOUSEHOLD
05 OTHER: SPECIFY	

50. * <u>LOOP FOR CHILD THEN SIBLING</u>: Does [CHILD'S NAME] reads the SMS stories from the notebook?
    01. Yes
    02. No
	|__|__|

51. Who else reads the SMS stories from the notebook? [ENUMERATOR: DO NOT READ THE RESPONSES, SELECT ALL THAT APPLY]
    01 I DO
    02 OTHER CHILDREN IN THE HOUSEHOLD
    03 OTHER MEMBERS OF THE HOUSEHOLD
    04 ADULTS FROM OUTSIDE THE HOUSEHOLD
    05 CHILDREN FROM OUTSIDE THE HOUSEHOLD
    06 NO ONE
	|__|__|
	|__|__|
	|__|__|

PROGRAM: ASK THIS QUESTION <u>ONLY</u> IF SIBLINGS ARE BEING SURVEYED
 	
52. In general, are voice recordings listened by [CHILD'S NAME] and [SIBLING'S NAME] at the same time? [how_listen_child]
    01 Yes **[ASK QUESTIONS Q53 TO Q54 ONLY ONCE]**
    02 No **[ASK QUESTIONS Q53 TO Q54 BY CHILD]**
    03 Stories are never listen by either children **[SKIP TO 57]**
	|__|__|

53. ^^Do <u>you or someone else in your household</u> listen to the voice recording with [CHILD'S NAME]?
    01 Yes
    02 No **[IF Q40 AND Q42 ARE "NEVER", SKIP TO END; OTHERWISE SKIP TO Q57]**	

54. **^^How often do <u>you or someone else in your household</u> listen to the voice recording when reading with [CHILD'S NAME]? [listen_voice]** 01 After we receive every SMS message (3 times/week) 02 After we receive 2 SMS messages (2 times/week) 03 After we receive 1 SMS message (1 time/week) 04 Less than once a week	
55. **When do you listen to the voice recording? [how_listen_parent]** 01 I listen <u>before</u> I open the SMS 02 I listen <u>after</u> I open the SMS 03 I do not read the SMS messages: I just listen to the voice recording 04 I listen while looking at the notebook where my child copied the story 05 Other: _____	|__|__| |__|__| |__|__| |__|__|
56. **Who else listens to the voice recording? [ENUMERATOR: DO NOT READ THE RESPONSES, SELECT ALL THAT APPLY AND PROBE 'WHOE ELSE?']** **[listen_voice_who]** 01. I DO 02. OTHER CHILDREN IN THE HOUSEHOLD 03. OTHER MEMBERS OF THE HOUSEHOLD 04. ADULTS FROM OUTSIDE THE HOUSEHOLD 05. CHILDREN FROM OUTSIDE THE HOUSEHOLD 06. NO ONE	|__|__|
57. **How do you discuss the stories? [READ OPTION ALOUD; SELECT ALL THAT APPLY] [discuss_how]** 01. We talk about the comprehension questions in the text/voice recording 02. I ask additional questions to test my child's comprehension 03. We have a conversation about the story 04. We do not talk about the stories at all **[SKIP TO Q60]**	|__|__| |__|__| |__|__| |__|__|
58. ***<u>LOOP FOR CHILD THEN SIBLING</u>: How often do you discuss these stories with [CHILD'S NAME]? [discuss_freq]** 01. We discuss these stories almost every day 02. We discuss after we receive every SMS message (3 times/week) 03. We discuss after we receive 2 SMS messages (2 times/week) 04. We discuss after we receive 1 SMS message (1 time/week) 05. Less frequently than once a week, but at least once a month	
59. **Who else participates in these discussions? [SELECT ALL THAT APPLY. ENUMERATOR: DO NOT READ THE RESPONSES] [discuss_who]** 01. OTHER CHILDREN IN THE HOUSEHOLD 02. OTHER ADULTS OF THE HOUSEHOLD 03. ADULTS FROM OUTSIDE THE HOUSEHOLD 04. CHILDREN FROM OUTSIDE THE HOUSEHOLD	|__|__| |__|__| |__|__| |__|__|
60. **How easy are these stories for <u>YOU</u> to read? [adult_read]** 01 Very difficult 02 Difficult 03 Neutral 04 Easy 05 Very Easy	|__|__|
61. How enjoyable are these stories for <u>YOU</u>? [parent_enjoy] 01 Very enjoyable 02 Enjoyable 03 Somewhat enjoyable 04 Not enjoyable at all	

[LOOP 62 THROUGH 64 FOR CHILD THEN SIBLING]
62. * How easy are these stories for [CHILD'S NAME] to <u>read</u>? [child_read]
  01 Very difficult
  02 Difficult
  03 Neutral
  04 Easy
  05 Very Easy
	|__|__|

63. * How easy are these stories for [CHILD'S NAME] to <u>understand</u>? [child_comp]
  01 Very difficult
  02 Difficult
  03 Neutral
  04 Easy
  05 Very Easy
	|__|__|

64. * How enjoyable does [CHILD'S NAME] find these stories? [child_enjoy]
  05 Very enjoyable
  06 Enjoyable
  07 Somewhat enjoyable
  08 Not enjoyable at all
	|__|__|

65. How helpful would you say this program is in helping your child learn to read? [help_read]
  01 Very helpful
  02 Helpful
  03 Somewhat helpful
  04 Not at all helpful
	|__|__|

66. Would you be willing to pay for SMS stories and call-in to voice messages programs?
  01 Yes
  02 No **[SKIP TO Q68]**	

67. How much would you be willing to pay per week for a service like the one provided by **Makhalidwe Athu?** _____	

68. What suggestions do you have on improving the program? [suggest_improve]
[CHECK ALL THEY APPLY; DO NOT READ THE BELOW CHOICES BUT PROBE 'WHAT ELSE?']
  01. PROVIDE MORE TRAINING SESSIONS
  02. PROVIDE CLEARER INSTRUCTIONS
  03. PROVIDE MORE STORIES
  04. PROVIDE FEWER STORIES
  05. SEND MORE TEXT MESSAGES
  06. SEND FEWER TEXT MESSAGES
  07. SEND MORE INTERESTING STORIES
  08. MAKE TEACHER RESPONSIBLE FOR READING AND DISCUSSING WITH CHILD
  09. PROVIDE ONE-ON-ONE SUPPORT TO HELP ME WITH MY SPECIFIC SITUATION
  10. PROVIDE BETTER QUALITY PHONE TRANSMISSIONS
  11. PROVIDE BETTER PHONE
  12. PROVIDE READING/STORY BOOKS FOR THE HOME
  13. PROVIDE PROGRAM FOR OTHER GRADES
  14. OTHER: _____
	|__|__|
	|__|__|
	|__|__|
	|__|__|

69. Have you or someone else in your household ever submitted a story or story idea for use by the Makhalidwe Athu program? [story_submit]
  01. Yes
  02. No
	|__|__|

70. Did you know that each week BreezeFM broadcasts a reading program for children and adults? [know_breeze] 01. Yes 02. No	|__|__|
71. Have you or anyone in your household ever listened to BreezeFM reading program on the radio? [listen_breeze] 01. Yes 02. No	|__|__|

SECTION D. SUBSTITUTION

72. Are you enrolled in any other program or participate in any other activity that uses cell phones to provide reading materials for [CHILD'S NAME]? 01 Yes→ *[SKIP TO Q74]* 02 No→ *[SKIP TO SECTION E]*	|__|__|
73. Are you enrolled in any program or participate in and activity that uses cell phones to provide reading materials for [CHILD'S NAME]? 01 Yes 02 No→ *[SKIP TO SECTION E]*	
74. What is the name of this program? _____	

SECTION E: FINISH INTERVIEW

We have now come to the end of interview. Do you have any immediate question/s about the interview?
Thank you very much for participating in this survey! Please reach out to the number listed on the information brochure if you have any questions or concerns. Also, please reach out to the number if you do move. [end]

75. ENUMERATOR: WHERE DID YOU CONDUCT THIS INTERVIEW? [location] 01 School [SKIP TO END] 02 Home 03 Other: Specify _____	|__|__|

76. **Household address/direction to household** [full_address]**: _____**

Enumerator comments [comments**]:**
_____
_____
_____

A3. EGRA/STUDENT SURVEY INSTRUMENT

STUDENT READING ASSESSMENT (EGRA) IN ENGLISH FOR *MAKHALIDWE ATHU* PROJECT

ENGLISH AND CHINYANJA

NORC at the UNIVERSITY *of* CHICAGO

PRESENTED TO:

Eres Converge

PRESENTED BY:

NORC at the University of Chicago
Institute of Economic and Social Research (INESOR), University of Zambia

Zambia Mini Early Grade Reading Assessment (EGRA)

General instructions

First, make sure you are surveying the right child. Then establish a playful and relaxed rapport with the child through a short conversation (see example topics below). The child should perceive the assessment almost as a game to be enjoyed rather than a test. Use this time to identify in what language the child is most comfortable communicating. Read aloud slowly and clearly ONLY the sections in boxes.

Hi, are you [CHILD'S NAME]?
01 Yes
02 No [END INTERVIEW]

Uli bwanji. Dzina langa 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 response; if pupil is reluctant, ask question 2, but if they seem comfortable continue to verbal consent].
2. **Kodi ndi masewera otani amene umakonda kusewera?** What games do you like to play?

Verbal Consent: Read the text in the box clearly to the child.

- **Ndifuna kukuuza cifukwa cake ndabwera kuno lero. Ndigwira nchito mu unduna wa maphunziro mu Zambia ndipo tikufuna kumvetsetsa mmene ana amaphunzirira kuwerenga . Iwe wasankhidwa mwamwai.** 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 surveyed in January 2016 and today we would like to ask you a few more questions.
- **Ndifuna thandizo lako 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 nkoloko iyi, ndizaona nthawi imene utenga kuwerenga .** Using this stopwatch/device/gadget, I will see how long it takes you to read.
- **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.
- **Ndizakufunsanso mafunso ena monga kumene umayeselera kuwerenga ndiponso ngati ukonda kuwerenga. .** I will also 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 wakonzeka kuti tiyambe?** Are you ready to get started?

Check box if verbal consent is obtained: ☐ *YES*

(If verbal consent is not obtained, thank the child and move on to the next child, using this same form)

IF CHILD AGREES, ASK HIM/HER TO WRITE NAME TO THE RIGHT		

Demographics

Date of assessment	
Assessor's Name	
District:	
School Name:	
School EMIS Number:	
Pupil ID	
Sex:	☐ Boy ☐ Girl
Grade:	☐ G2 ☐ G3

Assessor ID:		Date: / / /
Disposition code	**Complete** 01 Complete- school 02 Complete- home **Will interview later** 02 Absent (not at school) **No interview** 03 Not locatable (the child was not locatable at home) 04 Disability prevents taking the exam 05 Child too ill to take exam 06 Parent refuses interview	

TASK 1: ORIENTATION TO PRINT	📖 **Page X**	🕐 **X**

	Materials: a passage from the pupil stimuli packet
Show the child a story passage in the pupil stimuli packet. Read the instructions in the gray boxes below, provide the child 10 seconds to respond, recording the child's response before moving to the next instruction. If the child doesn't respond in the 10 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.

1.	*(Child puts finger on the top row, left-most word)*	O Correct	O Incorrect	O No Response

Tsopano ndionetse mbali imene udzawerenga motsatira.

Now show me in which direction you would read next.

2.	*(Child moves finger from left to right)*	O Correct	O Incorrect	O No Response

Ukafika kotsirizira kwa mzere, udzawerenga kuti motsatira?

When you get to the end of the line, where would you read next?

3.	*(Child moves finger to left-most word of second line)*	O Correct	O Incorrect	O No Response

Total Correct	/3

TASK 2: LETTER SOUND IDENTIFICATION	📖 **Page 1**	⏱ 60 seconds

🎤 **Pano ndili ndi tsamba limene liri ndi malembo a alifabeti ya muchinyanja. Coonde ndiuze MAMVEKERO a malembo a alifabeti amene ungathe 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 lembo ili ndi /a/.** For example, the sound of this letter is /a/.

[point to the letter p] **Tiye tiyese: ndiuze mvekero la lembo ili:** Let's practice: Tell me the sound of this letter.
✓🎤 **Cabwino, mvekero la lembo ili ndi /p/** Good, the sound of this letter is /p/.
✗🎤 **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.
✓🎤 **Cabwino, mvekero la lembo ili ndi /l/.** Good, the sound of this letter is /l/.
✗🎤 **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. Ngati wafika pa lembo limene sudziwa, pitiriza kupita ku lembo lotsatira. Ika cala cako pa lembo loyamba. Wakonzeka? Yamba.** When I say "Begin," 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. 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.

✍ (/) Mark any incorrect letters with a slash
(Ø) Circle self-corrections if you already marked the letter incorrect
( ]) Mark the final letter read with a bracket

Examples: A   p   L

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
m	N	K	I	d	k	A	J	m	u	*(10)*
C	d	b	o	L	I	U	K	A	w	*(20)*
G	n	a	e	s	E	A	D	I	g	*(30)*
l	r	A	a	v	f	A	T	W	i	*(40)*
D	a	t	L	N	a	A	M	i	Y	*(50)*
t	u	z	N	i	I	N	k	e	O	*(60)*
u	Z	P	i	U	N	i	M	i	l	*(70)*
A	p	A	a	B	W	T	k	c	M	*(80)*
a	w	N	m	E	R	a	A	h	a	*(90)*
n	A	o	l	O	n	a	U	T	S	*(100)*

✍ Time remaining on stopwatch at completion (number of SECONDS)	
✍ Exercise discontinued because the child had no correct answers in the first line	

Wacita bwino! Tiye tipitirize patsamba lotsatira Good effort! Let's go on to the next section.

Right column text:

Start the timer when the child reads the first letter.

➲ If a child hesitates or stops on a letter for 3 SECONDS, point to the next letter and say "Go on"

✋ When the timer reaches 0, say "stop."

✋ If the child does not provide a single correct response on the first line (10 items), say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

Task3: Non-word Reading

🕮 **Apa pali mau opangidwa mcinyanja. Ndifuna kuti uwerenge mau amene ungakwanitse kuwerenga, Uwerenge mau awa osati masipelingi.** 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: "ola"** For example, this made-up word is: "ola".

[point to the word "koki"] **Tiye tiyese: conde werenga liu ili.** Let's practice: Please read this word.
✓🕮 **wacita bwino. Liu ili ndi "koki"** Good, This made-up word is "koki."
✗🕮 **Liu 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.
✓🕮 **"wacita bwino, liu lopangidwa ili ndi "cota"** Good, This made-up word is "cota."
✗🕮 **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. 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. 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.

✎ (/) Mark any incorrect words with a slash
(Ø) Circle self-corrections if you already marked the word incorrect
( ]) Mark the final word read with a bracket

Examples: ola   koki   cota

1	2	3	4	5	
nipe	atapi	gelu	kelo	mdzimu	*(10)*
ninane	wondi	umbe	rizi	ninda	*(20)*
ledesi	fikiraku	tomo	ngalo	zirama	*(30)*
yu	ane	mwane	mukudi	dzimo	*(40)*
liraku	ia	anuli	wekusera	dzimoli	*(50)*
cofukwa	udi	kubu	anauna	mtisinaka	*(60)*
wera	eka	diko	amoi	kasuci	*(70)*
ateta	lia	nacho	komi	labo	*(80)*
menepa	ncheto	ndaako	nthua	balo	*(90)*
mtanyama	mtutu	ndokonda	mtingi	ko	*(100)*

✎ Time remaining on stopwatch at completion (number of SECONDS)	
✎ Exercise discontinued because the child had no correct answers in the first line	

Wacita bwino! Tiye tipitirize patsamba lotsatira Good effort! Let's go on to the next section.

Sidebar:
Start the timer when the child reads the first word.

➲ If a child hesitates or stops on a letter for 3 SECONDS, point to the next word and say "Go on"

✋ When the timer reaches 0, say "stop."

✋ If the child does not provide a single correct response on the first line (5 items), say "Thank you!", discontinue this subtask, check the box at the bottom, and go on to the next subtask.

TASK 4a: ORAL READING PASSAGE (2014)	☻ 60 seconds	
Show the child the sheet in the pupil stimulus booklet as you read the instructions. 🖁 **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. Ngati wapeza liu limene sudziwa, pita ku liu 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 as you can. 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.	➲ If a child hesitates or stops on a letter for 3 SECONDS, say "Go on" 🖐 If the child does not provide a single correct word on the first line of text. Do not ask any comprehension questions.	After the child is finished reading, REMOVE the passage from in front of the child. Ask the child only the questions related to the text read. A child must read all the text that corresponds with a given question. If the child does not provide a response to a question after 15 seconds, mark "no response" and continue to the next question. Do not repeat the question.
		🖁 **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.

✎ (/) Mark any incorrect letters with a slash
  (Ø) Circle self-corrections if you already marked the letter incorrect
  ( ]) Mark the final letter read with a bracket
		✎ (✓) 1 = Correct
		  (✓) 0 = Incorrect
		  (✓) . = No response.

		Questions [Answers]			
Amai anapita kumsika m'masana tsiku <u>lina</u>.	6	**Ndani anapita kumsika?** (Amai)	1	0	.
Anasiya mwana ndi mkulu wake D<u>olika</u>.	12	**Mwana anatsala ndi ndani?** (Dolika)	1	0	.
Anzake a Dolika anabwera kudzamtenga pamodzi ndi mwanayo. Dolika ndi anzake anaphunzitsa mwana <u>kuyimba</u>. Anamuphunzitsa nyimbo ya alifabeti.	30	**Kodi mwana anaphunzitsiwa kucita ciani?** (Kuyimba, Kuyimba nyimbo ya alifabeti)	1	0	.
Atabwerako kumsika amai, anapeza mwana ali <u>kuyimba</u>.	37	**Kodi mwana anadziwa bwanji kuyimba nyimbo ya alifabeti?** Dolika ndi anzake anamphunzitsa)	1	0	.
Amai anakondwera <u>kwambiri</u>.	40	**N'cifukwa ciani amai anakondwera?** (Mwana anali kuyimba)	1	0	.

✎ Time remaining on stopwatch at completion (number of SECONDS)	
✎ Exercise discontinued: the child had no correct answers in the first line	

Wacita bwino! Tiye tipitirize patsamba lotsatira Good effort! Let's go on to the next section.

TASK 5: LISTENING COMPREHENSION (2014)	📖 X			⏱ X
🗣 **Ndidzakuwerengera ka nthano/nkhani mokweza KAMODZI ndipo pambuyo pake ndidzakufunsa mafunso. Conde umvetsere mosamalira ndipo uyankhe mafunso mmene ungakwanitsire. Wakonzeka? Tiye Tiyambe.** I am going to read you a short story aloud ONCE and then ask you some questions. Please listen carefully and answer the questions as best as you can. Ready? Let's begin.				Remove the pupil stimuli booklet from the child's view.

✎ (✓) 1 = Correct
   (✓) 0 = Incorrect
   (✓) . = No response.
				Do not allow the child to look at the passage or the questions.

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.

Ndi tsiku liti pamene Mangani anapita kusukulu? (Pa Lolemba)	1	0	.
Ananyamula ciani mu chola cake? (Mabuku ndi nyama)	1	0	.
N'ciani cimene anapeza panjira? (Anapeza galu wamkulu)	1	0	.
Ndi cifukwa ciani Mangani anathawa galu? (Anaopa kuti galu angamulume, nyama, Anaopa, Galu wamkulu)	1	0	.
Ndi cifukwa ciani m'bale wake anamubwereka yunifomu Mangani? (Cifukwa yunifomu yake inada, Anagwa).	1	0	.

Wacita bwino! Tiye tipitirize patsamba lotsatira Good effort! Let's go on to the next section.

TASK 6: READING PRACTICES
ZOCHITA 6: MAWERENGEDWE

I am going to ask you some questions about the time you spend reading. If you do not want to answer any question, or do not know the answer to a question, you do not have to answer.

Ndizakufunsa mafunso pa nthawi imene umakhala uwerenga. Ngati siufuna kuyanka funso lilionse mwina kapena siudziwa yanko, usayankhe.

Can I begin asking you these questions?

Kodi ndiyambe kufunsa mafunsowa sopano?

(Inde 01, Ai 02➔ *SKIP TO END)* |__|__|

Some of the questions I am going to ask you are about what you like and do not like to do. Mafunso ena ndizakufunsa ndi a zinthu zimene ukonda ndi zimene siukonda kuchita

Q1	**Do you like to play?** Kodi ukonda kusewela?. 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|
Q2	**Do you like to go to school?** Kodi ukonda kupita ku sukulu? 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|
Q3	**Do you like to listen to stories?** Kodi ukonda kunvelera nthano? 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|
Q4	**Do you know how to read?** Kodi udziwa kuwerenga? 1 Inde **(Yes)** 0 Ai **(No)** 99 = Kulibe yanko **(Don't know/no response)**	|__|__|
Q5	**Do you like to read or like to practice reading?** Kodi ukonda kuwerenga kapena ukonda kuyetselera kuwerenga? 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|
Q6	**Do you read or try to read on your own at home?** Kodi umawerenga kapena umayetselera kuwerenga pa iwe wekha ku nyumba? 1 Inde **(Yes)** 0 Ai **(No)** [*SKIP TO QUESTION 11*] 99 = Kulibe yanko **(Don't know/no response)** [*SKIP TO Q11*]	|__|__|

Q7	**Do you like to read or try to read on your own at home?** Kodi ukonda kuwerenga kapena umayetselera kuwerenga pa iwe wekha ku nyumba? 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|
Q8 [multiple select]	**Last week, on which days did you read or try to read on your own at home? [ENUMERATOR SELECT ALL THAT APPLY]** Mu sabata lata, ndi matsiku ati yomwe unawerenga kapena unayetselera kuwerenga pa iwe weka ku nyumba? 1 MONDAY 2 TUESDAY 3 WEDNESDAY 4 THURSDAY 5 FRIDAY 6 SATURDAY 7 SUNDAY 8 I DID NOT DO THIS ACTIVITY LAST WEEK	|__|__|
Q11.	**Does anyone read or help you try to read with you at home?** Kodi kuli omwe ama werenga kapena omwe ama kutandiza kuyetselera kuwerenga ku nyumba? 1 Inde **(Yes)** 0 Ai **(No)** [*SKIP TO QUESTION 21*] 99= kulibe yanko **(Don't know/no response)** [*SKIP TO QUESTION 21*]	|__|__|
Q12.	**Who do you read with, or who helps you try to read, at home the most?** Kodi umawerenga ndi ndani, kapena mwina ndani omwe ama kutandizira kuyetselera kuwerenga ku nyumba kawiri-kawiri? 1 Amai **(Mother)** 2 Atate **(Father)** 3 Amai opezamo **(Stepmother)** 4 Atate opezamo **(Stepfather)** 5 Amai akulu kapena ang'ono **(Aunt)** 6 Atsibweni **(Uncle)** 7 Ambuya amuna **(Grandfather)** 8 Ambuya akadzi **(Grandmother)** 9 Akalongosi **(Sister)** 10 Abale **(Brother)** 11 M'zako **(Friend)** 12 Ena, simikiza **(Other, specify)** _____ 99= kulibe yanko **(Don't know/no response)**	|__|__|

Q13.	**Do you read or try to read with anyone else at home?** Kodi umawerenga kapena kuyetselera kuwerenga ndi munthu wina ku nyumba? 1 Inde 0 Ai [*SKIP TO Q15*] 99= kulibe yanko **(Don't know/no response)** [*SKIP TO QUESTION 15*]	|__|__|
Q14. [multiple select]	**Who else do you read with, or who else helps you try to read at home [DO NOT READ RESPONSES BUT PROBE WITH 'WHO ELSE?';** *SELECT ALL THAT APPLY*] Ndani wina omwe umawerenga naye kapena omwe ama kutandizira kuyetselera kuwerenga ku nyumba? [*sankani mayanko yonse omwe apasa*] 1 Amai **(Mother)** 2 Atate **(Father)** 3 Amai opezamo **(Stepmother)** 4 Atate opezamo **(Stepfather)** 5 Amai akulu kapena ang'ono **(Aunt)** 6 Atsibweni **(Uncle)** 7 Ambuya amuna **(Grandfather)** 8 Ambuya akadzi **(Grandmother)** 9 Akalongosi **(Sister)** 10 Abale **(Brother)** 11 M'zako **(Friend)** 12 Ena, simikiza **(Other, specify)** _____	|__|__|
Q15. [multiple select]	**Last week, on which days did anyone read to you at home, or help you try to read?** Mu sabata latha, ndi matsiku ati omwe munthu wina anakuwerengela kapena kukutandiza kuyetselera kuwerenga ku nyumba? 1 MONDAY 2 TUESDAY 3 WEDNESDAY 4 THURSDAY 5 FRIDAY 6 SATURDAY 7 SUNDAY 8 I DID NOT DO THIS ACTIVITY LAST WEEK 99 DON'T KNOW / NO RESPONSE	|__|__|
Q16.	**Do you like to read or try to read with someone at home?** Kodi ukonda kuwerenga kapena kuyetselera kuwerenga ndi munthu wina ku nyumba? 01=I like it 02=I do not like it or dislike it 03=I do not like it	|__|__|

Q17.	**Do you participate in any reading activities outside home after school, like reading clubs, visiting the school library, etc.?** *03* Yes *04* No → *[SKIP to Q19]*	
Q18.	**In which of these outside home reading activities do you participate?** *[SELECT ALL THE APPLY]* [readingactivities_type] 11 Reading Clubs 12 Reading activitys at the community center 13 Reading through church activities 14 Visit the school library 15 Reading with friends 16 Reading with a mentor or older pupil in the communicty 17 Story-telling competitionsRY-TELLING COMPETITIONS 18 READING COMPETITIONS/READ-A-THONS **19** WRITING CLUBS **20** OTHER:_____ 99= kulibe yanko **(Don't know/no response)**	
Q19. [multiple select]	**[CHOOSE 'NO SCHOOL' OPTION, DO NOT ASK IF ITS FIRST WEEK OF SCHOOL]** **Last week, which days did you attend school?** **INTERVIEWER: SELECT ALL THAT APPLY** Mu sabata latha, ndi matsiku ati omwe unapita ku sukulu? 9 NO SCHOOL LAST WEEK 0 ZERO DAYS 1 MONDAY 2 TUESDAY 3 WEDNESDAY 4 THURSDAY 5 FRIDAY	|__|__|
Q20.	"Last week, on the days you were in school, was your teacher in school present on all of those days?" Sabata latha pa masiku unapita kusukulu kodi aphunzisi anabwera masiku onse? 1 Inde **(Yes)** 0 Ai **(No)** 99 Don't know/No reponse	|__|__|

PARTICIPATION IN *MAKHALIDWE ATHU*

Mahkalidwe Athu is a program that sends short stories to your parents' cell phones so you and your relatives can read them together.
 	
Q21. Do you know if [CAREGIVER'S NAME] has received SMS stories and messages from Mahkalidwe Athu? [receive]
     04 Yes
     05 No → **[SKIP TO END]**
	|_|_|

Q22. Have you read any of these stories?
   01 Yes
   02 No → **[SKIP TO END]**	

Q23. The last week that you received SMS stories from Makhalidwe Athu, did you read an SMS story?
   01 Yes
   02 No → **[SKIP TO Q31]**	

Q24. Did you read this story more than once?
   01 Yes
   02 No → **[SKIP TO Q26]**	

Q25. How many times did you read this story?	|_| or DK
Q26. Did someone in your family help you read the SMS stories? 01 Yes 02 No	

Q27. Who did you read the SMS stories most often with?
 	
     11 MOTHER
     12 FATHER
     13 GRANDMOTHER
     14 GRANDFATHER
     15 AUNT
     16 UNCLE
     17 SISTER
     18 BROTHER
     19 COUSIN
     20 OTHER: Specify _____	

Q28. In general, when you read the SMS stories, how does it happen? [how_read]
     05 You read and someone else listens most of the time
     06 You read half the time and someone else reads the other half
     07 You listen and someone else reads most of the time
     08 You read on your own
	|_|_|

Q29. What do you do with the notebook you were provided?? READ OUT ONE BY ONE; CHECK ALL THAT APPLY [notebook]
     04. To record SMS stories
     05. To share stories with others
     06. To draw pictures about stories
     07. To write new stories
     08. Other school work _____
     09. Other activities _____
     10. You did not receive a notebook
	|_|_|

Q30.   Did any of your siblings, friends or classmates ever ask you to see the SMS stories or read your notebook?
 	
   01 Yes
 	
   02 No	

Q31.   Have you listened to the stories on the phone?
     03 Yes
     04 No **[SKIP TO Q60]**	

Q32. How often do <u>you</u> listen to the voice recording when reading the SMS stories? [listen_voice]
 	
     05 3 times/week
     06 2 times/week
     07 1 time/week
     08 Less than once a week	

Q33. After you listen to the SMS story on the phone, do you usually answer or discuss any of the questions with someone in your family? 01 Yes**[SKIP TO Q60]** 02 No	

Q34.Why not?
 	
     10 YOU WERE NOT VERY INTERESTED
     11 OTHER HOUSEHOLD MEMBERS WERE NOT VERY INTERESTED
     12 OTHER: _____	

Q35.   How easy are these stories for <u>YOU</u> to read?
 [child_read]
     06 Difficult
     07 Somewhat difficult
     08 Somewhat easy
     09 Easy
	|_|_|

Q36.   How easy are these stories for <u>YOU to</u> understand?
 [child_comp]
     06 Difficult
     07 Somewhat difficult
     08 Somewhat easy
     09 Easy
	|_|_|

Q37. Do you like the stories? Point to how you fell **[child_enjoy]**	|_|_|

Q38.   **Is it harder or easier to read stories on the phone than from a book?**
   01 Harder to read SMS stories on the phone than from a book
   02 Easier to read SMS stories on the phone than from a book
   03 The same	

Q39.   **Do you prefer to read stories on the phone or from a book?**
      01 Prefer reading SMS stories
      02 Prefer reading from a book
      03 The same	

Q40.   **Do you know that BreezeFM radio station features programs on reading?**
   01 Yes
   02 No **[SKIP TO END]**	

Q41.   **Did you listen to any of the BreezeFM radio programs on reading last week?**
   01 Yes
   02 No	

"Wow, you did a great job today! We are done now. Thank you for your help, here is a small token of thanks. You can go back to class now. Have a good day!"

ANNEX B. INTENT TO TREAT VERSUS TREATMENT ON THE TREATED

The parameters discussed in subsection IV.1 constitute the effect of merely offering the program. This effect might be different from the effect of actually taking up the program, that is, of participating in the program activities. In other words, the effect for children that live in participant households where every message is read and discussed, may be very different from the effect on children that live in participant households where most messages are ignored. In effect, more enthusiastic households may observe greater effects in reading skills than more apathetic ones.

Along these lines, the program evaluation literature usually refers to two different treatment effects; the effect of simply offering the services, regardless of the extent to which participants take up the treatment, is called the Intent-to-Treat estimator (ITT). This effect is the one estimated by equation (1). The second treatment effect is called Treatment-on-the-Treated effect (TOT), which takes into consideration the effect of the program on those that actually participate in the activities.

A naïve approach to estimate the impact of engaging in MA activities, defining engagement as, for example, showing the SMS to the MA children three times a week, could be to disregard the original treatment dummy, and estimate Equation (1) replacing D_s with a dummy variable D_s' that is equal to 1 if caregivers showed the SMS to the MA children three times a week and 0 otherwise. One threat to the internal validity of this approach is that households in the treatment group that showed the SMS three times a week may be different from households in the treatment group that did not engage in such behavior and showed fewer times the SMS to the MA children, and these differences could impact other outcomes, like reading skills. Hence, any estimate of the program impact using D_s' as the covariate of interest would confound the treatment effect with differences in these two groups that would have occurred even in the absence of treatment.

Instead of simply running Equation (1) using D_s' as the treatment variable, we can use D_s as an instrumental variable of D_s'. The objective of this approach is to purge D_s' from any factors that determined having showed SMS messages other than randomization. Therefore, the variation of D_s' used to estimate the impact of program participation or engagement only comes from randomization, not from other factors that could bias the estimate of the treatment effect.

In the context of the MA the differences between TOT and ITT should be relatively small, given the high take-up the program observed, as documented in section II. If all or most households in the treatment group engaged in the activities they were supposed to, then the parameters that we can estimate for ITT and TOT should be relatively similar.

The main challenge that needs to be addressed to produce TOT in the context of MA is how to define take-up. Take-up of the program could be defined as having showed the MA child the SMS three times a week. However, listening to the voice messages could also be considered program participation. We use the randomization dummy to evaluate treatment take-up for different definitions of treatment.

Table B1 shows results of this analysis. In each cell, the first number corresponds to the estimated impact, the second, in parentheses, to the standard error, and the third to the sample size. In the first row the results for the ITT, already discussed in subsection IV.1, are replicated as a reference. The second row shows the estimated impact of showing the SMS to child at least once a week. Just as we found for the ITT estimates, there are no significant effects for letter sounds or listening comprehension under any specification, and there are positive and significant effects for non-word reading, ORF and reading comprehension across all specifications. The estimated effects are slightly higher than those found using the ITT; for example the ITT estimate for ORF is 3.1 wpm, while for those households that showed the SMS at least once a week the impact is 3.9 wpm. As we define treatment take-up more restrictively the impact increases, so for households that showed every SMS to the child and listened to the voice recordings, the impact of the program on ORF was 6.4 words per minute, more than double the ITT effect.

Table B1. The effect of participating in MA using randomization as an instrument for participation

	Letter Sounds (Total Correct)	Non-Word Decoding (Total Correct)	Oral Reading Fluency	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
Intent-to-Treat	0.42	1.59***	3.11***	0.28***	0.015
	(0.69)	(0.43)	(0.65)	(0.065)	(0.065)
	[1966]	[1969]	[1968]	[1973]	[1973]
Showed at least one message per week to child	0.52	1.96***	3.86***	0.35***	0.018
	(0.84)	(0.53)	(0.78)	(0.078)	(0.079)
	[1966]	[1969]	[1968]	[1973]	[1973]
Shows every SMS to the child	0.64	2.42***	4.75***	0.43***	0.022
	(1.03)	(0.64)	(0.94)	(0.095)	(0.098)
	[1966]	[1969]	[1968]	[1973]	[1973]
Shows every SMS to child and listens to voice recording	0.86	3.27***	6.42***	0.58***	0.030
	(1.40)	(0.88)	(1.29)	(0.13)	(0.13)
	[1966]	[1969]	[1968]	[1973]	[1973]

Note: All models correspond to second stage regressions using the treatment dummy as an instrument for take-up. All regressions include as controls the lagged value of the dependent variable, child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a tv, bed, bicycle, radio, and has electricity). Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.

Standard errors clustered at the school level in parentheses and sample sizes in brackets.

* p<0.05 ** p<0.01 *** p<0.001

These results highlight the importance of engaging in program activities. While positive and significant effects were found using the ITT estimates for ORF and other two EGRA subtasks, the TOT estimates were larger. Interestingly, no effects were found for letter sound or listening comprehension under any specification; not even when we used the most comprehensive defection of program take-up we tried.

ANNEX C. DATA COLLECTION

C1. SAMPLE DESIGN

Sample size and school selection

While the program does not work through schools but directly with students and their caregivers, the sampling frame for data collection was created through schools. Schools were also the unit of randomization for treatment assignment.

During the evaluation design phase, NORC estimated that a sample of 2,400 students, specifically 30 students surveyed in 40 treatment schools and 30 students in 40 control schools, will allow estimation of an standardized effect size of 0.19, assuming an alpha of 5 percent, power level of 80 percent, an intracluster correlation coefficient of 0.1, and endline attrition of 20 percent.

The school sample frame was the list of all public schools in the districts of Lundazi and Chipata, which was provided by the corresponding District Education Board Secretaries (DEBS). These lists contain information on the number of students by grade. There are 210 schools in Chipata and 140 in Lundazi, for a total of 350 schools. We matched school pairs using the number of students and distance to DEBS for each school to make sure that, at least across these two characteristics, schools in the treatment and control groups were observationally equivalent.

We then randomly chose 40 of these school pairs and randomly assigned one school from each pair to the treatment group and the other to the control group.[20] Note that although the program interventions are not school-centered, to participate in the program children need to be enrolled in school.

Student Sample

Given that the MA intervention was centered on the delivery of text messages and voice recordings to parents and caregivers, participation in the MA intervention required access to a working cell phone. Zambian school rosters do not contain information on the cell phone ownership of parents. Therefore, the student sample frame (that is, the list of all eligible students from which to sample in treatment and control schools) was not readily available and needed to be constructed as part of the data collection effort.

To construct the sample frame, in October 2015 we sent Advanced Teams (AT) of enumerators to the field to survey all students in grades 1 and 2[21] across all 80 sampled schools to determine whether anyone in their home had a cell phone. In total, the ATs attempted interviewing 8,681 students, out of which 4,910 (56.5 percent) reported having a cell phone at

---

[20] To construct the school pairs we followed these steps: i) Calculate quartiles of the number of students and terciles of the distance to DEBS, and use these variables to divide the sample in twelve blocks or cells, ii) Randomly select 40 school pairs in proportion to the total number of schools in each cell, and iii) One school from each of the 40 pairs is randomly assigned to the treatment and the other to the control group, resulting in a sample of 40 schools in the treatment group and 40 schools in the control group.

[21] Because the sample was constructed at the end of the school year (October 2015), we interviewed students in grades 1 and 2.

home. 2,354 students (27.1 percent) were absent during the AT visit, so the cell phone ownership rate of 56.5 percent can be considered a lower bound of the true cell phone ownership rate.

After the student sample frame was constructed, enumerators randomly selected the sample of students in each school (15 from each grade and 10 replacements from each grade—for a total of 50 students per school). Enumerators left notes with the head teacher to distribute to the parents of the selected students, informing them about the possibility of being contacted in the following weeks regarding the survey and for participation in the MA program.

Parents were asked by the head teacher to return the notes within a week of receiving them. Follow-up cars visited all schools a week after notes were distributed to pick up returned notes. Returning the note was not required to participate in the survey (or the program). For the caregiver survey, enumerators had to locate and interview parents of sampled children even if they had not returned the note.

C2. BASELINE DATA COLLECTION

The baseline caregiver data collection occurred between mid-November and early December 2015. The caregiver questionnaire included a series of screener questions, intended to determine the respondent's eligibility to participate in the study. These screening criteria were agreed upon by NORC, Creative, USAID and World Vision, and included questions about the respondents' interest in participating in the intervention and the respondents' access to a working cell phone number. The same screening process was conducted in treatment and control groups to guarantee balance between the two groups across these dimensions. Respondents did not know to which group they were going to be assigned.

In the case that multiple children from the same household were sampled, household-level questions were only asked once, and only child-specific questions were asked twice. 2,222 household interviews were completed, corresponding to a total of 2,397 students.

In January 2016, enumerators visited schools to conduct the EGRA and the student survey, examining 2,260 students. Further baseline data collection details, including instrument design training, fieldwork and data quality review procedures, can be consulted in NORC (2016a).[22] All instruments were reviewed by a local IRB (ERES Converge) as well as by NORC's IRB.

C3. UPTAKE SURVEY

A subsample of caregivers from the treatment group was surveyed in June 2016 to measure program uptake; that is, the extent to which parents were receiving the messages and sharing the content with their children. Data on other implementation aspects were also collected.

NORC estimated that approximately 283 caregivers from the baseline sample would need to be surveyed at midline in order to measure a take-up of 40 percent with 95 percent confidence and a margin of error ±5 percent. We randomly selected 11 of the 40 treatment schools in our

---

[22] While this report contains below previously undocumented descriptions and details about the endline data collection, to avoid duplicative reporting, we have opted not to include a similar description about training, data quality control, and fieldwork at baseline, since these components are documented in the Baseline Report for this evaluation.

sample, encompassing 293 caregivers interviewed at baseline, and attempted to survey all of them.

NORC and the Institute of Economic and Social Research at the University of Zambia (INESOR), with input and feedback from Creative and USAID, designed an uptake survey to be administered to caregivers. The instrument included questions on whether households received SMS, the frequency at which they open and share the SMS with their children, attendance at community meetings, and other questions related to program implementation and uptake.

The caregiver uptake survey took place in June 6-10, 2016. 267 interviews were successfully completed. The main results can be consulted in NORC (2016b).

C4. ENDLINE DATA COLLECTION

Endline data was collected between mid-January and early February of 2017. Instruments administered for the endline data collection were:

1. **A Parent/Caregiver Household survey**. The purpose of this survey was to collect data on how much time parents spend reading with their child, how much time the child reads on her/his own, whether the child participates in reading activities after class in the community, and participation in MA activities (reading the SMS stories and attendance to MA meetings).

2. **An Early Grade Reading Assessment (EGRA) and student questionnaire.** The purpose of these instruments was to evaluate students' reading abilities; in particular, letter sound identification, oral reading, reading comprehension and listening comprehension, as well as ask children about their reading habits (e.g. if they like to read, how often they like to read) and their participation in the MA program.

In addition, a Telephone Script for Contacting Parents was also designed. The purpose of the script was to standardize the process that the enumerators used to call the parents for (a) collecting verbal consent to interview their child, and (b) setting up a time and place to interview the parent/guardian of the child. Given realities encountered in the field, described below, this instrument was discarded and other measures were taken in order accomplish the objectives for which it was designed.

All data collection activities were carried out in partnership with the Institute of Economic and Social Research (INESOR) at the University of Zambia. NORC's partner School-to-School International (STS) conducted the enumerator training, data quality oversight, and a data quality review for the EGRA components.

The remainder of this Annex is a description of the activities undertaken for the endline phase of the evaluation. Specifically it relates to the following: questionnaire development and programming; training; fieldwork and quality control; data quality review; challenges and lessons learned.

1. **Questionnaire Development**

For purposes of comparability, the same EGRA tool used for the MA baseline was used for the endline data collection. Since one year had passed since the children had last seen the tool at baseline, there was no need to re-shuffle the items within timed subtasks (a practice normally done to avoid having children remember the order from the last time they were administrated the tool). The student questionnaire, which immediately followed the EGRA, was developed by NORC based on the MA impact evaluation research questions and programmed in Tangerine by STS. The student questionnaire was pre-tested with approximately 5-10 children participating in the MA program during the week of enumerator training. The structure of the EGRA plus Student Questionnaire is presented in table C1.

Table C1: Sections of the Student Interview

Subtask	Title	Content
0	**Introduction and Consent**	Assent to participate in study
1	**Orientation to Print**	(Untimed) 3 questions. Indicate where one begins reading printed text on a passage and the direction one would read text.
2	**Letter Sound Identification**	(Timed) Produce sounds of 100 letters presented in written form. Presented in a grid of 10 rows and 10 columns.
3	**Non-Word Decoding**	(Timed) Sound out, or decode, unfamiliar words. Asked to read out 50 words without meaning.
4a	**Oral Reading Fluency**	(Timed) Read a passage of narrative text of ~40 words in length.
4b	**Reading Comprehension**	(Untimed) 5 questions. Respond to five questions asked about above passage.
5	**Listening Comprehension**	(Untimed) 5 questions. Oral response to listening comprehension questions.
6	**Student questionnaire**	Reading practices at home: both alone and with others. Frequency of reading and reading preference. Student attendance at school and teacher attendance at school. Level of involvement with MA program.

The endline parent/caregiver survey was adapted from the baseline and midline surveys conducted with parents. The home literacy environment components from the baseline survey were combined with the program uptake section from the midline survey to form a comprehensive endline survey. The final version was a collaboration between NORC, Creative, and USAID. Due to the fact that nearly all questions had been administered at an earlier phase in the evaluation, pre-testing of the questionnaire was not done. The sections of the parent/caregiver household survey are described in Table C2.

Table C2: Sections of the parent/caregiver survey

Section	Title	Content
A	**Introduction and Consent**	Consent to participate in the study Consent to administer EGRA to child
B	**Home Literacy Environment**	Frequency at which parents read with child Time and frequency child reads on his/her own Participation in community reading activities Parent attitudes and confidence for helping child learn to read
C	**Participation in MA**	Level of involvement with: SMS, audio recordings of stories, meetings, how the materials are used at home and by whom, suggestions for program improvement
D	**Program Substitution**	Other SMS programs to which the households are subscribed

As many of the survey questions were taken from past rounds of data collection, translation was minimal. INESOR completed translation of new questions and verified the past translations of previous work. To ensure high quality data, NORC deployed tablet data collection. Both English and Chinyanga versions of the parent questionnaire were programmed into the Nfield application by NORC's programming consultant, Moses Gitau. NORC and INESOR tested the instruments extensively on tablets prior to its deployment in the field. All instruments were reviewed by a local IRB (ERES Converge) as well as by NORC's IRB.

2. Training

The training was held for a period of eight days in Chipata from January 7th to January 14th, 2017. A total of 45 trainees attended the training out of which 8 supervisors and 32 enumerators were selected. The training was led by Alice Michelazzi, Sr. Evaluation Specialist and Kayla Nachtsheim, Data Capture Specialist of School-to-School International (all EGRA sessions) and Stacy Pancratz, survey specialist of NORC (all questionnaire sections) with support from INESOR. An initial "bootcamp" was conducted for new enumerators (enumerators that had not participated in the baseline data collection) on the first day of the training in order to familiarize them to EGRA rules and procedures as well as to tablet use. Starting from the second day of training, all enumerators worked together on EGRA and questionnaires (student and parent).

For the training of EGRA subtasks, the various subtasks goals and rules were presented in English to the group and then practice was conducted in the local language (Chinyanja). The student and parent questionnaires were first reviewed on paper to familiarize the team with the components and skip patterns of the survey. Afterwards, the full team reviewed the tablet versions of the surveys question by question, establishing a consensus on what data was meant to be captured by each question and the best translation into Chinyanja to do this (as a verification of the translation already completed prior to training).

The same marking conventions established at baseline were adopted at endline in order to ensure comparability of results. In particular:

- In the non-word reading and the oral reading fluency subtasks enumerators were prompted to consider syllabicated words as wrong.

• "L" and "R" letter substitution were marked as correct in non-word reading and oral reading fluency subtasks.

• In the oral-reading comprehension and listening comprehension sub-task, enumerators were prompted to consider as wrong, answers given in a language different than Chinyanja if the pupil did not self-correct after being prompted to answer in Chinyanja

During the review of the student questionnaire, enumerators that had participated in baseline activities shared that the show card with the three smiley faces was unclear to students and explaining the faces and their significance to the children was time consuming. With this feedback, NORC decided to drop the use of the show card and instead administer the response options verbally to the child.

Practice of EGRA and the questionnaires continued throughout the week including demonstrations by pairs with review following each subtask. Three enumerator agreement exercises were conducted during the training to determine the participants' inter-rater reliability (IRR). A quiz took place on day six of the training. A full agenda for the training is in Table C3.

Table C3: Endline Data Collection Training Schedule Fieldwork

Day	Activities Covered
Day 1	**Bootcamp for new enumerators** Introductions Summary of Makhalidwe Athu Project What is EGRA? Tablet use
Day 2	Overview of the enumerators' manual, principles of data collection (NORC) Orientation to Print subtask Letter Sound Identification subtask Non-Word Decoding subtask
Day 3	Oral Reading Passage subtask Oral Reading Comprehension subtask Listening Comprehension subtask EGRA Practice Student Questionnaire (NORC) IRR
Day 4	EGRA and Student Questionnaire practice Enumerator Agreement Practice
Day 5	Parent Questionnaire (NORC) Teams and logistics Calling procedures (NORC)
Day 6	School Practice Feedback from school practice QUIZ

Day	Activities Covered
Day 7	Review of Quiz Roles during data collection EGRA practice
Day 8	EGRA on paper (with selected enumerators) How to provide feedback Role of supervisors Preparation of materials

Field work started on the 16th of January, 2017 and ended on the 4th of February, 2017. We had a total of eight teams with each team consisting of a supervisor and four enumerators. There were four teams in Lundazi and another four in Chipata. A total of 80 schools were visited in Chipata and Lundazi districts.

This was the first time that the field teams were prepared to administer the parent survey and the EGRA/student questionnaire during the same field visit (as opposed to two field visits during the baseline). Overall the combined approach proved to be more cost-effective and efficient. Interviewing the parents first and seeking their consent immediately prior to interviewing the child also gave a sense of assurance to the children.

In January and February 2017, enumerators visited schools and households to conduct the EGRA and parent and student surveys. Two days prior to a team's visit to a school and community, enumerators were assigned to call or text parents notifying them of the visit and asking them to be at school on that day with their child. Reaching parents by phone proved to be difficult. Phones could not be reached because they were turned off, the network was bad, or the parent did not answer. Learning this issue early on, the teams adapted, asking any parent they were able to reach by phone to spread the word throughout the community of the team's visit. The MA program households were usually well aware of the other participating households. From previous rounds of data collection as well as social networks of the field team, teams were also able to call and notify head teachers prior to the school visit.

The one-year time lapse between the baseline and endline had a direct impact on the response rate, as demonstrated in Table 1C4, resulting from attrition for various reasons. In total 1,942 caregivers were surveyed at baseline and endline. These 1,942 caregivers correspond to 2,091 children (there are more students than caregivers because 149 siblings were also interviewed). Table 1C4 shows final disposition codes for the caregiver data at the children level. Out of these 2,091 children for which there is caregiver data, 2,054 were interviewed for the EGRA and student survey at endline. Along these lines, the completion rate for the caregivers was 87 percent, and for the students 86 percent.

Table C4. Endline Caregiver Interview Completion Rates

	N	Percent
Completed	2,091	87.2
Not locatable	231	9.6
Refused	28	1.2
Child transferred/dropped	4	0.2
Other	43	1.8
Total	2,397	100

Source: Own calculations using MA endline data

The STS and NORC team followed field work at the schools for 5 days; one or two teams were visited each day. During the field visits STS EGRA specialists and NORC survey specialist conducted observations on:
- EGRA administration: each enumerator was observed once and feedback was provided. When an enumerator produced a sub-par performance, STS conducted a second observation and feedback was shared with team supervisors to ensure ongoing monitoring.
- Space setting and organization of activities at the school
- Use of materials by enumerators (EGRA stimuli, tablets, adequate provisions of incentives for children being assessed, supervisor control sheets, enumerator control sheets)
- EGRA home visits

3. Data Quality Review

The STS team worked with field data manager hired by INESOR, Vincent Kapotwe, to monitor data quality. Daily tasks included:
- Cross-checking between downloaded data, supervisor control sheets, and the student sample.
- Monitoring disposition codes were correctly added to all students in the original student sample.
- Recording invalid cases and duplicates to be deleted.

Ms Nachtsheim, the data capture specialist from STS worked with INESOR field data manager to establish a daily data intake process.
- Master tracker: This document was used to record the number of assessments conducted by school. This document was also used to track the number of disposition codes that supervisors added in Tangerine for students who were not assessed due to illness, absence, disability, parent refusal, not locatable at school or home, or for other reasons. Any changes to the codes that supervisors entered (for example if a child marked as not assessed for "other" reasons but was actually ill) were also included in this document.
- Daily issues tracker: This document was used to record duplicate cases due to assessor errors that needed to be deleted, as well as other errors that needed to be addressed during data cleaning.

4. Challenges and Lessons Learned

While overall the fieldwork can be described as a success, there were a number challenges, lessons and opportunities for improvement. During field work, the following observations were made:

<u>Long distances and bad roads</u>: Some schools are located far from Lundazi and Chipata towns and require travel on poorly maintained roads.

<u>Solution</u>**:** The teams used 4x4 vehicles to navigate the poor conditions of the roads. It is also advisable that in the future, fieldwork should be undertaken during the dry season, which occurs May through October each year.

<u>Farming activities</u>**:** During the rainy season, most parents are occupied with tending to their field during the morning, requiring their child or an enumerator to find them there for the interview. This cost the child or the enumerator significant amount of time.

<u>Solution</u>**:** The team should consider undertaking fieldwork during the dry season. During this period farming activities are not so intensive so it is easier for the parent to participate in the survey.

<u>Poor phone network</u>**:** A number of schools were located in areas with very poor phone reception; therefore communication by phone was difficult. This affected making appointments with parents and even head teachers.

<u>Solution</u>**:** The solution was two-fold:

1. Where the team was able to reach the head teacher, he or she was asked to contact the parents and notify them of the survey team's visit date;

2. Where even the head teacher was impossible to contact, the teams could physically pass through the school on the way back from another school visit, or the teams used contacts from office of the District Education Board (DEB) to send messages through the school network.

<u>Timing/School Calendar</u>: Field work started on the opening day of school, hence there was poor turnout at the first schools visited. A number of non-interviews were due to the fact that the children were still on holiday.

<u>Solution</u>**:** Fieldwork involving both learners and teachers should not start until after at least two weeks of the opening of the schools.

<u>Language Barrier</u>**:** Most of the pupils in Lundazi were at a disadvantage in terms of the Chinyanja language because they speak Tumbuka at home and are also taught in Tumbuka at school.

<u>Solution</u>**:** The language barrier in most Lundazi schools should be put into consideration of future questionnaire translations. Enumerators with Tumbuka skills should be allocated to visit these schools.

Recommendations of practices to continue in possible future waves of data collection:

- The involvement of school teachers as enumerators should be promoted in order to have enumerators with enough experience working with children and with the Zambian new curriculum for reading (i.e. letter sounds). The teachers were also knowledgeable about the school system and often helped set school appointments by tapping into their social networks.

ANNEX D. UPTAKE REPORT

USAID Impact Evaluation of the *Makhalidwe Athu* Project (Zambia)
Uptake Survey Report

Background

The *Makhalidwe Athu* project (MA) is an 18-month intervention aimed at improving the reading skills of 1,200 students in 2nd and 3rd grade in the Chipata and Lundazi districts of Zambia's Eastern province. The project, funded by the All Children Reading Partners (USAID, World Vision and the Australian Government), and implemented by Creative Associates, provides reading materials in ChiNyanja (the predominant local language) and supports reading activities through SMS text messaging. Participant households receive three text messages on their mobile phones each week, transmitted on Mondays, Wednesdays and Fridays. These three messages comprise a short story (e.g. 160 characters each) for children to read with their families. Children have been provided a notebook and are encouraged to transcribe the stories. In addition, participants can call in for a pre-paid recorded voice message (IVR), which includes comprehension questions, as well as a recording of the story itself.

Following an initial orientation meeting with participants about the program and how to access the SMS stories and IVR, community mobilizers, aided by community volunteers, meet monthly with participants to discuss and address any issues that have arisen, both technological and how to read and listen to the stories with their children. Each Saturday, the local radio station broadcasts a "variety show" on reading, which features the story of the week, as well as call-in questions and presentations by local experts.

Stories are "crowdsourced" from the local communities, which have learned about the program through local media and community meetings. Local reading specialists then "level" the stories for Grade 2 and 3 readers and to fit the SMS format and at the same time making them culturally relevant and in line with the school curriculum.

NORC at the University of Chicago is conducting the impact evaluation of MA. Baseline data was collected in December 2015 and January 2016 on reading resources at home, reading habits, and reading skills, as measured by the Early Grade Reading Assessment (EGRA). A subsample of treated households was surveyed in June 2016 to measure program uptake; that is, the extent to which parents are receiving the messages and sharing the content with their children. Data on other implementation aspects was also collected. Endline data will be collected in January 2017.

In this report we describe the sample design and instruments fielded in June 2016, the data collection process and the main results. Baseline results are presented in a separate report.

Sample design and instruments

During the design phase of the impact evaluation, NORC conducted power calculations to establish the number of households that would need to be visited at midline, in order to measure take-up of the program with 95 percent confidence and a margin of error ±5 percent. Take-up can be defined as the fraction of households receiving the SMS, or the fraction of households opening and showing the SMS to their children at least once a week. Assuming a 40 percent take-up rate, NORC estimated that approximately 283 households would need to be surveyed at midline, given the parameters just discussed. NORC randomly selected 11 of the 40 treatment schools in our sample, encompassing 293 households interviewed at baseline, and planned to survey all of them.

NORC and the Institute of Economic and Social Research (INESOR), its local data collection partner, designed a caregiver instrument to be administered in households. Feedback from Creative and USAID was also incorporated in the survey. The instrument included questions on whether households received SMS, the frequency at which they open and share the SMS with their children, attendance to the community meetings, and other questions related to program implementation and uptake.

Data Collection

The parent uptake survey took place in June 6-10, 2016 in Chipata and Lundazi, Eastern Province, Zambia. Two teams of 4 enumerators and 1 supervisor each were deployed, one to Chipata and one to Lundazi; on average, each team covered one school per day. Cell phone numbers were used to contact caregivers one day in advance to schedule an interview time. The instrument instructs enumerators and respondents to identify "the person in the household most knowledgeable about the Makhalidwe Athu Program or who participates most in the Makhalidwe Athu Program. This is typically the person who attends monthly meetings about Makhalidwe Athu, helps the child read the SMS stories and listen to the voice messages, and discusses the stories with the child."

Main results

Out of 293 target households, complete interviews were conducted in 267 (91 percent) of them. Table 6 shows the reason why enumerators could not survey the remaining 26 households.

Table 6.   Survey results

	Count	Percentage
No one home after third visit	1	0%
Most knowledgeable person unavailable after third visit	2	1%
Refused to participate	6	2%
Most knowledgeable person absent for extended period	17	6%
Complete	267	91%
N	293	

Source: MA Midline Survey

Meetings

Of the respondents that completed the interview, 95 percent (254) said that they had attended at least one Makhalidwe Athu (MA) meeting or training session. Of those, 93 percent (235 cases) said that someone in the household attended the first meeting of the year. All respondents that attended the meetings found them to be very useful. Figure 1 shows a breakdown of what respondents learned at MA meetings.

Figure 1.   What was learned at MA meeting? (N=254) (Multiple responses allowed)

Category	Percentage
How to help my child read the SMS stories	73%
How to help my child write the SMS story in a notebook	42%
How to access SMS stories on my phone	32%
What is expected of me/the household and student	23%
What the program will do	23%
Other	12%
How much time I should be spent helping my child read the stories	12%
How to beep for a voice recording of the stories	9%
How and where to submit stories	6%
How many times a week the household will receive the SMS stories	5%

Interruption of cell phone service

Forty-eight percent of interviewees (128 respondents) reported that their phone broke or had at least one interruption of service since the start of the school year. Six percent (15 respondents) changed their phone number/SIM cards. Six of these parents did not receive any Makhalidwe Athu text messages. Enumerators reported that these parents did not always know how to communicate the new number to Makhalidwe Athu.

SMS Messages

Ninety-two percent of interviewees (245 respondents) said they received stories via SMS. The remaining tables all concern only this subset of the sample. Table 7 shows that 61 percent of the households that received SMS always have a clear phone reception.

Table 7. How often does the respondent have clear phone reception?

	Count	Percentage
Always	149	61%
Usually	63	26%
About half the time	13	5%
Seldom	16	7%
Never	4	2%
N	245	

As Table 8 shows, seventy-six percent of the respondents that received SMS received three messages a week. The 2 respondents that answered 'Don't know' to the number of days they received SMS are not included in the subsequent tables.

Table 8. Number of days a week messages are received from MA

	Count	Percentage
One	12	5%
Two	43	18%
Three	185	76%
Four	3	1%
Don't know	2	1%
N	245	

Only one household answered they never could open the SMS (Table 9). The subsequent tables pertain only to the 242 households that received and were able to open the SMS.

Table 9. How often can someone in the household open messages?

	Count	Percentage
Can open every message	210	86%
Only 2 out of 3 messages	24	10%
Only 1 out of 3 messages	8	3%
Can never open messages	1	0%
N	243	

Table 10 indicates the frequency at which someone in the household shows messages to the child by whether or not the mother knows how to read. Perhaps surprisingly, it is more likely that a child would see every message if the mother does not know how to read than if she does, but fractions in both groups are pretty high. When the sample is split by whether the father can read (Table 11), there seems to be no strong relationship. In Table 12 frequency at which messages are shown to child by whether the oldest sibling knows how to read are displayed; in this case, literacy of the oldest sibling and frequency at which messages are shows is positively correlated.

Table 10. Frequency at which messages are shown to child by whether or not the mother can read.

	Mother can read (N=74)	Mother can't read (N=137)
Every message (3 times/week)	69%	80%
Two messages a week	26%	15%
One message a week	5%	4%
Less than once a week	0%	0%
Never	0%	0%

Note: To measure whether the mother can read we use to the baseline data. The respondent of the uptake survey is not necessarily the mother.

Table 11. Frequency at which messages are shown to child by whether or not the father can read.

	Father can read (N=149)	Father can't read (N=46)
Every message (3 times/week)	76%	76%
Two messages a week	21%	15%
One message a week	3%	9%
Less than once a week	0%	0%
Never	0%	0%

Note: To measure whether the father can read we use to the baseline data. The respondent of the uptake survey is not necessarily the father.

Table 12. Frequency at which messages are shown to child by whether or not the oldest sibling can read.

	Oldest sibling can read (N=149)	Oldest sibling can't read (N=130)
Every message (3 times/week)	84%	73%
Two messages a week	15%	20%
One message a week	2%	7%
Less than once a week	0%	0%
Never	0%	0%

Note: To measure whether the oldest sibling can read we use to the baseline data. The respondent of the uptake survey is not necessarily the oldest sibling.

As Figure 2 demonstrates, children's mothers read the SMS with the child the most, but fathers and siblings participate often as well.

Figure 2. Who participated in reading messages with the child? (N=242) (Multiple responses allowed)

Mother	51%
Father	45%
Sister	43%
Brother	42%
Uncle	14%
Children from outside household	9%
Aunt	7%
Grandmother	7%
Other	6%
Cousin	6%
Grandfather	5%
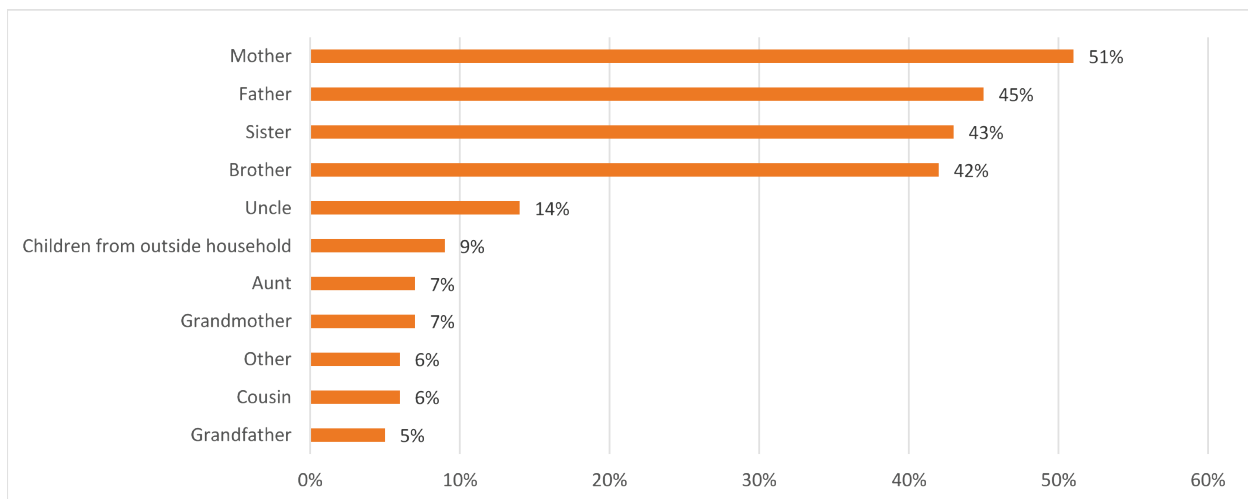
Table 13 shows how reading happens by whether or not the mother knows how to read. There are no clear differences in how reading happens by mothers' literacy. Most frequently the respondent (which is not necessarily the mother) reads half of the story and the child the other half. The second most frequent way of reading the stories is that the child does all the reading. Similar results can be observed when the sample is divided by whether the father knows how to read (Table 14), or whether the oldest sibling knows how to read (Table 15).

Table 13. How reading happens by whether or not the mother can read.

	Mother can read (N=61)	Mother can't read (N=124)
I read and child listens most of the time	3%	12%
I read half the time and child reads the other half	59%	53%
I listen and child reads most of the time	36%	34%
I am not present when child reads messages	2%	1%

Note: To measure whether the mother can read we use to the baseline data. The respondent of the uptake survey is not necessarily the mother.

Table 14. How reading happens by whether or not the father can read.

	Father can read (N=132)	Father can't read (N=41)
I read and child listens most of the time	11%	5%
I read half the time and child reads the other half	55%	61%
I listen and child reads most of the time	32%	34%
I am not present when child reads messages	2%	0%

Note: To measure whether the father can read we use to the baseline data. The respondent of the uptake survey is not necessarily the father.

Table 15. How reading happens by whether or not the oldest sibling can read.

	Oldest sibling can read (N=51)	Oldest sibling can't read (N=112)
I read and child listens most of the time	2%	12%
I read half the time and child reads the other half	55%	54%
I listen and child reads most of the time	39%	34%
I am not present when child reads messages	4%	1%

Note: To measure whether the oldest sibling can read we use to the baseline data. The respondent of the uptake survey is not necessarily the oldest sibling.

Notebook

As shown in Table 16, the majority of children use the provided notebook to record stories. While there is a correlation between knowing how to read (reported at baseline by the child) and recording the stories, even among the students who reported they did not know how to read the fraction recording the stories in their notebooks is very high (76 percent). Note that half a year has passed since data on children's literacy was collected, so it is possible that

children that were not able to read in January were able to do some writing when the midline data was collected.

Table 16. Does the child use the provided notebook to record stories (by whether or not the child knows how to read)?

	Child can read (N=109)	Child cannot read (N=126)
Yes	84%	76%
No	9%	17%
Did not receive notebook	7%	7%

Voice Recordings

Reports from the field indicate that several parents either were unaware of the audio file option or did not know how to access the audio file. When teams discovered this, they communicated with Angela Musonda, the Makhalidwe Athu project coordinator, to obtain the dial-in number for the audio file and shared with the parents. However, it was clear that these parents did not know that this number existed or how to access it.

Sixty-eight percent of respondents that received text messages (164 respondents) said they listened to the voice recording with their child. Table 17 shows how recordings are listened. In 72 percent of the households, recordings are listened after opening the SMS.

Table 17. How do you listen to the recording?

	Count	Percentage
Listen before opening SMS	20	12%
Listen after opening SMS	118	72%
Do not read SMS, just listen to recording	1	1%
Listen while looking at notebook where child copied the story	21	13%
Other	4	2%
N	164	

(percentages are of 164 respondents that listened to voice recordings)

Engagement

The uptake survey also asked how stories were discussed. Table 18 shows that half of the respondents answered that they have a conversation about the story, and the same fraction answered that they ask the child additional questions to test comprehension.

Table 18. How do you discuss the stories? (Multiple responses allowed)

	Count	Percentage
We talk about the comprehension questions	90	37%
I ask additional questions to test my child's comprehension	123	51%
We have a conversation about the story	125	51%
We do not talk about the stories at all	2	1%
Don't know	2	1%
N	243	

(percentages are of 243 respondents that received messages)

Table 19 shows data on who participates in these discussions. In the majority of cases (90 percent) the child participates in the conversation. Other children and other adults also participate frequently.

Table 19. Who participates in these discussions? (Multiple responses allowed)

	Count	Percentage
Child	215	90%
Other children in household	149	62%
Other adults in household	156	65%
Adults from outside household	3	1%
Children from outside household	5	2%
N	239	

(percentages are of 239 respondents that discussed stories)

Caregivers were asked how easy the stories were for the child to read, and how easy they found the stories for themselves to read. As Figures 3 shows, we find what could be called a 'normal' or bell-shaped distribution for children, as the majority of parents report their children find the stories neutral or easy to read, smaller fractions find the stories either very easy or difficult for the child to read, and only a very small fraction (3 percent) find the stories very difficult to read. For caregivers, on the other hand, the distribution seems to be bimodal, as the bulk of caregivers find the stories easy or very easy to read, but 10 percent find the stories very difficult to read.

Figure 3. How easy are these stories to read? (N=242)
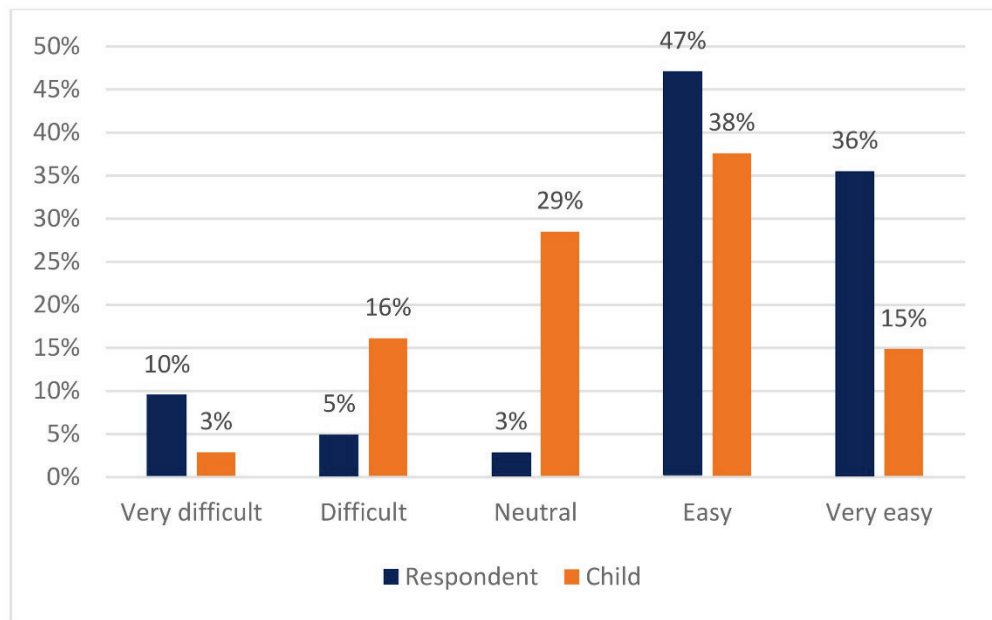
All caregivers said the program is either helpful or very helpful in supporting their children's reading. When asked how the program could be improved, substantial fractions of participants suggested scaling-up the program, either by increasing the number of stories (31 percent), or by providing the program for other grades (28 percent). Forty-one percent of the respondents provided 'Other' suggestions, many of those were simply to continue the program.

Figure 4. How would you suggest improving the program? (N=242) (Multiple responses allowed)
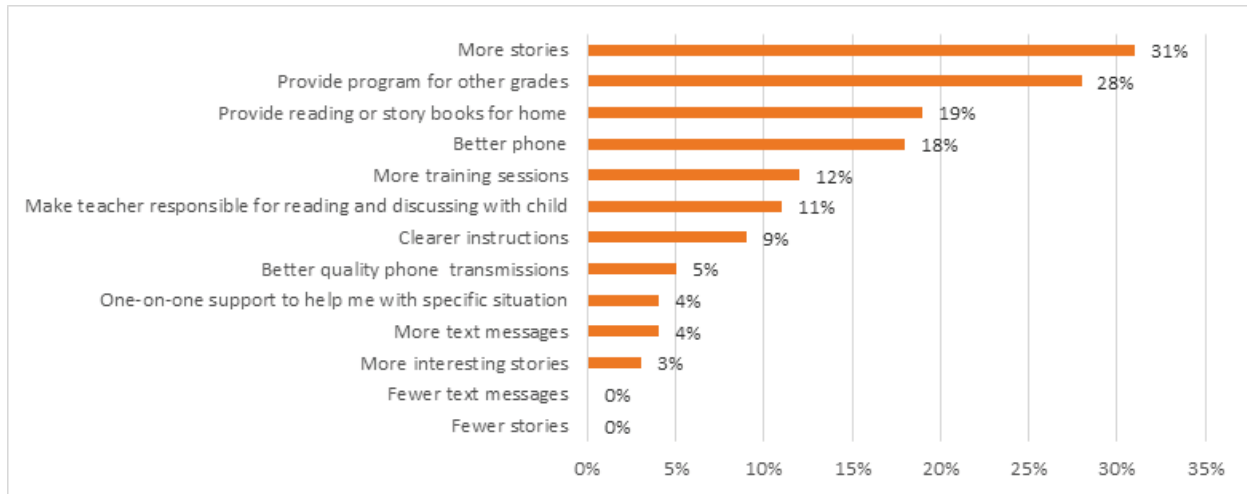
Radio programming

Twenty-one percent of interviewees who received stories (50 respondents) said they had submitted a story or idea to MA.

Seventy-nine percent of interviewees who received stories (192 respondents) said they knew about the reading program broadcasted each week by the local radio station.

Sixty percent of interviewees who received stories (145 respondents) said they had listened to a reading broadcast.

Key Takeaways
- Based on the findings of the uptake survey, MA seems to be operating as planned. Most households are receiving SMS (92 percent) and of these, 78 percent are receiving three SMS each week. Participation in the program is pretty high, all caregivers said they read SMS with the child at least once a week, and 78 percent that they do it every time an SMS is received. Also, 95 percent said that they had attended at least one MA meeting.

- The program is positively perceived by parents. All caregivers reported that MA helps their children learning process, and many household members are actively participating in the SMS reading activity.

- Forty-eight percent of parents reported that they had an interruption of cell phone service, while 6 percent reported changing their phone number. Several parents commented to enumerators that they were not sure how to communicate changes in phone numbers to MA. Establishing a clear system for communicating changes in SIM cards and phone numbers to caregivers would likely solve this problem.

- Sixty-eight percent of respondents that received text messages (164 respondents) said they listened to the voice recording with their child. While the uptake survey did not ask if parents received voice recordings or knew how to access voice recordings, reports from the field indicate that several parents either were unaware of the audio file option or did

not know how to access the audio file. When teams discovered this, they communicated with Angela Musonda, the Makhalidwe Athu project coordinator, to obtain the dial in number for the audio file and shared with the parents. However, it was clear that these parents did not know this number existed or how to access it. Making sure this information is widely distributed (even via SMS) to parents would likely help address this issue.

- The most commonly suggested recommendations parents had for improving the program included providing more stories (31 percent), providing the program for other grades (28 percent) and providing reading or story books for the home (19 percent).

ANNEX - ADDITIONAL FREQUENCY TABLES

Table A1. Respondent's relationship to child

	Count	Percentage
Mother	121	45%
Father	87	33%
Grandmother	15	6%
Grandfather	3	1%
Aunt	6	2%
Uncle	10	4%
Sister	6	2%
Brother	5	2%
Cousin	4	2%
Other	10	4%
N	267	

Table A2. Who else participates in MA activities with child? (Multiple responses allowed)

	Count	Percentage
Mother	107	40%
Father	102	38%
Grandmother	12	4%
Grandfather	16	6%
Aunt	15	6%
Uncle	29	11%
Sister	59	22%
Brother	60	22%
Cousin	8	3%
Other	17	6%
No one else	13	5%
N	267	

Table A3. Meetings' attendance

	Count	Percentage
Someone in household attended at least one MA meeting	254	95%
Someone in household attended the first MA meeting	235	88%
Respondent personally attended first MA meeting	203	76%
N	267	

Table A4. Who attended the first MA meeting, if respondent did not attend? (Multiple responses allowed)

	Count	Percentage
Mother	10	31%
Father	21	66%
Grandmother	0	0%
Grandfather	2	6%
Aunt	2	6%
Uncle	3	9%
Sister	2	6%
Brother	1	3%
Cousin	0	0%
Other	0	0%
N	32	

(percentages are of the 32 households where someone attended the first MA meeting, but not the respondent.)

Table A5. How prepared did you feel after this meeting to open and read SMS stories?

	Count	Percentage
Very prepared	139	68%
Somewhat prepared	61	30%
Not at all prepared	3	1%
N	203	

(percentages are of the 203 households where respondent personally attended the first meeting)

Table A6. Average number of additional MA meetings attended by someone in household

	Mean	N
Number of meetings	3.1	253

(percentages are of the 254 households where someone in household attended at least one meeting, minus a single "don't know" response)

Table A7. Whether respondent personally attended additional meetings

	Count	Percentage
I attended them	226	89%
Another household member attended them	27	11%
N	253	

(percentages are of the 254 households where someone in household attended at least one meeting, minus a single "don't know" response)

Table A8. Who attended other MA meetings, if respondent did not attend? (Multiple responses allowed)

	Count	Percentage
Mother	9	33%
Father	15	56%
Grandmother	0	0%
Grandfather	2	7%
Aunt	1	4%
Uncle	2	7%
Sister	2	7%
Brother	2	7%
Cousin	0	0%
Other	1	4%
N	27	

(percentages are of the 27 households where someone attended other MA meetings, but not the respondent.)

Table A9. What was learned at MA meeting? (Multiple responses allowed)

	Count	Percentage
What the program will do	59	23%
What is expected of me/the household and student	57	23%
How many times a week the household will receive the SMS stories	12	5%
How to access SMS stories on my phone	80	32%
What to do if we cannot receive SMS stories	6	2%
How to beep for a voice recording of the stories	23	9%
How to help my child read the SMS stories at home	184	73%
How much time I should dedicate every week to helping my child read the stories at home	31	12%
How to help my child write the SMS story in a notebook	106	42%
Who to contact if I have questions	3	1%
How and where to submit stories that I created or my ideas for stories	16	6%
That BreezeFM broadcasts a MA reading program every Saturday	2	1%
How I might receive a short survey on my phone asking me about the MA program	5	2%
Nothing	1	0%
I cannot remember	6	2%
Other	10	4%
N	253	

(percentages are of 253 households that attended at least one meeting)

Table A10. Usefulness of meetings

	Count	Percentage
Very useful	252	100%
Somewhat useful	1	0%
Not at all useful	0	0%
N	253	

(percentages are of 253 households that attended at least one meeting)

Table A11. Why did no one in household attend meetings?

	Count	Percentage
Did not know meetings occurred	5	36%
Did not think they would be useful	0	0%
Already knew information	1	7%
Did not have time	4	29%
Other	4	29%
N	14	

(percentages are of the 14 households where no one attended a meeting)

Table A12. Problems with phone

	Count	Percentage
Phone broke or had interruption of service since start of school year	128	48%
Phone number changed since start of school year	15	6%
N	267	

Table A13. Owner of cell phone used to received messages

	Count	Percentage
Mother	54	22%
Father	110	45%
Grandmother	12	5%
Grandfather	7	3%
Aunt	6	2%
Uncle	29	12%
Sister	2	1%
Brother	15	6%
Cousin	1	0%
Other	9	4%
N	245	

(percentages are of 245 households that received messages)

Table A14. Number of days a week messages are received from MA

	Count	Percentage
One	12	5%
Two	43	18%
Three	185	76%
Four	3	1%
Don't know	2	1%
N	245	

(percentages are of 245 households that received messages)

Table A15. Are you able to read and understand the SMS stories on your own?

	Count	Percentage
Yes	195	81%
No, I can read but don't understand	8	3%
No, I cannot read	39	16%
N	242	

(percentages are of 242 households that opened messages)

Table A16. Last week, what days of the week did you or someone else help your child read the SMS stories? (Multiple responses allowed)

	Count	Percentage
Monday	142	59%
Tuesday	103	43%
Wednesday	144	60%
Thursday	84	35%
Friday	138	57%
Saturday	84	35%
Sunday	64	26%
None of the above	27	11%
N	242	

(percentages are of 242 households that received messages)

Table A17. In general, how does reading happen?

	Count	Percentage
I read and child listens most of the time	20	9%
I read half the time and child reads the other half	116	54%
I listen and child reads most of the time	74	35%
I am not present when child reads messages	4	2%
N	214	

(percentages are of 214 households that read with child)

Table A18. Average number of minutes spent reading one SMS story with child, by gender of the child

	Mean	N
Number of minutes (boys)	21.6	93
Number of minutes (girls)	22.7	119
Number of minutes (all)	22.2	212

(Of 214 households that read with child, minus two "don't know" responses)

Table A19. What prevented you or someone else from reading with your child last week? (Multiple responses allowed)

	Count	Percentage
Did not have enough time in day/had other tasks to do	9	32%
Did not know how to use SMS or voice recording	1	4%
Did not receive/could not access the SMS story or voice recording	1	4%
Do not feel confident reading with child	3	11%
Did not understand the story well enough	0	0%
Child did not have enough time (e.g., had to work)	0	0%
No interest from child	0	0%
Too difficult for child	0	0%
Other	18	64%
N	28	

(percentages are of 28 households that read with child less than three days a week)

Table A20. Child uses the provided notebook to record stories, by gender of the child

	Boys	Girls	Total	Count
Yes	83%	80%	81%	196
No	10%	14%	12%	29
Did not receive notebook	8%	7%	7%	17
N	104	138	242	

(percentages are of 242 households that received messages)

Table A21. Who writes down stories in the notebook by gender of the child

	Boys	Girls	Total	Count
Respondent	21%	23%	26%	50
Child	61%	54%	65%	128
Other children in household	5%	8%	8%	15
Other adults in household	10%	13%	14%	27
Other	2%	2%	3%	5
N	98	127	196	

(percentages are of 196 households that used notebooks)

Table A22. When does the child write the story in their notebook?

	Count	Percentage
After every SMS story message	141	72%
After some SMS story messages	24	12%
After listening to voice recording	5	3%
Never, my child does not copy the story	26	13%
N	196	

(percentages are of 196 households that used notebooks)

Table A23. Who reads SMS stories from the notebook? (Multiple responses allowed)

	Count	Percentage
I do	78	40%
Child	176	90%
Other children in household	51	26%
Other adults in household	36	18%
Adults from outside household	3	2%
Children from outside household	3	2%
No one	1	1%
N	196	

(percentages are of 196 households that used notebooks)

Table A24. Someone in household listens to voice recording

	Count	Percentage
Someone in household listens to voice recording with child	165	68%
N	243	

(percentages are of 243 households that received messages)

Table A25. How do you listen to the recording?

	Count	Percentage
After every SMS message (3 times/week)	111	67%
After we receive two SMS messages (2 times/week)	33	20%
After we receive one SMS message (one time/week)	16	10%
Never	5	3%
N	165	

(percentages are of 165 households that listened to voice recordings)

Table A26. How do you listen to the recording?

	Count	Percentage
Listen before opening SMS	20	12%
Listen after opening SMS	118	72%
Do not read SMS, just listen to recording	1	1%
Listen while looking at notebook where child copied the story	21	13%
Other	4	2%
N	164	

(percentages are of 164 households that listened to voice recordings)

Table A27. How do you discuss the stories? (Multiple responses allowed)

	Count	Percentage
We talk about the comprehension questions	90	37%
I ask additional questions to test my child's comprehension	123	51%
We have a conversation about the story	125	51%
We do not talk about the stories at all	2	1%
N	243	

(percentages are of 243 households that listened to voice recordings)

Table A28. How often do you discuss stories with the child?

	Count	Percentage
Almost every day	127	52%
After every SMS message (3 times/week)	69	29%
After we receive two SMS messages (2 times/week)	35	14%
After we receive one SMS message (one time/week)	10	4%
Less than once a week, but at least once a month	1	0%
N	242	

(percentages are of 242 households that discussed stories)

Table A29. How easy are the stories for the child to read, by gender of the child

	Boys	Girls	Total	Count
Very difficult	4%	2%	3%	7
Difficult	17%	15%	16%	39
Neutral	26%	30%	29%	69
Easy	39%	36%	38%	91
Very easy	13%	16%	15%	36
N	104	138	242	

(percentages are of 242 households that received messages)

Table A30. How easy are the stories for the child to understand, by gender of the child

	Boys	Girls	Total	Count
Very difficult	2%	1%	1%	3
Difficult	7%	7%	7%	17
Neutral	19%	12%	15%	36
Easy	50%	53%	52%	125
Very easy	22%	28%	25%	61
N	104	138	242	

(percentages are of 242 households that received messages)

Table A31. How enjoyable does the child find the stories, by gender of the child

	Boys	Girls	Total	Count
Very enjoyable	70%	67%	68%	165
Enjoyable	28%	32%	30%	73
Somewhat enjoyable	1%	1%	1%	2
Not enjoyable at all	1%	1%	1%	2
N	104	138	242	

(percentages are of 242 households that received messages)

Table A32. How helpful would you say this program is in helping your child learn to read?

	Count	Percentage
Very helpful	214	88%
Helpful	29	12%
Somewhat helpful	0	0%
Not at all helpful	0	0%
N	243	

(percentages are of 243 households that received messages)

Table A33. Story submission and broadcast awareness

	Count	Percentage
Someone in household submitted a story to MA	50	19%
Know that BreezeFM broadcasts a reading program	192	72%
Have listened to BreezeFM reading program	145	54%
N	267	

ANNEX E. FRACTIONS OF STUDENTS REACHING MEANINGFUL THRESHOLDS

To analyze the fraction of students that are able to reach certain meaningful reading thresholds, we use performance levels proposed by Ministry of Education, Science, Vocational Training and Early Education (2014). For 2nd grade non-word reading, the Ministry defines reading fewer than 10 words as 'below minimum', between 10 and 14 words as 'minimum', between 15 and 19 words as 'desirable', and 20 words or more as 'outstanding'. No thresholds were produced for 3rd grade so only results for 2nd graders are shown for non-word reading.

Figure E1 shows fractions of 2nd graders in treatment and control groups divided by non-word reading categories. The fraction of students in the control group that were not able to read a single word was 88 percent at baseline and 72 percent at endline. This decline was compensated by increases in the fractions of all other reading categories, even 'Outstanding' readers increased their participation, going from 1 percent at baseline to 3 percent at endline. The results for the treatment group are similar but more pronounced. While the decline in the fraction of students that were not able to read a single word in the control group was 16 percentage points, the decline for the treatment group was 24 percentage points.

Figure E1. Students in 2nd grade treatment and control groups by reading level – non-word reading
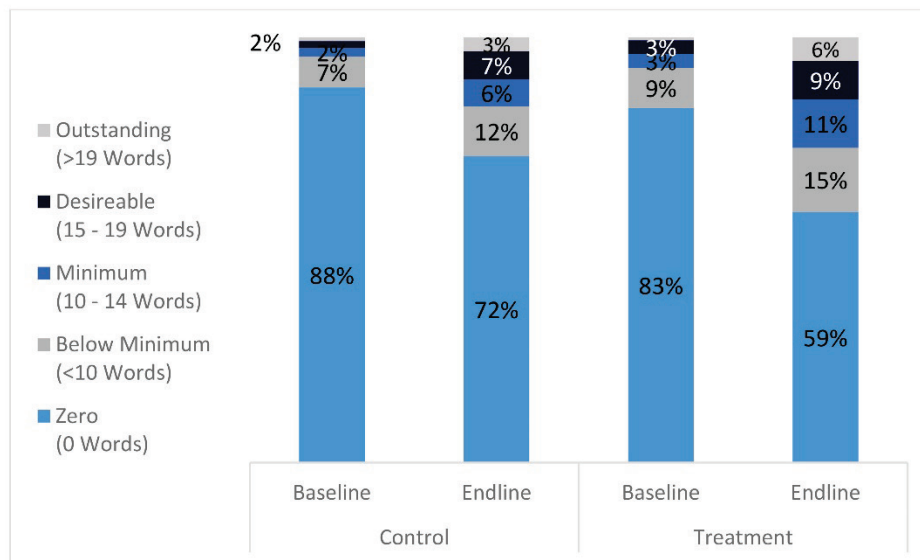
Note: Fraction labels of 1 percent or less are dropped from the graph for clarity.
Source: Own calculations using MA baseline and endline data and Ministry of Education, Science, Vocational Training and Early Education (2014).

For reading comprehension the Ministry of Education proposes the following thresholds for 2nd grade: Answering 1 or 2 reading comprehension questions correctly is defined as 'below minimum' , answering 3 questions correctly is 'minimum', answering 4 questions correctly is 'desirable', and answering all 5 questions correctly is 'outstanding'. For 3rd grade the thresholds are: One question correctly answered is defined as 'below minimum', 2 questions correctly

answered is 'minimum', 3 questions correctly answered is 'desirable', and answering 4 or 5 questions correctly is 'outstanding'. Note that thresholds are lower for 3rd grade than for 2nd grade, which is counterintuitive. The reason for this is that the Ministry of Education proposed these thresholds relative to grade-appropriate passages, so thresholds are not directly comparable between grades because they refer to passages that differ in complexity. However, for this EGRA we used the same passage for both 2nd and 3rd graders. Therefore the results look more favorable for 3rd graders than for 2nd graders, compared to what would be the case if two different passages, appropriate for each grade, had been used.

Figure E2 shows fractions of students in treatment and control groups divided by reading comprehension categories, by grade. For 2nd grade the fraction of students in the control group that were not able to answer a single question correctly was 91 percent at baseline and 76 percent at endline. This decline was compensated mostly by an increase in the fraction of 'below minimum' readers between baseline and endline. With respect to students in the treatment group, the decline in the fraction of students that were not able to read a single word was 22 percentage points, from 88 percent to 66 percent. This was compensated mostly by an increase of 'below minimum' and 'minimum' readers.

For 3rd grade the fraction of students in the control group that were not able to answer a single question correctly was 76 percent at baseline and 60 percent at endline. This decline was compensated by increases in the fractions of 'minimum', 'desirable' and 'outstanding' readers between baseline and endline. With respect to students in the treatment group, the decline in the fraction of students that were not able to read a single word was 30 percentage points (from 68 percent to 38 percent), almost twice the decline observed for the control group. This was compensated mostly by increases in the fractions of 'minimum', 'desirable' and 'outstanding' readers between baseline and endline.

Figure E2. Students in treatment and control groups by reading level and by grade – reading comprehension
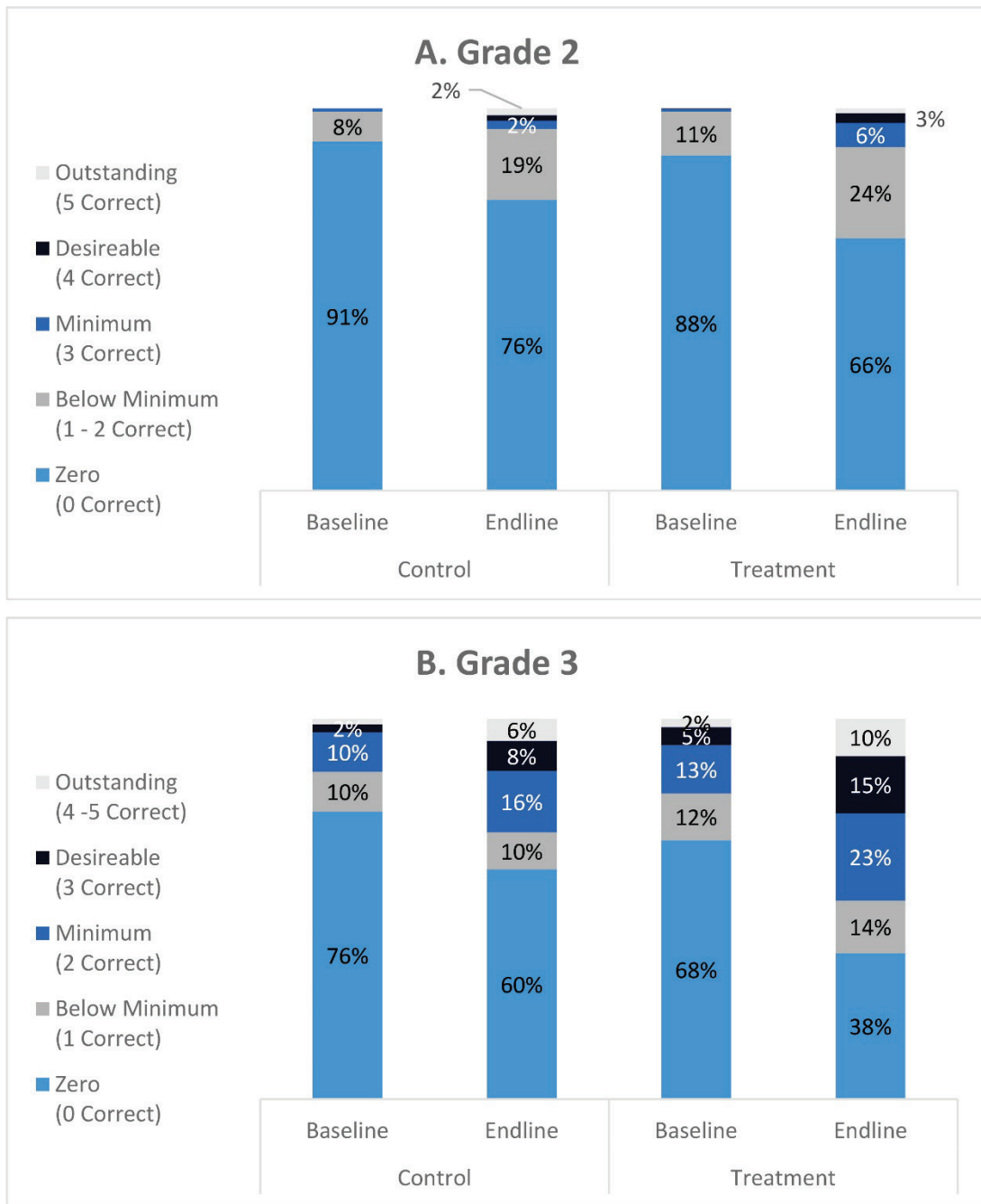
A. Grade 2

Legend:
- Outstanding (5 Correct)
- Desireable (4 Correct)
- Minimum (3 Correct)
- Below Minimum (1 - 2 Correct)
- Zero (0 Correct)

Control: Baseline (8%, 91%), Endline (2%, 2%, 19%, 76%)
Treatment: Baseline (11%, 88%), Endline (3%, 6%, 24%, 66%)

B. Grade 3

Legend:
- Outstanding (4 -5 Correct)
- Desireable (3 Correct)
- Minimum (2 Correct)
- Below Minimum (1 Correct)
- Zero (0 Correct)

Control: Baseline (2%, 10%, 10%, 76%), Endline (6%, 8%, 16%, 10%, 60%)
Treatment: Baseline (2%, 5%, 13%, 12%, 68%), Endline (10%, 15%, 23%, 14%, 38%)

Note: Fraction labels of 1 percent or less are dropped from the graph for clarity.
Source: Own calculations using MA baseline and endline data and Ministry of Education, Science, Vocational Training and Early Education (2014).

Finally, Figure E3 shows results for listening comprehension. For each grade the thresholds are the same as the ones for reading comprehension. It is apparent that students were much more likely to answer one or more questions correctly compared to their performance on the

reading comprehension subtask. It is also clear that no clear improvement was observed for this subtask over time or between treatment and control groups. This is consistent with the analysis on the means for this subtask discussed in the main body of this report. For grade 2 for the control group perhaps the main change was a relatively small decline in the fraction of 'below minimum' readers, compensated by mild increases in the fractions of 'minimum', 'desirable' and 'outstanding' readers. For the treatment group small decline observed in the fraction of 'below minimum' readers was compensated by an increase in the fractions of 'minimum' readers.

For 3rd grade the observed changes between baseline and endline are also small. For the control group the decline in the fraction 'below minimum' readers is compensated by an increase in the fraction of 'Outstanding' readers. For the treatment group, small declines in the fractions of 'below minimum' and 'minimum' are mostly compensated by an increase in the fraction of 'outstanding' readers.

Figure E3. Students in treatment and control groups by reading level and by grade – listening comprehension

A. Grade 2

Legend:
- Outstanding (5 Correct)
- Desireable (4 Correct)
- Minimum (3 Correct)
- Below Minimum (1 - 2 Correct)
- Zero (0 Correct)

Control – Baseline: 2%, 22%, 33%, 40%, 4%
Control – Endline: 5%, 24%, 36%, 33%, 2%
Treatment – Baseline: 5%, 24%, 32%, 36%, 3%
Treatment – Endline: 5%, 23%, 38%, 33%, 2%

B. Grade 3

Legend:
- Outstanding (4 -5 Correct)
- Desireable (3 Correct)
- Minimum (2 Correct)
- Below Minimum (1 Correct)
- Zero (0 Correct)

Control – Baseline: 32%, 38%, 20%, 8%, 2%
Control – Endline: 37%, 38%, 20%, 4%
Treatment – Baseline: 38%, 35%, 21%, 6%
Treatment – Endline: 42%, 36%, 18%, 3%

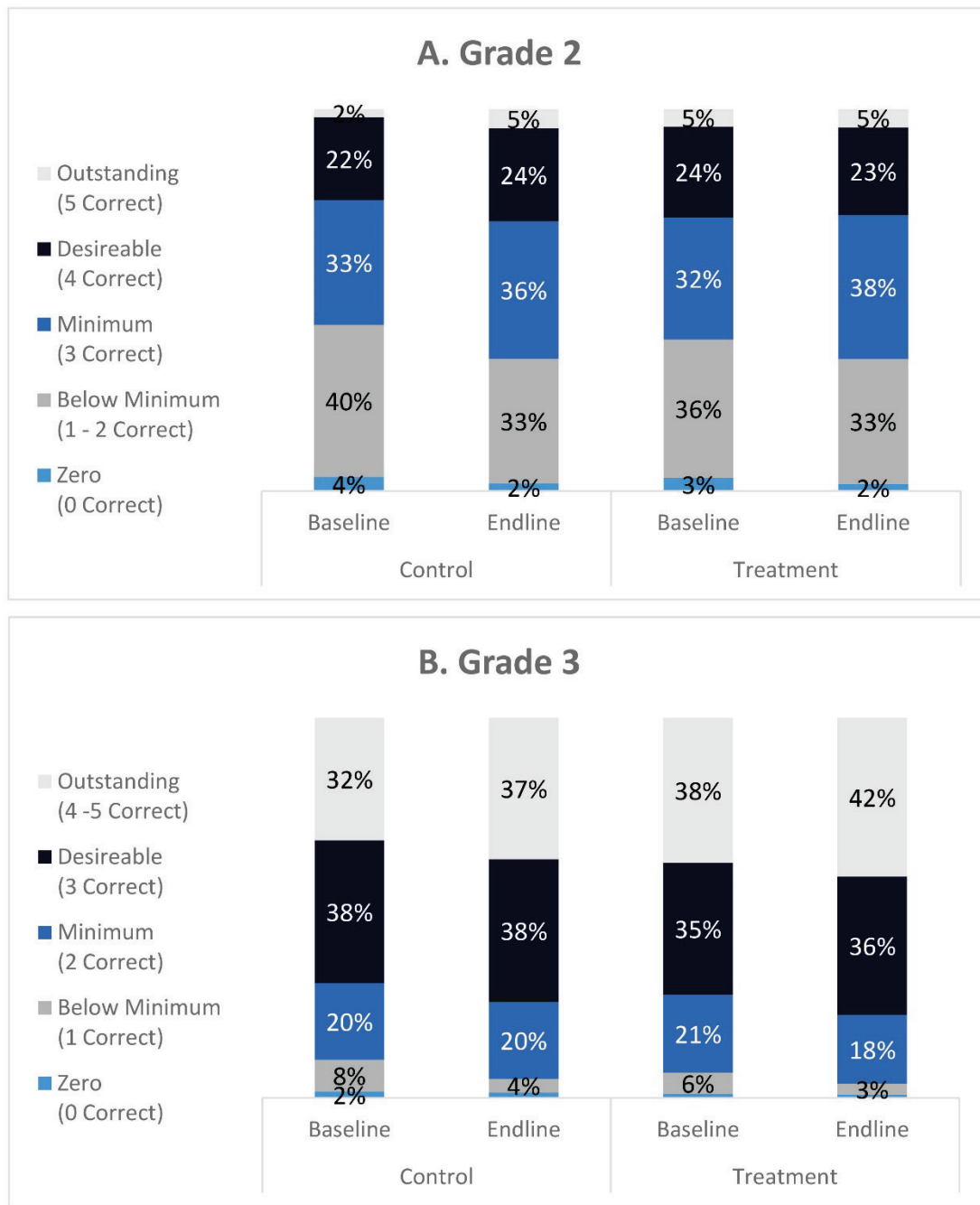Note: Fraction labels of 1 percent or less are dropped from the graph for clarity.
Source: Own calculations using MA baseline and endline data and Ministry of Education, Science, Vocational Training and Early Education (2014).

ANNEX F. DESCRIPTIVE STATISTICS

Table F1. Descriptive statistics at baseline by treatment status

	Control	Treatment	N
Child characteristics			
Age (years)	8.82	8.85	2037
Female	0.52	0.51	2054
Household characteristics			
Mother's literacy	0.54	0.58	1752
Father's literacy	0.79	0.79	1613
Household Size	5.39	5.25	2054
Household connected to electric grid	0.06	0.04	2053
Asset ownership			
Television	0.25	0.29	2054
Bed	0.52	0.55	2053
Bike	0.75	0.75	2054
Radio	0.60	0.62	2054

Source: Own calculations using MA baseline and endline data. Only observations for which panel data is available (households were surveyed both at baseline and endline) are included.

Table F2. Descriptive statistics by baseline ORF levels

	Zero	Below Minimum	Minimum	Desirable	Outstanding
Child characteristics					
Age (years)	8.7	9.0	9.1	9.4	9.3
Female	0.53	0.47	0.57	0.44	0.54
Household characteristics					
Mother's literacy	0.54	0.61	0.68	0.63	0.47
Father's literacy	0.77	0.84	0.80	1.00	0.82
Household Size	5.34	5.21	5.58	5.67	5.27
Household connected to electric grid	0.04	0.06	0.09	0.00	0.05
Asset ownership					
Television	0.24	0.35	0.32	0.44	0.24
Bed	0.50	0.61	0.58	0.56	0.48
Bike	0.76	0.73	0.80	0.44	0.68
Radio	0.58	0.65	0.67	0.67	0.64
N	1331	516	113	9	85

Source: Own calculations using MA baseline and endline data. Only observations for which panel data is available (households were surveyed both at baseline and endline) are included. Performance levels are from Ministry of Education, Science, Vocational Training and Early Education (2014).

	Zero	Below Minimum	Minimum	Desirable	Outstanding
Child characteristics					
Age (years)	8.7	9.0	9.1	10.2	9.3
Female	0.53	0.47	0.56	0.50	0.54
Household characteristics					
Mother's literacy	0.54	0.61	0.69	0.40	0.46
Father's literacy	0.77	0.84	0.80	1.00	0.81
Household Size	5.34	5.21	5.52	6.33	5.30
Household connected to electric grid	0.04	0.06	0.08	-	0.05
Asset ownership					
Television	0.24	0.35	0.32	0.50	0.23
Bed	0.50	0.61	0.58	0.67	0.47
Bike	0.76	0.73	0.79	0.33	0.69
Radio	0.58	0.65	0.67	0.67	0.64
N	1,331	516	118	6	83

Source: Own calculations using MA baseline and endline data. Only observations for which panel data is available (households were surveyed both at baseline and endline) are included. Performance levels are from Ministry of Education, Science, Vocational Training and Early Education (2014).

Table F3. Descriptive statistics on outcomes of interest by wave of data collection – Reading attitudes

	Baseline	Endline
Child likes to listen to stories (%)	55.7	92.9
Child likes to read or practice reading (%)	62.1	90.0
Child reads alone at home (according to child) (%)	70.0	78.8
Child likes to read or try to read on his/her own at home (%)	66.4	75.8
Someone helps or tries to help the child read at home (%)	78.1	80.8
Child likes to read or try to read with someone at home (%)	69.3	75.1
There are reading activities for child to participate in outside school (%)	12.0	11.8
Child participates in reading activities outside school (%)	7.4	7.9
N	2044	2046

Source: Own calculations using MA baseline and endline data. Sample sizes are smaller for come outcomes due to item-specific missing data.

Table F4. Descriptive statistics on outcomes of interest by wave of data collection – Reading patterns

	Baseline	Endline
According to the caregiver (%)		
How often parents read with children		
Four Days A Week Or More	13.5	20.3
Two Or Three Days A Week	37.0	44.8
One Day A Week	14.9	11.3
Once Or Twice A Month	5.0	1.1
Less Than Once A Month	1.2	0.8
Never	28.4	21.8
How often children read alone at home		
Four Days A Week Or More	14.4	25.0
Two Or Three Days A Week	25.2	33.3
One Day A Week	12.4	15.4
Once Or Twice A Month	3.1	1.8
Less Than Once A Month	1.6	0.9
Never	43.4	23.5
According to the child		
Days per week child reads with a parent	2.3	1.7
Days per week child reads alone at home	2.4	2.0
N	2025	2067

Source: Own calculations using MA baseline and endline data. Sample sizes are smaller for come outcomes due to item-specific missing data.

ANNEX G. VALUE-ADDED MODELS

In this section we discuss results of the Value-Added models. The key feature of these models is that we control for baseline test scores, as well as a series of other control variables. Specifically, at the student level we control for age, gender and grade level, and at the household level we control for household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity. The main improvement of this approach, compared to the non-parametric difference-in-difference just discussed, is that it controls for a range of sociodemographic characteristics, improving the precision of the estimates.

The results obtained through the Value-Added models do not differ much compared to the non-parametric difference-in-difference estimates, which is not surprising given that treatment status was randomized. Table G1 displays results pooling grades 2 and 3. We found no effects for the number of correctly pronounced letters or listening comprehension, but found positive and significant impacts for the other three evaluated subtasks. Specifically, as a result of MA, students in the treatment group are able to read 1.6 more non-words and 3.1 more real words per minute due to the program, and are able to answer 0.28 more reading comprehension questions.

Table G1 also shows the coefficients for children's age, gender, grade and parental literacy. It can be seen that being a female is positively correlated with EGRA scores, as being in grade 3 (as opposed to being in grade 2). Parental literacy is also positively correlated with children's EGRA scores.

Table G1. Impact of MA on reading skills

	Letter Sounds (Total Correct)	Non-Word Decoding (Total correct)	Oral Reading Fluency (per minute)	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
Impact	0.42	1.59***	3.11***	0.28***	0.015
	(0.69)	(0.43)	(0.65)	(0.065)	(0.065)
Gender	0.61	0.61*	1.35**	0.095*	-0.0059
	(0.39)	(0.30)	(0.45)	(0.047)	(0.042)
Age	-0.072	0.060	0.19	0.023	0.0065
	(0.13)	(0.10)	(0.14)	(0.017)	(0.017)
Grade	-0.035	0.63	0.83	0.16**	0.22***
	(0.52)	(0.35)	(0.57)	(0.055)	(0.053)
Father Literacy	0.39	0.67*	0.64	0.14**	0.11
	(0.47)	(0.31)	(0.46)	(0.052)	(0.055)
Mother Literacy	1.10*	0.41	0.69	-0.0089	0.088
	(0.47)	(0.31)	(0.50)	(0.059)	(0.053)
Effect Size	0.041	0.20	0.26	0.22	0.014
Observations	1966	1969	1968	1973	1973

Note: Value-added models on EGRA scores. All regressions include child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a tv, bed, bicycle, radio, and has electricity). The regressions also include dummy variables for the child's grade-level. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.
Standard errors clustered at the school level in parentheses
*
 p<0.05 ** p<0.01 *** p<0.001

Differential impacts

In this section we explore whether the program had different impacts across four dimensions of interest. First, we estimate separate models for students that were able to read at least one word at baseline and students that were not. Given that the program was providing reading materials, it is reasonable to expect that it had a greater impact on students that were already able to read, than on students that were not able to read a single word at baseline. We also analyze if the program had different impacts depending on whether or not there were any reading materials at home at baseline. Considering that there were no reading materials for children in almost half the surveyed households, it can be expected that the program had greater impacts in households where no reading materials are available, than in households were children have some reading resources. As discussed in the main body of the report, we also estimate separate models by grade, and for boys and girls.

Table G2 shows the impact of the program for students that were not able to read a single work at baseline in Panel A, and for students that were able to read at least one word at baseline in Panel B. No clear pattern can be distinguished. The impact coefficients are higher in panel A than in Panel B for letter sounds, non-word reading and listening comprehension, but lower for ORF and reading comprehension. Moreover, the differences are not statistically significant for any subtask. These results indicate that there are no clear differences between the impacts for students that were not able to read a single word at baseline and students that were.

Table G2. Impact of MA on reading skills by baseline reading level

	Letter Sounds (Total Correct)	Non-Word Decoding (Total correct)	Oral Reading Fluency (per minute)	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
A. Zero WPM at BL					
Impact	0.81	1.52***	2.70***	0.24***	0.027
	(0.63)	(0.36)	(0.57)	(0.048)	(0.082)
Effect Size	0.093	0.30	0.35	0.31	0.026
Observations	1327	1329	1329	1331	1331
B. One or more WPM at BL					
Impact	-0.82	1.28	3.31**	0.28*	-0.044
	(1.17)	(0.71)	(1.12)	(0.12)	(0.085)
Effect Size	-0.075	0.16	0.26	0.19	-0.044
Observations	638	639	639	640	640

Note: Value-added models on EGRA scores. All regressions include child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity). The regressions also include dummy variables for the child's grade-level. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.
Standard errors clustered at the school level in parentheses
* p<0.05 ** p<0.01 *** p<0.001

Table G3 shows impact results for students with no reading resources at home in Panel A and for the rest of students in Panel B. We define having reading resources at home as having a reader's book or other school reading materials, at least one children's books in ChiNyanja at home, and/or having access to reading activities in the community. The results show that there are no clear differences in treatment effects for these two groups of students. The parameters in Panel A are higher for non-word reading, reading comprehension and listening comprehension, but lower for letter sounds and ORF, and none of the differences are statistically significant except for letter sounds.

Table G3. Impact of MA on reading skills by availability of reading materials at baseline

	Letter Sounds (Total Correct)	Non-Word Decoding (Total correct)	Oral Reading Fluency (per minute)	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
A. No reading materials at BL					
Impact	-0.85	1.69**	2.96***	0.34***	0.035
	(0.90)	(0.58)	(0.83)	(0.081)	(0.086)
Effect Size	-0.082	0.24	0.25	0.28	0.033
Observations	824	825	825	828	828
B. Pooled (all grades together - some reading materials at BL)					
Impact	1.25	1.53**	3.29***	0.26***	0.0095
	(0.68)	(0.47)	(0.72)	(0.075)	(0.073)
Effect Size	0.12	0.19	0.25	0.19	0.0088
Observations	1120	1122	1121	1123	1123

Note: Value-added models on EGRA scores. All regressions include child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity). The regressions also include dummy variables for the child's grade-level. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.

Standard errors clustered at the school level in parentheses

Table G4 shows results disaggregating by grade. Students in grade 3 made more progress than students in grade 2. All the differences between the grade 2 and grade 3 impact parameters are statistically significant except for listening comprehension. This suggests that the program has a greater impact on students with higher reading levels at baseline, and/or that the materials students see in class in 3rd grade supplement MA activities better than those covered in 2nd grade. Further research should address the source of this difference in treatment effects across grades.

Table G4. Impact of MA on reading skills by grade

	Letter Sounds (Total Correct)	Non-Word Decoding (Total Correct)	Oral Reading Fluency (per minute)	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
A. Grade 2					
Impact	-0.51	1.09*	2.07*	0.16*	-0.051
	(0.77)	(0.48)	(0.82)	(0.074)	(0.084)
Effect Size	-0.051	0.16	0.19	0.14	-0.047
Observations	1023	1025	1025	1027	1027
C. Grade 3					
Impact	1.41	2.11***	4.18***	0.41***	0.085
	(0.88)	(0.59)	(0.84)	(0.090)	(0.071)
Effect Size	0.14	0.25	0.31	0.29	0.083
Observations	943	944	943	946	946

Note: Value-added models on EGRA scores. All regressions include child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity). The regressions also include dummy variables for the child's grade-level. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.

Standard errors clustered at the school level in parentheses

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Finally, Table G5 shows results for boys and girls. For non-word reading the estimates are basically equivalent by gender, while for ORF and reading comprehension girls progressed more than boys due to the program. However, none of the differences between girls and boys are statistically significant.

Table G5. Impact of MA on reading skills for boys and girls

	Letter Sounds (Total Correct)	Non-Word Decoding (Total Correct)	Oral Reading Fluency (per minute)	Reading Comprehension (Total Correct)	Listening Comprehension (Total Correct)
A. Boys					
Impact	0.42	1.56**	2.77***	0.23**	-0.017
	(0.77)	(0.49)	(0.72)	(0.075)	(0.078)
Effect Size	0.041	0.21	0.23	0.19	-0.016
Observations	951	954	953	955	955
B. Girls					
Impact	0.21	1.55**	3.28***	0.32***	0.035
	(0.79)	(0.54)	(0.83)	(0.087)	(0.076)
Effect Size	0.020	0.19	0.25	0.23	0.032
Observations	1015	1015	1015	1018	1018

Note: Value-added models on EGRA scores. All regressions include child's characteristics (age and gender) and a household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity). The regressions also include dummy variables for the child's grade-level. Three outliers for letter sounds were dropped, as well as 1 for Non-words and 2 for ORF.

Standard errors clustered at the school level in parentheses

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

ANNEX H. ATTITUDES TOWARDS READING AND READING HABITS

ATTITUDES TOWARDS READING

To analyze children's attitudes towards reading we asked them questions on whether they like reading or practicing reading in different contexts, and provided them with three response options: like, indifferent, and do not like. Children in both treatment and control groups were surveyed at baseline and endline. To simplify the analyses, we recoded these responses and created one dummy variable for each question, so if the child likes reading in a given context the dummy variable is 1, and if the child does not like it or is indifferent it is 0. We ran Linear Probability Models on the resulting bivariate variables, hence the estimated parameters can be directly interpreted as changes in percentage points in the likelihood that children like reading in a given context attributable to MA.

Table H1 shows the regression results for this exercise. We did not find any evidence that the program changed children's attitudes towards reading. In column (1) the treatment effect for whether the child likes to read is displayed. The coefficient is positive but not significant at 5 percent of confidence. A similar result can be observed for whether the child likes to read or practice reading (in general); the coefficient is positive but small and not significant. Note that the fraction of children that reported they like to read at baseline was 62 percent, suggesting that there was room for improvement.

We also failed to find a significant impact on whether the child likes to read at home on his or her own, or if they like reading with a household member at home. These results are displayed in columns (3) and (4), respectively. Note that the sample size of these last two regressions is smaller than what we have for the first two regressions; this is because questions on whether children like reading at home (alone or with a family member) were only asked to children that answered affirmatively that they engage in reading at home. Along these lines, for the regressions on whether children like reading on their own at home, we restricted the sample to those who answered affirmatively that they read at home on their own at baseline, and an analogous restriction was conducted for whether children like reading with a family member at home.[23]

We also ran regressions recoding the dependent variable so if the child likes reading or is indifferent the binary variable is 1, and if the child does not like it is 0, but the results are also negligible.

---

[23] Cases where there are no responses at endline because the child answered s/he did not engage in the corresponding activity at endline are coded as 0. This implicitly assumes that children that engaged in the given activity at baseline but not at endline does not like the activity or is indifferent.

Table H1. Impact of MA on Attitudes towards Reading

	Does the child like to:			
	Listen to stories (1)	Read or practice reading (2)	Read at home on his/her own (3)	Read with someone at home (4)
Impact	0.027	0.0097	0.028	0.019
	(0.016)	(0.020)	(0.032)	(0.028)
Observations	1840	1860	1329	1452
Mean of outcome (baseline)	0.56	0.62	0.66	0.69

Note: All estimates correspond to Linear Probability Models. Each model includes the treatment dummy, the lagged value of the dependent variable child's age, gender and grade, and household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity.

Standard errors clustered at the school level in parentheses

* p<0.05 ** p<0.01 *** p<0.001

To evaluate the effect of the program on the likelihood that children participate in out-of-school reading activities, we questioned caregivers about the availability of such activities in their communities, as well as their children's participation in them. These activities include reading clubs, reading activities at the community center and visiting the school library, among other options.

The program did not have an effect on children's participation in out-of-school reading activities either. In this case it would seem that the driving factor is the actual lack of out-of-school activities available in the communities. In effect, both at baseline and endline only 12 percent of caregivers reported having access to out-of-school reading activities for the children; hence a program like MA probably had little chance to affect the likelihood that children increased their participation in this type of activities, considering that program did not generate these activities.

Table H2 shows results for whether there are activities in the communities and students' participation in these activities. We show results on whether there are reading activities available not because the program was supposed to affect the actual availability of this type of programs, but just to check if maybe the program increased the likelihood that parents were aware that such activities were available in their communities, because ultimately parents need to know first these activities exist before having their children participate in them. The program had no impact on the likelihood that there are out-of-school reading activities in the community. We also found that the program did not have any impact on whether students participate or not in such programs. If we restrict the sample to those (few) cases where caregivers reported that there are out-of-school reading activities in their communities (both at baseline and endline), the impact of MA on the likelihood that children participate in these activities is also negligible.

Table H2. Impact of MA on Availability of and Participation in Out-of-School Reading Activities

	Reading activities are available	Child participates
Impact	0.035	0.0040
	(0.020)	(0.012)
Observations	2036	2034
Mean of outcome (baseline)	0.12	0.074

Note: All estimates correspond to Linear Probability Models. Each model includes the treatment dummy, the lagged value of the dependent variable child's age, gender and grade, and household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity. Standard errors clustered at the school level in parentheses
* p<0.05 ** p<0.01 *** p<0.001

READING HABITS - READING WITH PARENTS

To analyze if as a result of MA parents spend more time supporting their children's reading activities, we collected data on how often caregivers read with their children according to the respondent of the caregiver survey (who was a parent 80 percent of the cases at endline), and according to the child. As was the case with children, caregivers in both treatment and control groups were surveyed at baseline and endline.

In the caregiver survey we asked whether parents read with the MA child four days a week or more, two or three days a week, one day a week, once or twice a month, less than once a month or never. We also asked the caregiver how long each reading session lasted in minutes.

Table H3 shows results using this data as well as the data collected through the student survey. In column (1) the dependent variable corresponds to the parent that reads with the child most frequently. To accommodate the categorical nature of the dependent variable we used an interval regression type of model, with the intervals in the regression coded such that the resulting coefficients can be interpreted in number of days per week. We can see that the effect is positive and significant, the parameter indicates that the parent that spends the most time reading with the child in the treatment group reads with the child 0.5 more days per week than their counterparts in the control group as a result of MA, that is 1 day every two weeks.

In column (2) we evaluate the impact of the program on the number of minutes each reading session lasts. The program did not have a significant impact on session duration. Note that the regression is restricted to cases where caregivers reported a positive number of minutes. This allows analyzing separately the effect of the program on reading session frequency, already documented by the results in column (1), and the effect on session duration.

We used the results from the interval regression in column (1) and the data on number of minutes each session lasts to estimate how many minutes per week parents read with their

children. We found that parents in the treatment group read with their children 16 more minutes per week due to the program. The difference is significant at 1 percent. [24]

In column (3) the analysis using the child's responses is shown. In this case rather than asking the child the options provided to the caregiver in terms of reading frequency, we opted for a simplified version of the question and only asked which days in the previous week someone read with him or her at home. The displayed regression shows the results on the sum of the reported days per week. In this case the coefficient is much smaller and not significant. To keep the student survey simple, we did not ask how long sessions lasted.

Table H3. The effect of MA on children's reading habits at home – Reading with a parent

	According to Caregiver		According to Child
	Number of days per week	Minutes per session (for min>0)	Number of days per week
	(1)	(2)	(3)
Impact	0.52***	0.65	0.047
	(0.099)	(1.14)	(0.093)
Observations	2066	1162	1968
Mean of outcome (baseline)	2.00	28.4	2.30

Note: The model in column (1) corresponds to an interval regression, the models in columns (2) and (3) correspond to OLS. Each model includes the treatment dummy, the lagged value of the dependent variable (in the case of the model in column (1) this corresponds to a set of dummy variables that categorize the dependent variable at baseline) a subset of child's characteristics (age and gender) and a subset of household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity), as well as the child's grade-level.

Standard errors clustered at the school level in parentheses

* p<0.05 ** p<0.01 *** p<0.001

It is rather puzzling that the effect of MA on joint reading as reported by the caregiver indicates a clear positive impact, while the effect on the same underlying construct as reported from the child's perspective indicates a null result. It is possible that some of these children may be too young to accurately report how often they read with someone at home. We ran regressions on the number of days children read with someone at home, as reported by the child, restricting the sample to older children (9 years old, which is the median age at baseline), but again, did not find any impact of the program, which suggests that the lack of results is not driven by some of these children being too young to assess correctly the number of days they read with someone at home. Of course they could be all too young to report with any accuracy.

---

[24] To calculate the number of minutes per week parents read with children we multiplied the number of days predicted by the interval regression in column (1) of Table H3, times the number of minutes reported. Then we calculated the difference between treatment and control for this product and bootstrapped the standard errors.

Another possible explanation is that caregivers in the treatment group tend to overstate how frequently they read with the child, given that they participate in the MA program and they may feel there is an expectation that they behave in a particular way.

Ideally we would like to determine which respondent provided more reliable information: the child or the caregiver.

To try to assess which respondent is providing more reliable data, we analyze the correlation between how often children read with a parent at baseline and endline, according to the children's responses and according to the caregivers' responses. In principle, at least for the control group, the frequency at which children read with their parents should not change too much over time, and therefore analyzing the correlation between baseline and endline responses provided by each type of respondent can shed some light on how reliable each type of survey respondent is.

Figure H1 shows two graphs that summarize children's and caregivers' responses. Only control observations are included. In Panel A we present children's responses. The figure shows the average number of days that children read with a family member at endline by the number of days they read with a family member at baseline. There is no apparent correlation between children's responses at baseline and endline. For example, children that reported reading with a family member one day per week at baseline answered they read with a family member 2.2 days per week at endline on average, while children that said they read with a family member 7 days per week at baseline, answered they read with a family member 1.9 days per week at endline on average. Not surprisingly, the correlation coefficient is pretty low, 0.07. In other words, the number of days children said they read with a family member at baseline does not seem to be a good predictor of the number of days they said they were read at endline.

In Panel B the results for the caregiver's responses are shown. Given the interval nature of the question we used to record caregivers' responses on how often they read with children, we transformed the endline variable so we could calculate averages in terms of number of days per week.[25] In this case the correlation coefficient is still low (0.26), but at least there is certain correspondence between the caregiver's report at baseline and endline. In effect, for the most part, the higher the frequency at which caregivers reported they read with the child at baseline, the higher the average number of days they reported they read with child at endline.

Judging by these results, it would seem that data provided by the caregiver on how often children read with their parents is more reliable than children's data.

---

[25] Specifically, we coded each interval using its midpoint, so 'Four days a week or more' is 5.5, 'Two or three days a week' is 2.5, 'One day a week' is 1, 'Once or twice a month' is 0.4, 'Less than once a month' is 0.1 and 'Never' is 0. An alternative transformation can be obtained by running an interval regression and use the predicted number of days of the model; the problem with this approach is that, as the resulting number of days the model produces is naturally compressed by the regression, we are left with very little variation in the (expected) number of days, which is crucial for the purpose of this graphical exercise.

Figure H1. Baseline and endline responses on how often parents read with children in the control group
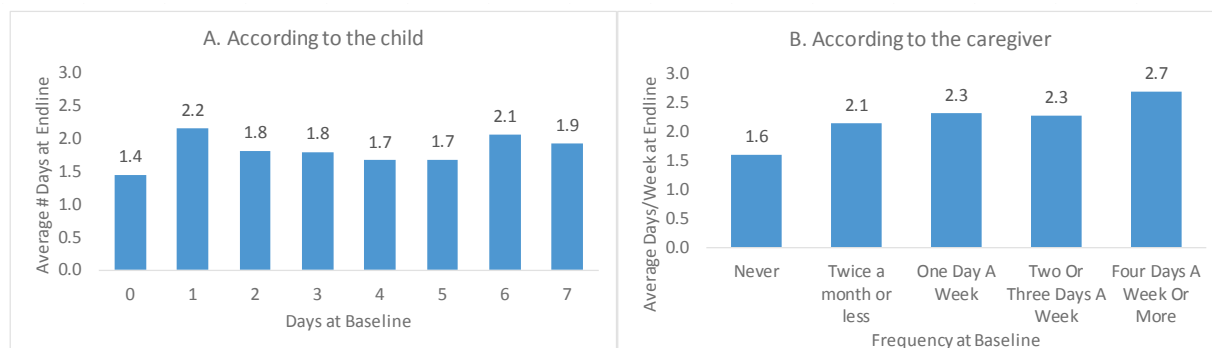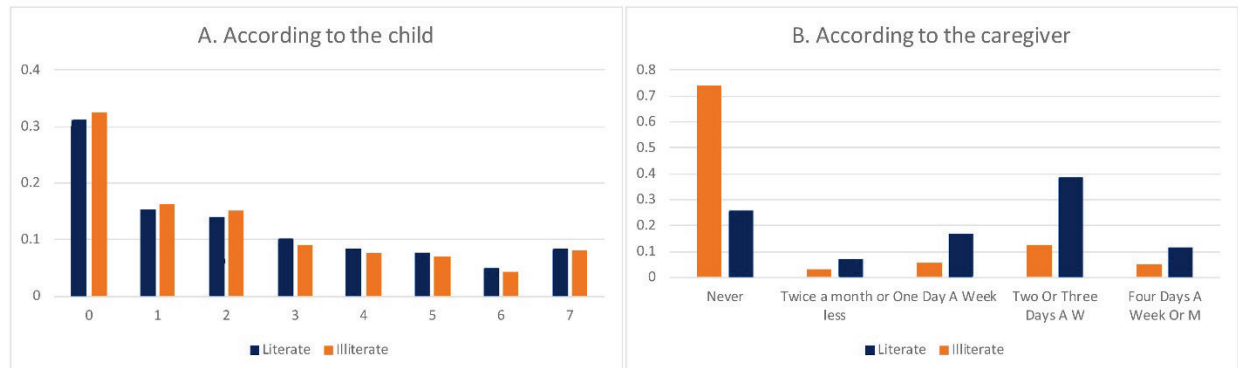

Note: In Panel B the baseline category 'Less than once a month', with only 14 observations, is pooled with the 'Once or twice a month' category.
Source: Own calculation using MA baseline and endline data.

We also analyzed how children's and caregivers' responses are associated with whether the caregiver knows how to read. Although it is possible that illiterate caregivers would read with the child, it is natural to expect that if a caregiver knows how to read it is more likely that s/he will read with the child than if s/he is illiterate. Therefore, if survey responses are reflecting the actual frequency at which parents read with children, we should expect that literate parents read more often with their children than illiterate parents.

To conduct this analysis Panel A in Figure H2 shows the distribution of the number of days children said a household member read with them in the previous week at baseline, by whether the caregiver knows how to read. A similar exercise, using caregiver's data, is displayed in Panel B. The source for determining literacy of the caregiver is always the caregiver survey. According to the child the results show that a little over 30 percent of children did not read with another household member, regardless of whether the caregiver is illiterate or not; in fact, the distributions are not very different depending on whether the caregiver knows how to read or not. The results using caregiver's data are a little different. In this case there is more correspondence between literacy of the caregiver and the frequency at which s/he reads with the child. In effect, while more than 70 percent of caregivers that do not know how to read said they never read with the child at baseline, only 26 percent of caregivers that know how to read said they never read with the child. As a consequence, small fractions of illiterate caregivers read with the child once a week, two or three times a week or four or more days a week, while the fractions of literate caregivers reading at these frequencies are much higher.

Figure H2. Frequency at which children read at baseline by literacy of the caregiver (including treatment and control groups)

Note : In Panel B the baseline category 'Less than once a month', with only 14 observations, is pooled with the 'Once or twice a month' category.
Source: Own calculation using MA baseline and endline data.

While these analyses are far from constituting proof that caregivers' responses reflect the actual frequency at which parents read with children and that children's responses do not, given the available data we would be inclined to consider the results using caregiver's data more reliable than those derived using children's data.

READING HABITS - READING ALONE

To analyze the impact of the program on the frequency at which MA children read at home on their own, we collected data from both the child and the caregiver. Children and caregivers in both the treatment and control groups were surveyed at baseline and endline. In the caregiver survey, we asked if the child read four days a week or more, two or three days a week, one day a week, once or twice a month, less than once a month or never. In other words, we used the same response structure that we used to record the frequency at which household members read with the child. We also asked the caregiver how long lasted each reading session in minutes. In the student survey, we asked how many days they read at home on their own last week. Therefore, in this case we also used the same question structure that we used in the student survey to record reading with a household member.

Table H4 shows regression results using this data. In column (1) the dependent variable is the number of days the child reads alone at home, as reported by the caregiver. As before we used an interval regression to accommodate the categorical nature of the dependent variable, and the intervals are coded so the coefficients can be interpreted in number of days per week. The parameter indicates that children in the treatment group read on their own at home 0.36 days per week more than children in the control group. No significant effect is observed for the number of minutes each session lasts. To calculate the number of minutes per week children read alone at home we followed the same procedure described before to calculate the number of minutes children read with their parent. We found that children read 12 more minutes per week due to the program. This difference is significant at 1 percent.

When we look at the effect of the program on the number of days children read alone according to themselves, the result indicates that children in the treatment group read 0.23 more days due to the program.

Table H4. The effect of MA on children's reading habits at home – Reading alone

	According to Caregiver		According to Child
	Number of days per week	Minutes per session for min>0	Number of days per week
	(1)	(2)	(3)
Impact	0.36**	0.83	0.23*
	(0.12)	(1.34)	(0.10)
Observations	1940	834	1970
Mean of outcome (baseline)	1.63	19.5	2.39

Note: The model in column (1) corresponds to an interval regression, the models in columns (2) and (3) correspond to OLS. Each model includes the treatment dummy, the lagged value of the dependent variable (in the case of the model in column (2) this corresponds to a set of dummy variables that categorize the dependent variable at baseline) a subset of child's characteristics (age and gender) and a subset of household characteristics (household size, mother's literacy, father's literacy, and whether their household owns a TV, bed, bicycle, radio, and has electricity), as well as the child's grade-level.

Standard errors clustered at the school level in parentheses

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

While the parameters on the number of days children read alone derived from the child and caregiver data are different, they are both positive and significant, and the actual difference between the two coefficients is only about one standard error. These results suggest that the program had a positive impact in the number of days students read at home on their own.

Note that for this research question we did not find contradicting results between the responses provided by the caregiver and the responses provided by the child. It is possible that children were able to provide more reliable data when they were asked about reading alone than when they were asked about reading with someone at home.

In effect, as can be seen in Figure H3, children's responses on how often they read alone between baseline and endline (Panel A) seem to be more correlated than when the outcome is reading with someone at home (Panel B). In effect, the correlation coefficient between baseline and endline for reading alone is 0.21 and for reading with a family member is 0.07. This suggests that data provided by children on reading alone may be more reliable than data on reading with a parent, which could explain why there are no major discrepancies between the results using caregivers and children's data on the effect of the program on reading alone.

Figure H3. Baseline and endline responses on how often children read on their own at home according to the child for the control group
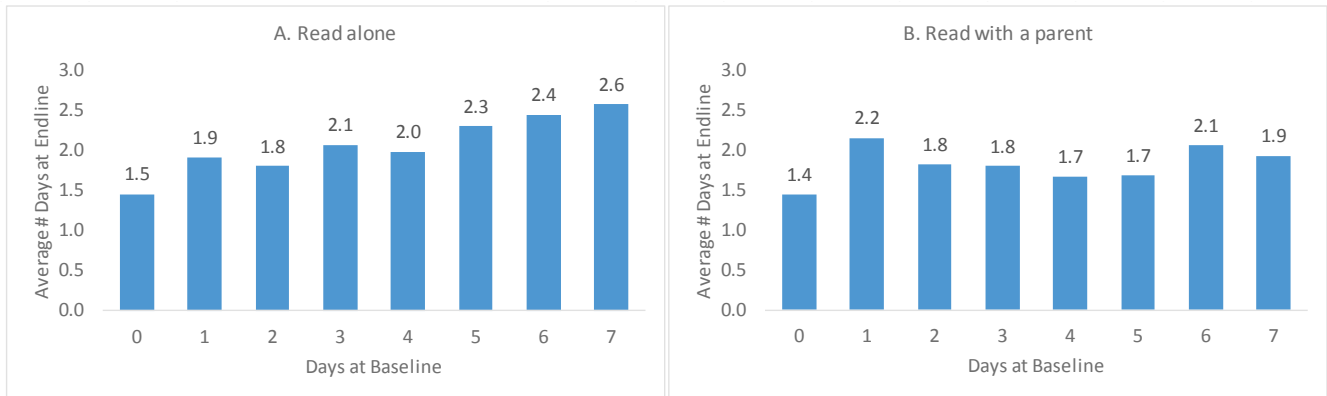
Source: Own calculation using MA baseline and endline data.

We can only speculate as to why children's responses would be more reliable for reading alone than for reading with someone else. Albeit the questions follow the same structure, perhaps it is easier for a child to remember how often s/he reads on her/his own, than remembering how often anyone at the household reads with them, which involves remembering interactions with probably several people.

ANNEX I. SCALE-UP COST ESTIMATE

In this section we discuss the costs of the project. Specifically, we estimate the costs that would be observed if the project was scale up at the national level. For this analysis we consider the population of interest children in grades 2 and 3 in Zambia, as only children in these grades were treated during the MA pilot. We divide this population between children living in Lusaka and the Eastern province, where ChiNyanja is the language predominantly spoken, and the rest of the country where six other local languages are spoken. The reason for this division is that the costs for these two groups will be rather different, at least initially. In effect, while for ChiNyanja a bank of stories has already been constructed as a result of the MA pilot, for the rest of the country new stories will need to be developed, which will entail costs similar to the ones observed for the MA pilot.

Moreover, we restrict the population of interest to the fraction that we estimate have access to a cell phone at home. According to the results discussed in section III.2, 56.5 percent can be considered a lower bound of the true cell phone ownership rate.

We estimate that by 2018 the number of 2nd and 3rd graders in the country will be 1,123,691, and of those 629,267 will have access to a cell phone at home. The Eastern Province and Lusaka account for approximately 29 percent of the country's population. Along these lines, we estimate that the program could reach 184,065 2nd and 3rd graders in these two provinces in 2018, and 445,202 in the rest of the country.26

Table 11 summarizes the costs of the scale-up. A few assumptions are worth highlighting:

1. Cost of SMS and IVR are estimates provided by Creative, these do not correspond to the unitary prices observed during the program.
2. To manage the program, we are assuming that one Program Coordinator (PM) for each of the 7 official languages will need to be hired by the Ministry of Education.
3. To replicate what was fielded during the pilot, Community Mobilizers (CM) will need to be hired to conduct monthly meetings with parents. According to Creative, during the pilot there were 4 community mobilizers that conducted one meeting per month in each of the participating school communities (one mobilizer per 10 school communities). The average total monthly cost for all 4 community mobilizers was $3,616 (or $904/mobilizer/month).
4. CM will contact parents to capture cell phone numbers and conduct monthly meetings with them. CM will also contact eligible parents that do not attend the monthly meetings to see if they can be persuaded to do so.[27]
5. For each of the 6 official languages for which no stories have been developed, the costs to develop stories is imputed as the total development cost observed during the pilot. This information was provided by Creative.

---

[26] Own calculations using Census data for the total and by-region population, UNESCO data for the number of 2nd and 3rd graders over the last few years, and MA baseline data for the cell phone ownership rate.
[27] In the pilot contacting parents that were not attending meetings was the role of community volunteers.

6. Office space and equipment assumed as available at no cost

The estimated scale-up costs per student are US 20.1 for students in Lusaka or Eastern Province, and US 21.6 for students in other parts of the country.

There are several considerations worth highlighting regarding these estimated costs. First, MA was piloted in predominantly rural areas. Several things could change when the intervention is launched in urban areas. First, cell phone ownership could be higher. This would imply that program's fixed costs could be lower per student, improving the cost-effectiveness of the program. Second, it is possible that the lack of reading materials is less of a problem in urban areas than in rural areas. This could increase the frequency of refusals because reading resources are not in such need; on the other hand, if children have better reading skills in urban areas than in rural areas, the impact of the activity could be higher as the program had greater impacts on abler students, which would increase the effectiveness of the program per dollar invested.

We are also assuming that cell phone coverage is acceptably reliable across the country, which may not be the case in some areas.

Finally, if stories are translated rather than developed from scratch, then the development costs could be reduced. Moreover, even if stories are developed from scratch for each official language, after the first year the costs would fall anyway because no new stories would need to be developed (at least not as many as at the beginning).

Table I1. Scale-up cost estimation assumptions*

		Number of students		
		Lusaka & Eastern Province 184,065	Rest (6 other languages) 445,202	
1. Program Management	*Cost*	*Total $*	*Total $*	*Assumption*
• 7 MOE Program Coordinators (one per official language)	15,000/each	15,000	90,000	The MOE will hire 7 full time Program Coordinators, one per official language.
• Travel for Coordinators	1 trip/year/district @ $650/trip = $6500	1300	5200	MOE Coordinator will visit each province once per year for training
• Training for community mobilizers	Each PC trains 51 CM during 10 days @ $100/day	1000	6000	PC will train an average of 51 CM. Each PC will have 10 days for training, so probably two groups of 5-day trainings.
• Guidebook production and distribution for above	357 CM + 5017 schools x $1= $5374	1,572	3,802	10-page guidebook at $0.10/page
Sub-total		*18,872*	*105,002*	
2. Publicity/Marketing				
• Flyers/posters printing and distribution	5017 schools x $3 = $15,051	4,403	10,648	Posted by CM to encourage service sign-up
• Radio time (marketing)	1 station/province x 10 provinces @$500 x 9 time/year =$45,000	9,000	36,000	Monthly publicity/information about the program. Additional time (3 more monthly broadcasts) donated by radio stations
• Public Events	10 provinces x $1,000 = $10,000	2,000	8,000	MOE in provinces will organize one public event per year in different location each year. Additional funding will be obtained from local businesses.

		Number of students		
		Lusaka & Eastern Province 184,065	Rest (6 other languages) 445,202	
• Community mobilizer	One mobilizer conducts monthly meetings in 10 school communities. The monthly rate for each CM is $904.	1,591,952	3,850,489	This replicates what happened in the pilot. CM will also work with teachers to contact parents and capture cell phone numbers.
Sub-total		*1,607,355*	*3,905,138*	
3. Story Development				
• Selection, levelling, adaptation workshops	Development costs for the pilot were US 97,577	-	*585,459*	SMS story library to be developed in 6 local languages of instruction for early grades and in English for grades 2-3
Sub-total		-	*585,459*	
4. Delivery/Transmission of SMS				
• Wi-fi/Internet for IVR and SMS	$1,200	351	849	
• Frontline Cloud subscription	$300	88	212	$25 per month
• Bulk SMS fees	156 SMS/student, at $0.02/message	574,283	1,389,030	
• Website domain	50	15	35	A fee once per year to maintain our website domain.
• Website admin	US 15000	3,000	12,000	
Sub-total		*577,736*	*1,402,127*	
4'. Other				
• Notebooks	30 cent/notebook	55,220	133,561	
Sub-total		*55,220*	*133,561*	
Total (SMS)		**2,259,182**	**6,131,287**	
Cost per student (SMS)		*12.3*	*13.8*	
5. Delivery/Transmission of Voice Recording				

| | | Number of students | | |
		Lusaka & Eastern Province 184,065	Rest (6 other languages) 445,202	
• IVR	155 minutes/student @ $0.05/minute	1,426,504	3,450,316	
• Audio recording	6,000	6,000	36,000	Professional Audio Recording of stories and messages ($500/month).
Sub-total		1,432,504	3,486,316	
Total (SMS+Voice Recording)	*15,524,226*	*3,691,686*	*9,617,602*	
Cost per student (SMS+IVR)		*20.1*	*21.6*	

*Adapted from framework developed by Creative.

ANNEX J. DISCLOSURE OF ANY CONFLICTS OF INTEREST

Name	Alicia Menendez
Title	Senior Research Scientist
Organization	NORC at the University of Chicago
Evaluation Position?	X Team Leader Team member
Evaluation Award Number *(contract or other instrument)*	AID-OAA-M-13-00010 – Impact Evaluation of Zambia Makhalidwe Athu
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID/Zambia Makhalidwe Athu Impact Evaluation
I have real or potential conflicts of interest to disclose.	☐ Yes X No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	10/10/2017
Name	Alejandro Ome
Title	Senior Research Scientist
Organization	NORC at the University of Chicago

Evaluation Position?	☐ Team Leader X Team member
Evaluation Award Number *(contract or other instrument)*	AID-OAA-M-13-00010 – Impact Evaluation of Zambia Makhalidwe Athu
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID/Zambia Makhalidwe Athu Impact Evaluation
I have real or potential conflicts of interest to disclose.	☐ Yes X No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	10/10/2017

Name	Varuni Dayaratna
Title	VP International Projects
Organization	NORC at the University of Chicago

Evaluation Position?	☐ Team Leader X Team member
Evaluation Award Number *(contract or other instrument)*	AID-OAA-M-13-00010 – Impact Evaluation of Zambia Makhalidwe Athu
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID/Zambia Makhalidwe Athu Impact Evaluation
I have real or potential conflicts of interest to disclose.	☐ Yes X No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	10/10/2017

Name	Russell Owen
Title	Senior Research Analyst
Organization	NORC at the University of Chicago

Evaluation Position?	Team Leader X Team member
Evaluation Award Number *(contract or other instrument)*	AID-OAA-M-13-00010 – Impact Evaluation of Zambia Makhalidwe Athu
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID/Zambia Makhalidwe Athu Impact Evaluation
I have real or potential conflicts of interest to disclose.	☐ Yes X No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated. 2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation. 3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project. 4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated. 5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated. 6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	10/10/2017

Name	Carlos Fierros
Title	Research Analyst
Organization	NORC at the University of Chicago
Evaluation Position?	Team Leader X Team member
Evaluation Award Number (contract or other instrument)	AID-OAA-M-13-00010 – Impact Evaluation of Zambia Makhalidwe Athu
USAID Project(s) Evaluated (Include project name(s), implementer name(s) and award number(s), if applicable)	USAID/Zambia Makhalidwe Athu Impact Evaluation
I have real or potential conflicts of interest to disclose.	☐ Yes X No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	*[signature]*
Date	10/10/2017