



USAID
FROM THE AMERICAN PEOPLE

MALAWI

ABE/LINK

**MALAWI TEACHER PROFESSIONAL
DEVELOPMENT SUPPORT**

DRAFT

**MALAWI READING INTERVENTION
EARLY GRADE READING ASSESSMENT
(EGRA)**

FINAL ASSESSMENT - 2012

Contract No.: EDH-I-00-05-00026-02

Task Order No: EDH-I-04-05-00026-00

This report was prepared for review by the United States Agency for International Development. It was submitted to the COR, Malawi Teacher Professional Development Support project by Creative Associates International, RTI International, and Seward Inc.

ABE/LINK
Malawi Teacher Professional Development Support
(MTPDS)

Malawi Reading Intervention
Early Grade Reading Assessment (EGRA)
Final Assessment — 2012

Prepared by:

Sarah Pouezevara, Michael Costello, Odala Banda

Submitted by:

Creative Associates International, RTI International and Seward Inc.

under

Contract No.: EDH-I-00-05-00026-02

Task Order No: EDH-I-o4-05-00026-00

Creative Associates International

5301 Wisconsin Avenue, NW, Suite 700

Washington, DC 20015

Date submitted: April 23, 2013

Disclaimer

This report is made possible with the support of the American People through the United States Agency for International Development (USAID). The contents of this report are the sole responsibility of the author and do not necessarily reflect the views of USAID or the United States Government.

Table of Contents

Foreword.....	iv
Acknowledgments.....	v
Abbreviations.....	vi
Executive Summary.....	1
Background.....	1
Did the MTPDS reading intervention— Maziko a Kuwerenga —improve reading outcomes?.....	1
What contributed to this achievement?	2
How were outcomes rigorously evaluated?	4
Introduction	1
Background.....	1
Purpose and Research Design.....	2
Methodology and Limitations.....	4
Methodology	4
Sampling	4
Limitations	5
Comparison of Gains from Baseline to Endline	7
Results Overview by Subtest, EGRA 2012.....	10
Summary of Average Scores by Subtest	10
2012 Endline Zero Scores.....	12
Oral Reading Fluency and Comprehension.....	14
Descriptive Statistics	17
Learner Background.....	17
Teacher Interview.....	18
Head Teacher Interview.....	21
Conclusions and Recommendations.....	23
Factors that affect reading outcomes.....	23
Summary and recommendations	25

List of Tables

Table 1. Endline Results on key EGRA subtests, by Intervention Group	2
Table 2. Summary of Three-Year Research Design Outputs.....	3
Table 3. 2010 and 2012 RCT Intervention Study Sample, by Division and Standard.....	5
Table 4. Endline Gains (intervention group only)	8
Table 5. Endline Results on EGRA Subtests with a Fluency Component, by Intervention Group	10

Table 6. Results on EGRA Subtests Measured by Percent Correct, by Intervention Group, at Endline	11
Table 7. Comprehension Scores (intervention group only)	15
Table 8. Comprehension Scores (control group only)	16
Table 9. Learner Characteristics (% of sample with this characteristic or average value)	17
Table 10. Teacher Characteristics (% of sample with this characteristic or average value)	19
Table 11. Materials Used in Class	20
Table 12. Head Teacher and School Characteristics (% of sample with this characteristic, or average value)	21
Table 13. How Do You Know Your Children Are Progressing?	23

List of Figures

Figure 1: Comparison of Coaching Visits in Each Intervention District	6
Figure 2: Comparison of Intervention and Control School Scores at Baseline and at Endline on Timed Measures	8
Figure 3: Change in Zero Scores from Baseline to Endline, by Group	9
Figure 4: Comparison of Intervention Sample and National Sample	12
Figure 5: Percentage of Zero Scores by Group	13
Figure 6: ORF Scores by Range	14
Figure 7: Fluency vs. Comprehension (intervention group only)	16
Figure 8: Relationship Between Coaching Visits and Oral Reading Fluency (Intervention sample only, endline)	24
Figure 9: Relationship Between Class Size and Oral Reading Fluency (Intervention sample only, endline)	25

Foreword

In recent years, Malawi has made great strides in realizing the Millennium Development Goal (MDG) of providing universal access to primary education to all children of school-going age by 2015. The 2012 school census report showed that net enrollment had reached 99%. However, recent studies such as those carried out by the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the 2010 Early Grade Reading Assessment (EGRA) baseline have revealed a worrying picture of poor performance in Malawian primary schools, especially in literacy. Now is the time to focus upon improving educational attainment in all schools nationwide. The Early Grade Reading Assessment provides us with a valuable tool for measuring progress towards this goal.

The Ministry of Education, Science and Technology (MoEST) believes that reading is the most important skill that any child can learn at school. Without the ability to read, it is impossible for a child to access the school curriculum and it is impossible for an adult to participate fully as a productive member of society. Literacy is instrumental for national development, and the earlier that learners master literacy skills the better.

EGRA measures a learner's progress in developing the essential component skills of reading from the earliest stages, when interventions will have the most effect. MoEST is committed to developing internal capacity to administer EGRA as an integral part of its efforts to monitor learner achievement and to track the impact of interventions. For this reason it is heartening to know that the results in this report are based upon fieldwork conducted and supervised by MoEST staff.

This intervention study is an important companion to the 2012 EGRA midterm assessment report, which documents the performance of 3,000 Standard 2 and 4 learners, from a random selection of 150 schools nationwide. The results of these two samples present a challenge to everyone with an interest in primary education in Malawi.

—MacPherson Magwira, Permanent Secretary for Education, Science and Technology

Acknowledgments

The USAID Malawi Teacher Professional Development Support (MTPDS) project thanks the Ministry of Education, Science and Technology for its participation in piloting, training, data collection, data analysis, and report preparation, during the 2012 Early Grade Reading Assessment exercise. We also thank the students, teachers, and head teachers who welcomed us into their schools and participated in the study.

This report follows two comprehensive reports prepared by MTPDS for the baseline and mid-term. Much of the original format, guidance, and explanation is due to the efforts of the original authors—Jessica Mejia (2010) and Emily Miksic (2011), supported by Steve Harvey, the project’s Chief of Party.

MTPDS deeply regrets the loss of the Senior M&E Officer, Demis Kunje, just as the final data collection was underway in November 2012. Mr. Kunje directed the M&E activities of the project while also providing technical assistance to the MoEST in developing an M&E system that monitors the implementation of the NPC along with supporting communities to monitor the performance of their schools. The interpretation of the results would no doubt have been deeper and richer with his input.

Abbreviations

ABE/LINK	Assistance to Basic Education/Linkages to Education and Health Initiative
CERT	Centre for Education, Research and Training
clpm	correct letters per minute
CLS	Centre for Language Studies
clspm	correct letter sounds per minute
cnwpm	correct non-words per minute
CPD	Continuous Professional Development
Creative	Creative Associates International
csspm	Correct syllable sounds per minute
cwpm	correct words per minute
DBE	Directorate of Basic Education
DIAS	Directorate of Inspectorate and Advisory Services
DTED	Department of Teacher Education and Development
EGRA	Early Grade Reading Assessment
EMIS	Education Management Information System
JCE	Junior Certificate of Education
M&E	monitoring and evaluation
MIE	Malawi Institute of Education
MoEST	Ministry of Education, Science and Technology
MSCE	Malawi School Certificate of Education
MTPDS	Malawi Teacher Professional Development Support Project
n/a	not available
NPC	National Primary Curriculum
ORF	oral reading fluency
PA	phonological awareness
PCAR	Primary Curriculum and Assessment Reform
PEA	Primary Education Advisor
PTA	Parent–Teacher Association
RTI	RTI International (trade name of Research Triangle Institute)
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SES	socioeconomic status
Std	Standard
USAID	United States Agency for International Development

Executive Summary

Background

In 2010, the United States Agency for International Development (USAID), in collaboration with the Malawi Ministry of Education, Science and Technology (MoEST), developed a project to improve primary education and implementation of the National Primary Curriculum. The Malawi Teacher Professional Development Support (MTPDS) project, as it is known, included five major results areas related to improving teaching policy, teacher performance, early grade literacy, primary teaching and learning materials, and monitoring and evaluation (M&E) systems. While many of the inputs designed to support these objectives took place on a national scale, a specific reading improvement program was designed and delivered in two districts in order to determine its effectiveness on a small scale.

This reading intervention, Maziko a Kuwerenga (Foundations of Reading), was rolled out to schools in Salima and Ntchisi September 2011 reaching 238 schools and an estimated 46,000 Standard 1 learners. It has since been expanded to five more districts (Mzimba North, Ntcheu, Zomba Rural, Blantyre Rural, and Thyolo) and now reaches 1,310 schools with over 200,000 Standard 1 learners. The program was designed to integrate into Standard 1 classrooms “the five T’s,” or five major inputs that have been shown effective in improving reading scores¹: increased time to read, appropriate and sufficient texts in the mother tongue, improved direct teaching methods, and testing to measure progress and inform practice.

Did the MTPDS reading intervention— Maziko a Kuwerenga —improve reading outcomes?

Measurements using the Early Grade Reading Assessment (EGRA) methodology² compared baseline (2010), midterm (2011), and endline (2012) reading scores for the schools participating in the program and against a control group of schools not participating in the program. The results showed that while overall performance remains low, large absolute and relative gains in reading performance were achieved in the intervention schools that were not achieved in control schools. Children in control schools are demonstrating little, if any, measurable pre-reading skill, including in ‘naming the letters of the alphabet’. Although average scores for children in the intervention schools are still very low, student scores in intervention schools beginning Standard 2 are close to the level of achievement demonstrated by beginning Standard 4 learners nationwide.³ Students with measurable reading ability (those who read at least 1 correct word per minute by Standard 2), were reading an average of 19 correct words per minute. There is no statistically significant difference in scores based on gender, indicating that boys and girls are being affected equally by the intervention.

¹ Gove, A. and P. Cvelich (2011). Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute.

² www.eddataglobal.org

³ See separate report of a nationwide EGRA study in Standards 2 and 4.

While these results are encouraging, actual levels of performance in all reading skills are still very low (although higher than what is otherwise being achieved in Malawian schools), and below the benchmarks that the MoEST has set for the school system. A majority of children are still not demonstrating any measurable skills, and oral reading fluency for those who do still falls well below what is needed for children to read with comprehension. The following table presents the average scores for a few select measurements, as well as the national benchmark for that skill and results disaggregated by gender.

Table 1. Endline Results on key EGRA subtests, by Intervention Group

Subtest	Intervention group	Mean (correct per min. or percent correct)	National benchmark (Std 1)	Standard error	Boys	Girls
Letter naming	Intervention	21.5	24	0.9	21.3	21.7
	Control	1.5				
Oral reading fluency	Intervention	7.4	20	1.4	7.1	7.0
	Control	0.2				
Initial sound identification	Intervention	26%	80%	0.5	27%	26%
	Control	2%				
Reading comprehension	Intervention	5	40%	0.0	6%	5%
	Control	0%				

¹ All differences between intervention and control group are statistically significant (p-value =<.05)

Therefore, although this type of intervention is having an effect, it alone is not enough to overcome many of the underlying constraints to quality education in Malawian schools—large class sizes (average 122 in Grade 2, according to the national study conducted by MTPDS in 2012), teachers with limited training in teaching reading, a very short school day, high teacher and learner absenteeism, and persistent shortages in teaching materials and infrastructure (many classes are held under trees), for example. Moreover, children come from resource-poor environments, both economically and academically—a minority of parents are educated beyond primary school, there are few reading materials at home, they may have started school late without any prior school readiness preparation, and could be in poor health, all of which contribute to a poor foundation on which to build a learning trajectory.

What contributed to this achievement?

The MTPDS *Maziko a Kuwerenga* reading program was able to significantly improve foundational reading skills in one year. This program includes:

- One *Nditha Kuwerenga* (“I can read”) reader provided to each child in Standard 1.
- Structured routines (lesson plans) for literacy teaching provided to all Standard 1 teachers. These lesson plans brought about increased time on task; that is, an additional hour of literacy per day was implemented by teachers in all intervention schools.
- Continuous assessment of learners in specific skills.

- Eight days of Continuous Professional Development (CPD) training for Standard 1 teachers on early reading skills and on the use of the materials provided. Apart from the intervention CPDs and coaching the teachers have also been exposed to 8 days of literacy training that was spread across all schools in the country.
- Recurrent school- based coaching and support to teachers (i.e., answering teacher questions, modeling lessons, and offering feedback to improve teachers' knowledge and understanding) by Primary Education Advisors (PEAs) and project officers. Coaching was implemented as permitted by PEA's schedule and availability.
- Community mobilization in support of literacy. In this component, parents and community members were made aware of the intervention to gain their support as well as monitoring learner performance in literacy. At the end of the term teachers, learners and parents organized literacy fairs where learners and teachers showcase what has been learnt over the term.

The instruments used in the baseline and endline measurements could not detect the influence of each and every one of these components, nor other elements such as implementation fidelity and intensity of implementation in each school. However, linear regression was used to examine the effect of certain characteristics of children and schools, including the number of coaching visits a school received from MTPDS or PEAs. The number of coaching visits was the only variable that stood out as statistically significant.

Program implementation records provide additional qualitative evidence of how the reading intervention was implemented across schools over a 14 month period; that is, starting from September 2011 up to the time of EGRA 2012⁴.

- During the intervention trainings, there was high participation and attendance of teachers.
- Use of intervention materials was quite good although there were instances where teachers were not effectively following the routines.
- Standard 1 teachers have been supported by either the PEA or the project staff on effective implementation of the program.

In the course of implementing the intervention, a number of factors hampered smooth progress of the program:

- In large classes teachers do not effectively support learners with varying learning needs
- Schools are understaffed; in most cases Standard 1 teachers were assigned to teach another class thus compounding workload and limiting their ability to concentrate on the program.
- Some teachers were re-allocated to teach upper classes.

⁴ Implementation started in schools in Salima and Ntchisi in September; however, preparations started in January 2011 with the first training being conducted in April 2011 on 'Effective Literacy Teaching Practices' for Std 1-4 teachers.

- Some teachers, despite the training program, still found it difficult to distinguish between phonemic awareness, alphabetic principle and fluency routines and did not implement them correctly.
- Teachers were exposed to multiple interventions currently implemented in schools such as Read Malawi, Tikwere IRI program, and Maziko a Kuwerenga as well as the normal curriculum Chichewa lessons; all of these demand the teacher's time. For many teachers, even one hour was too much as they are used to teaching 30 minute lessons. Due to this, some teachers resorted to teaching only a few routines which they understand, thus undermining the impact of the program which build skills in a deliberate progression. However, those that are implementing it fully have their learners being able to read and write.

Though we do not necessarily have quantitative evidence of what, out of all of the inputs listed above, is most influencing the results, we can talk about what factors we examined and did not find to be related; for example, gender did not appear to be a determining factor, nor is socio-economic status. The same is true for region, language, school size, preschool attendance or a supportive home environment. Being in a school receiving the *Maziko a Kuwerenga* reading program is the only significant factor, and it is likely that it will be effective in a variety of environments and contexts.

From this we can conclude that by addressing fundamental reading skills such as phonological awareness and alphabetic principle in the first year of primary school, in a systematic and sustained way, reading can be measurably improved. The quantitative analysis confirms that children who succeed in basic skills such as letter naming and familiar word reading have higher oral reading fluency and comprehension scores. Therefore it is important to focus on these basic skills early in Standard 1. Yet carefully scripted lessons, teacher training, coaching and providing text books are only one aspect of the solution, as the evidence shows. The gains have begun to pave the road for students to becoming readers, but schooling will require continued, substantial changes to reach that goal. Additional efforts need to be made to improve classroom and we must learn more about students with zero learning gains in order to multiply the effects of the school reading programs and meet established benchmarks.

How were outcomes rigorously evaluated?

We are confident in concluding that the USAID MTPDS reading intervention, *Maziko a Kuwerenga*, was responsible for the gains measured in Standard 2 learners (who benefitted from a reading program during Standard 1). The outcomes were evaluated based on a rigorous three-year randomized control trial (RCT) design that selected learners at random from the intervention and control schools and measured reading ability using the same internationally recognized and validated methodology and instrument protocol (adapted specifically for Malawi under MTPDS). In order to ensure that the baseline and endline evaluations accurately measure changes in average reading ability, the instruments were subject to an equivalency analysis, and any changes detected in the difficulty of the items resulted in an equivalent weighting of scores. Characteristics of control and intervention schools at baseline were equivalent; most importantly all children started with the same reading ability. The two groups were also separated geographically in order to minimize the possibility of contamination, or intervention methods being adopted by control schools inadvertently.

Some limitations to the methodology include:

- The Hawthorne effect, or the possibility that intervention schools could realize they are being evaluated and may have been more faithful to the intervention program because they had been selected at baseline and were assessed again at endline. However, there were some mitigating factors. First, the same control schools were assessed at baseline and endline as well, so they also had the possibility of realizing that they may be assessed again and could also have tried to increase their performance. Secondly, all the schools in the district were treated, so schools wouldn't necessarily feel exceptional as far as being included in the program, as compared to others around them. Third, the baseline was taken two years before the endline, so as compared to same-year testing, the initial assessment could have been forgotten.
- Data collection took place at the beginning of Standard 2, in an attempt to measure gains due to inputs in Standard 1. This means that many teacher and classroom characteristics could not be measured and correlated directly with outcomes.
- Importantly, 11% of the sample of Standard 2 students were repeaters, who were in Standard 2 the previous year and not in Standard 1 where the intervention took place. This could have the potential effect of lowering the average, given that Standard 2 students nationally, and in the baseline tend to be below the scores that were registered in this endline assessment.
- It is also possible that some students in the intervention did not go on from Standard 1 to Standard 2. However, this would not likely have an effect on the comparison, since the comparison is being made from beginning of the year of Standard 2 in 2010 to beginning year of Standard 2 in 2012. So, unless school policies changed, theoretically the same profile of students would be moving from Standard 1 to Standard 2. And if it was a national policy, any change in cohort should also be reflected in the control schools.
- The methodology of data collection changed from a paper-and-pencil administration to electronic administration at midterm and endline. However, the instrument stayed the same across the two types of assessment, and the same content, stimuli and technical procedures applied in both cases. Also, the experience of the assessment does not change for the child. For example, for both paper and electronic data collection the child is tested orally then presented with a paper to read from, the child hears the same prompts, and the timed tests are given the same amount of time.
- In addition, the team of enumerators that collected data in 2010 baseline and endline were different. In 2010, the enumerators consisted of school leavers and private research assistants while in 2012 it was mainly composed of PEAs and other Ministry of Education officers who have lot of experience as teachers in primary school. The effect of this difference is uncertain, but measures were put in place to

prevent PEAs from collecting data in their own jurisdiction in order to reduce potential bias⁵.

This report presents more details on the background, methodology, context, and results of the intervention. **Section 1** introduces the program, the research design, sampling, methodology, and limitations. **Section 2** shows the differences between baseline and endline for intervention and control schools, and then **Section 3** presents the results of the different EGRA subtests for both control and intervention schools at endline only. **Section 4** provides descriptive statistics for the sample of control and intervention schools. Although these characteristics did not turn out to be directly associated with learning outcomes, they nevertheless provide a useful picture of the context learners and teachers face daily, as well as confirmation that the control and intervention schools were not significantly different. The report concludes with **Section 5**, discussing factors that have contributed to observable gains and suggesting actions that could sustain the achievements and expand on these findings.

⁵ It was emphasized that PEAs should not assess children within their own zone; however, in some cases this couldn't be avoided.

Introduction

Background

The United States Agency for International Development (USAID)-Malawi Teacher Professional Development Support (MTPDS) project is a three-year activity,

This report presents outcomes of the USAID MTPDS reading intervention—*Maziko a Kuwerenga*—for Standard 1 students that measured the evolution of primary school learners’ early grade reading skills in the Chichewa language. A random sample of learners enrolled in schools receiving the intervention and a random sample of students from control schools (not receiving the intervention) were tested using the EGRA. This section provides some background on the program, as well as the methodology used

supporting the professional development of teachers in Malawi and implementation of the National Primary Curriculum (NPC) with the goal of improving early grade reading and performance of learners. The program is made up of five major results areas:

- Result I: Strengthened Teacher Policy, Support, and Management Systems
- Result II: Enhanced Teacher Performance
- Result III: Improved Early Grade Literacy
- Result IV: Enhanced Quality of Primary Teaching and Learning Materials
- Result V: Improved Monitoring and Evaluation (M&E) Systems, focusing on teacher competencies and learner outcomes

One component of the project, as a part of Result III, has been to conduct annual, national Early Grade Reading Assessment (EGRA) studies administered in collaboration with the Malawian Ministry of Education, Science and Technology (MoEST).⁶ The reading skills tested are: letter naming, syllable segmentation, initial sound identification, syllable reading, familiar word reading, nonsense word reading, oral reading fluency, reading comprehension, and listening comprehension. Each of these subtests contains important component skills in early reading and is predictive of later performance in literacy.

Also as part of Result III, above, an early grade reading intervention, *Maziko a Kuwerenga* (Foundations of Reading), was designed in collaboration with the MoEST. In September 2011, this intervention began in schools in the districts of Salima and Ntchisi, reaching 268 schools and an estimated 46,000 Standard 1 learners. The program aims to improve Chichewa reading, and targets the “big five” reading competencies: phonemic awareness, alphabetic principal, fluency, vocabulary, and comprehension. The program includes:

- One *Nditha Kuwerenga* reader provided to each child in Standard 1.
- Structured routines (lesson plans) for literacy teaching provided to all Standard 1 teachers.
- Eight days of Continuous Professional Development (CPD) training for Standard 1 teachers on early reading skills and on the use of the materials provided.

⁶ Results of these studies are available in separate reports.

- Recurrent school-based coaching and support to teachers (i.e., answering teacher questions, modeling lessons, and offering feedback to improve teachers' knowledge and understanding) by Primary Education Advisors (PEAs). Actual level of support varies; see next section.
- Community mobilization in support of literacy.

The program was subsequently expanded to five more districts (Mzimba North, Ntcheu, Zomba Rural, Blantyre Rural, and Thyolo) in March 2012 and now reaches 1,310 schools and over 200,000 Standard 1 learners.

To determine the effectiveness of the program's design and implementation, a randomized controlled trial (RCT) with pre- and post- intervention measurements was put in place (see the next section for research design). The evaluation model included a pre-intervention

Review of Baseline and Mid-term Findings

Baseline findings at the national level showed that early reading in Malawi is very weak. Students regardless of grade level knew few letter names, read few words, and thus comprehended little of the text that was presented to them. A total of 72.8% of Standard 2 students and 41.9% of Standard 4 students could not read a single word of a short story. The percentage of zero scores on the reading comprehension questions was even higher: 97.1% of Standard 2 students and 69.3% of students in Standard 4 could not answer one comprehension question correctly.

The mid-term snapshot report concluded that the learners in the intervention area consistently outperformed learners assessed in 2010 in the same districts, and this was true across all the basic skills. The proportion of learners scoring zero was cut approximately in half. It is highly likely that these results can be attributed to the *Maziko a Kuwerenga* program because no other significant reading intervention program was taking place at that time. Furthermore, no similar gains were seen in the study's two control districts, nor in a national assessment conducted using the same instruments at the same time. Recurrent coaching (2 or more visits) by MTPDS staff appeared to have an effect on student performance

Yet although the intervention improved scores measurably, the average learner was still struggling, likely due to contextual factors such as large class sizes, inadequate nutrition and a loss of instructional time in schools, and insufficient training and follow-up for teachers.

Source: MTPDS Baseline Report and Midterm "Snapshot of progress"¹

(November 2010) baseline assessment; a mid-term "snapshot" of progress after one year of intervention (May 2012); and a final evaluation (November 2012) of a random sample of children in intervention and control schools. This report describes the outcomes of this final evaluation and discusses some of the factors responsible for differences between baseline and final outcomes. This report will show average gains of children beginning Standard 2, in schools that have been in the program for two full years.

Purpose and Research Design

The goal of the RCT design was to determine the impact of the *Maziko a Kuwerenga* intervention in which Standard 1 teachers were trained to implement an early grade reading program using direct and explicit methods to teach students how to read (see components described above). The study was designed to test two different models of intensity in the component related to recurrent school-based coaching and support to teachers:

- First, a **full intervention** group would receive training and the support of PEAs as well as a small amount of coaching from a coach dedicated to the literacy program;

- A second group, **partial intervention**, would receive training and coaching from PEAs;
- A third **control group** would not receive any intervention, which would allow for a valid counterfactual comparison.

The difference between the full and partial intervention groups was meant to determine if a light but sustainable coaching model helps to support (scaffold) teacher learning any better than support that is already available through the national education system.

In November 2010, MTPDS conducted a **baseline assessment** of 976 Standard 2 learners in Salima and Ntchisi (Intervention 1 and Intervention 2) and a control group of 480 learners randomly selected from schools in Dedza and Mwanza. The data collection was conducted at the same time, and using the same instrument, as the national EGRA study baseline, but the sampled schools were entirely separate. The sample drawn was of 48 schools: 17 schools selected from Ntchisi, 16 from Salima, and 8 from each of the control districts. Partial and full intervention groups were from both Salima (8 partial, 8 full intervention) and Ntchisi (8 partial, 8 full intervention), and the control groups were drawn from two districts that were likely to have economic and linguistic characteristics similar to the intervention groups. The two selected control districts were Mwanza (8 Control) and Dedza (8 Control).

For the May 2012 **mid-term snapshot of progress**, a random sample was taken of 210 Standard 1 learners derived from 20 of the same schools in Salima and Ntchisi that were sampled at baseline and where the intervention had been taking place for the duration of the school year. These learners were assessed using the same EGRA instrument that was used during the baseline. The performance of these 210 learners, who were in the final term of Standard 1 in May of 2012, was compared with the performance of those 595 learners who were sampled from the same 20 schools in November 2010. These two groups were treated as comparable, assuming that any possible loss of learning that may have occurred during the long holiday at the end of the school year (July to September) would have been overcome during the course of the first term of Standard 2. The separate group of control schools in Dedza and Mwanza was not included in the snapshot of progress, but was included in the final assessment carried out in November 2012. **Table 2** provides an overview of the three-year research design.

Table 2. Summary of Three-Year Research Design Outputs

Date	Study details and sample	Study write-up
Intervention baseline	<ul style="list-style-type: none"> • Intervention: 976 learners from 33 schools in Salima and Ntchisi • Control: 480 learners from 16 schools in Dedza and Mwanza • Beginning of Standard 2 • November 2010 	Unpublished internal report
Intervention midterm snapshot	<ul style="list-style-type: none"> • Intervention: 210 learners from 20 schools in Salima and Ntchisi • End of Standard 1 • May 2012 	"EGRA: Snapshot of Progress in Learner Achievement in Salima and Ntchisi Districts – May 2012," Unpublished internal report.

Intervention endline	<ul style="list-style-type: none"> • Intervention: 1,322 learners from 33 schools in Salima and Ntchisi • Control: 512 learners from 16 schools in Dedza and Mwanza • Beginning of Standard 2 • November 2012 	Present report
-------------------------	---	----------------

Between the May 2012 midterm report and this November 2012 final assessment, the reading program in schools has been ongoing, providing coaching and retraining of teachers during coach visits, and also participation in the national CPD program for literacy (that is, three 2-day training sessions in teaching reading, which have been delivered nationwide). These CPD programs have taken on board Standard 1-4 teachers to discuss and demonstrate specific literacy routines, drawing on lessons from the experiences from the literacy intervention districts.

The final data collection for a separate, nationally representative assessment series also took place in November 2012. The results of the national study are available in a separate report⁷, which includes information about the additional five districts to which the reading program was expanded in March 2012 (Mzimba North, Ntcheu, Zomba Rural, Blantyre Rural, and Thyolo).

Methodology and Limitations

Methodology

The national EGRA study, although a separate and distinct research design, can serve as a reference for this report. The same instruments were used, and the same rigorously trained enumerators were deployed. The instruments include the standard EGRA protocol,⁸ a short learner context questionnaire, and more detailed teacher and head teacher interviews. In order to ensure that the baseline and endline evaluations accurately measure changes in average reading ability, the instruments were subject to an equivalency analysis, and any changes detected in the difficulty of the items resulted in an equivalent weighting of scores. For full details of the EGRA methodology, equivalency assessments, explanation about weighting, and limitations of the instrument, please refer to the National Study Final Report (MTPDS, 2013).

Sampling

The overall sample sizes for the baseline and final data collections are presented in **Table 3**. The sample was drawn by the project's Technical Advisor using a random number generator and the complete list of schools from the 2009 education management information system (EMIS) database, which was the most recent available at the time the sample was drawn. Schools in the sample were stratified by district and then selected randomly within stratum. This ensured that some schools from each stratum were selected. Within schools, 10 children were randomly selected, stratified by gender, across all Standard 2 classrooms in the school, ensuring an equivalent number of boys and girls were selected.

⁷ MTPDS, 2013. "Malawi National Early Grade Reading Assessment Survey. Final Assessment – November 2012".

⁸ See www.eddataglobal.org for examples of the "standard" and variations across countries.

The decision to use control schools in different districts than the intervention schools was an attempt to ensure that there was little or no contamination effect between the two groups. Often times, especially when teachers from intervention and control schools gathered together for trainings similar to the CPD trainings that were being done nationally, there was an opportunity for sharing of materials, resources, and pedagogical practices between teachers and thus between intervention groups.

The numbers presented in the table below, from Ntchisi and Salima, combine Treatment 1 and Treatment 2 schools. As explained below, coaches were not consistent about visiting schools within Treatment 1 groups, so Treatment 1 and 2 were eventually combined and presented together as “treatment”. The original design was to compare the three treatment groups using equivalent samples (8 schools each), but since the two groups were merged, the result is that the overall intervention sample is larger than the control sample, as shown in Table 3 below. Additionally, due to a mis-communication with field teams during data collection, the intervention schools were over-sampled at endline resulting in about 100 more children per gender per district in the intervention schools.

Table 3. 2010 and 2012 RCT Intervention Study Sample, by Division and Standard

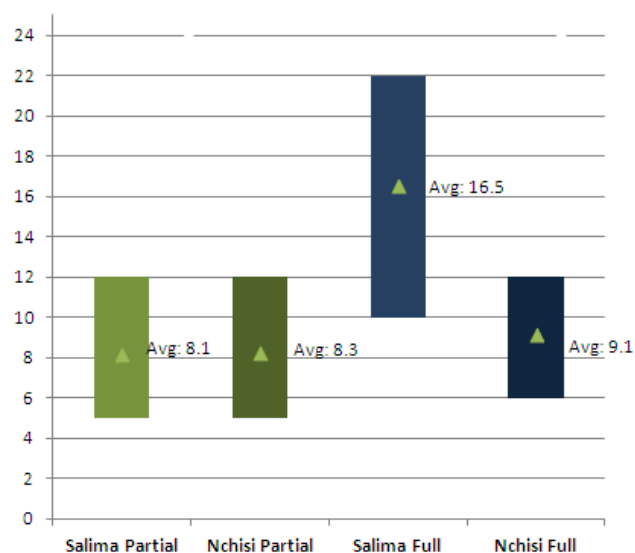
District	Schools	Baseline 2010		Final 2012	
		Male	Female	Male	Female
Dedza (Control)	8	122	117	124	116
Mwanza (Control)	8	118	123	136	136
Ntchisi (Intervention)	17	259	239	339	339
Salima (Intervention)	16	240	238	319	325
Total	49	739	717	918	916

Limitations

The original design of the RCT, as described above, intended to differentiate between two levels of intensity of coaching visits; however, in practice, coaches—both MTPDS coaches and government PEAs—visited schools both within and outside of their group. Furthermore, PEA coaching visits were not always consistent in number per school. This could be for a variety of reasons, including the very busy schedule of PEAs, and the fact that some may have more rural and distant schools. Detailed records of coaching visits were kept, and the number of coaching visits per school per academic year in each region is displayed in **Figure 1**, below. The chart shows the range of visits to any intervention school by “full” (PEA training and coaching plus MTPDS coaching) or “partial” (PEA training and coaching) intervention groups, as well as the average number of visits across schools in that group. For example, schools in the partial intervention group in Salima may have received between 5 and 12 coaching visits, with the average being 8 visits. In other words, partial treatment schools in Salima received, on average, a little less than 1 visit per month during the school year. Although there is an overall difference in the average number of schools visited in full and partial intervention groups in Salima, there are nevertheless schools in the full

intervention group that received fewer visits than schools in the partial intervention group. In Ntchisi, there is almost no difference in the range or average number of visits to the schools in the partial or full intervention group.

Figure 1: Comparison of Coaching Visits in Each Intervention District



The variation in the number of coaching and support visits in the 2 districts is due to differences in characteristics of district team, i.e. availability of PEAs due to commitment to other equally demanding tasks and lack of fuel to move around the schools. One notable factor was that for a few months the project officer was on leave on health grounds with periodic support from the project officers from Kasungu, Salima and the MTPDS central office who had equally their core duties. In the course of the implementation of the reading intervention MTPDS initiated coaching by PEAs in the districts that made most of the schools that were not supposed to be coached receiving full treatment. This consequently made no clear difference between partial and full treatment. With the support of fuel being provided to PEAs, schools which were supposed to be non-coached were coached by the PEAs as they were modeling teachers on the implementation of the reading programme.

Because this key difference in treatment groups was not maintained during implementation, this report has combined the data from the full and partial intervention groups into one single intervention—schools receiving the *Maziko a Kuwerenga* program, with the components described in Section 1 above. The number of coaching visits has been converted into a continuous variable that has been analyzed using regression techniques to determine the effect of more or less frequent coaching on reading results.

Another limitation to the ability of this research to detect and explain differences due to the intervention is that the assessment took place at the beginning of Standard 2, rather than at the end of Standard 1. Thus there may have been some loss in learning over the break period. Because data collection took place one month into the Standard 2 school year, it may be possible that any learning loss was made up for during that first month. More importantly, however, assessing Standard 2 students introduced the possibility that the random sample would select children who were in Standard 2 the previous year (repeaters) and thus may not have benefitted from the intervention inputs, which focused on Standard 1. This could cause the average scores to under represent the effect of the intervention.

Additionally, the Standard 2 teachers were interviewed, but these are not the teachers who were most responsible for what the children learned in Standard 1, and thus there is less possibility to correlate teacher characteristics with learner outcomes.

As in any research situation, the Hawthorne effect is difficult to avoid. In the case of this early grade reading program, there is a possibility that schools and specifically teachers may have been more faithful to the intervention program because they were selected at baseline. This could be even truer for the 20 schools also selected as part of the midterm snapshot, reinforcing the idea that there was going to be continued and repeated follow-up of teacher practices and student achievement. This can be avoided in the future by selecting different schools at midterm and endline, but it was felt that, due to the small number of schools being selected, it would be better to minimize other possible lurking variables such as changes in educational philosophy or socio-economic status of the school community or differences linked directly to schools, such as teachers, school communities or influence of specific PEAs from the equation.

The endline study utilized the same instrument and administration as baseline and midterm; however the data collection took place through electronic data entry software rather than paper and pencil tests. This is expected to have increased completeness and accuracy of the data collection. It is not possible to know if the use of the digital data collection method had an influence on overall scores, since errors (in administration or in data entry) could have favored either higher or lower overall scores. The most important thing to keep in mind is that the experience for the child did not change—they are reading the same stimuli under the same conditions—so the data is more accurate than at baseline because of fewer errors introduced during data entry.

Comparison of Gains from Baseline to Endline

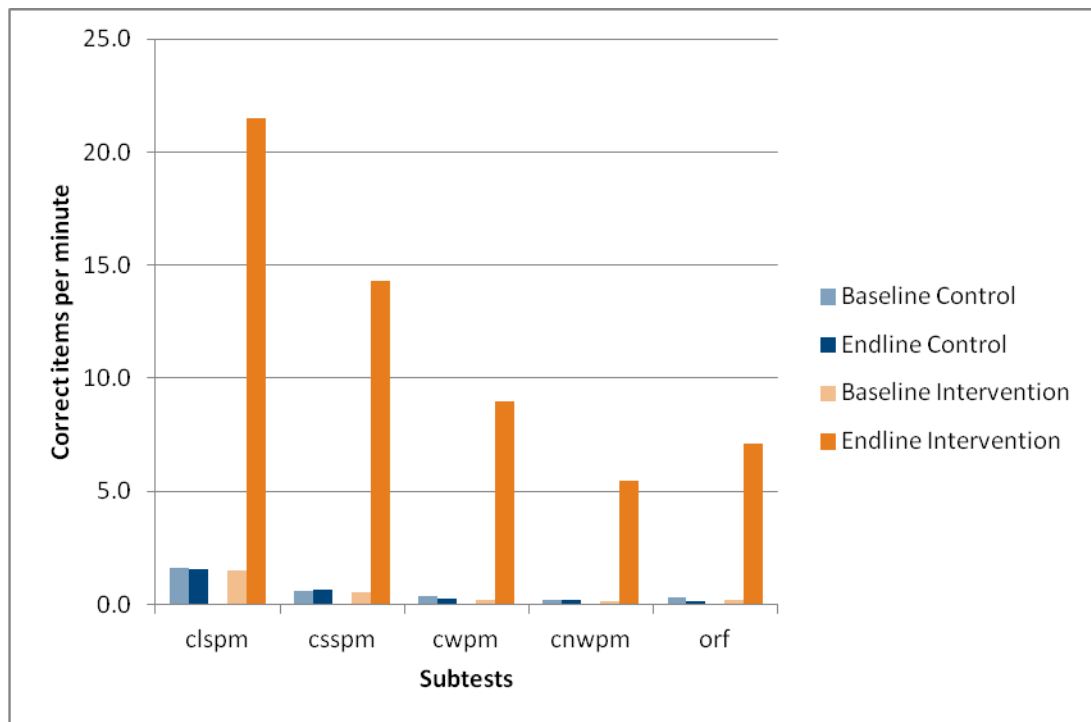
As described in the methodology section above, the design of the research aimed to detect with confidence how much of the change in reading scores was due to the MTPDS reading

Intervention and control groups started out at baseline with equivalent low scores on all fluency measures. By endline, the intervention group had increased scores significantly, whereas control schools had not increased at all. All of the effect sizes in this sample are notable, indicating that there were large actual and relative gains in ability for the Intervention group that can be attributed to the intervention

inputs. For this reason, the intervention schools and control schools were in separate districts, but were comparable in terms of socio-demographic indicators and more importantly, were similar in terms of reading skills. The only differences in scores that were statistically significant were listening comprehension and letter sound zero scores, and this was in favor of the control schools (control schools performed better at baseline).

Figure 2 below shows scores on fluency measures by intervention group and by phase—baseline or endline. It is clear that although intervention and control groups started out at equivalent levels of competency, the intervention group scores have improved, whereas the control group scores have not.

Figure 2: Comparison of Intervention and Control School Scores at Baseline and at Endline on Timed Measures



The actual **percent increase** and **effect sizes** for the intervention group are presented in **Table 4**, below. The percent increase calculation shows how much larger the gains of the intervention group were than the gains of the control group (where there were any gains in the control group). This helps us distinguish gains in ability that are most likely due to the intervention from overall gains made by the entire population for reasons other than the intervention. The effect size is a measurement of how important that gain is given the distribution of scores in each group. When describing differences relative to a specific intervention, effect size tells us more than just, ‘Does it work or not?’, but also ‘How well does it work in a range of contexts?’ (Coe, R. 2002⁹). Generally, any effect size larger than 0.5 is worth noting, since this means that 69% of individuals in the control group would fall below the average person in the experimental group (ibid.) All of the effect sizes in this sample were well above the .5 mark (with the exception of listening comprehension), thus indicating that there were large actual and relative gains in ability. This is, in some ways, not surprising given the very low starting point of the students, and should not mask the fact that scores are still very low. However, we can feel confident that the MTPDS reading intervention is effective in raising scores.

Table 4. Endline Gains (intervention group only)

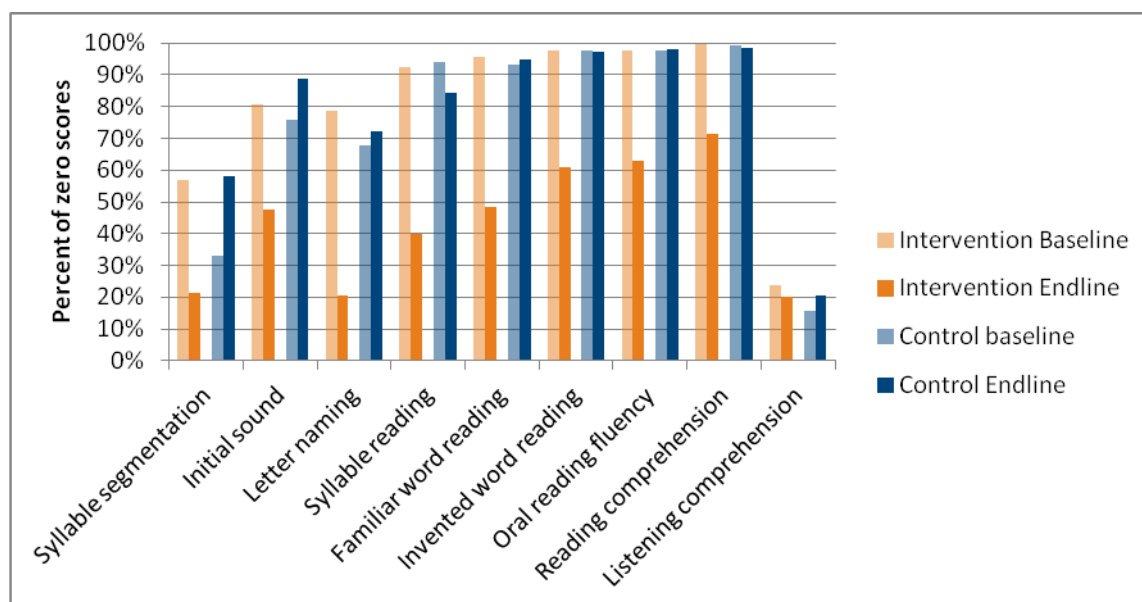
Subtest	Actual Scores (means)		% increase ($\Delta_{\text{Intervention}} - \Delta_{\text{Control}}$) / Δ_{Control}	Effect size
	Baseline	Endline		
Syllable segmentation	27%	63%	1.9	1.2

⁹ <http://www.leeds.ac.uk/educol/documents/00002182.htm>

Subtest	Actual Scores (means)		% increase ($\Delta_{\text{Intervention}} - \Delta_{\text{Control}}$) / Δ_{Control}	Effect size
	Baseline	Endline		
Initial sound identification	5%	26%	5.2	0.8
Letter naming	1.5	21.5	13.7	1.1
Syllable reading	0.5	14.3	27.0	0.8
Familiar word reading	0.2	8.9	41.8	0.8
Nonsense word reading	0.1	5.4	34.7	0.7
Oral reading fluency	0.2	7.4	35.6	0.7
Reading comprehension	0%	5%	42.7	0.6
Listening comprehension	28%	43%	0.3	0.2

Figure 3 is a graph of the change in zero scores from baseline to endline for the intervention group and control group. The EGRA administration procedures require a subtest to be discontinued when a child fails to answer correctly the first 5-10 items (depending on the task). For these children, the total score on the subtest will be zero. Whereas the zero scores declined greatly for children in the intervention schools, they actually increased or varied only marginally for the control schools.

Figure 3: Change in Zero Scores from Baseline to Endline, by Group



Note that unlike other graphs presented in this report, a small bar is desirable, since it means that fewer children showed no reading ability. For example, for the letter naming subtest, Figure 3 shows that nearly 80% of children in the intervention group could not identify a single letter name at baseline, but by endline only 21% of children scored zero. In contrast, close to 70% of children in control schools could not identify a single letter name at baseline and at endline.

Results Overview by Subtest, EGRA 2012

Summary of Average Scores by Subtest

The sample of learners assessed for this study is a random sample of students in schools from across the intervention and control districts. The size of the random sample allows us to make generalizations about the larger population.

Children in intervention schools are demonstrating large and statistically significant gains in reading compared to their counterparts in control schools. Children in control schools are demonstrating little, if any, measurable pre-reading skills, including letter identification. This is consistent with national study findings across Standard 2 children (see separate report). Although scores for children in the intervention schools are still very low, they are close to the level of achievement demonstrated by Standard 4 learners in the national study. There is no statistically significant difference in scores based on gender, indicating that boys and girls are being affected equally by the intervention.

The next two tables summarize the results of all the subtests of the 2012 assessment. **Table 5** presents the fluency-based subtests, with results expressed in terms of mean correct items per minute; and **Table 6** presents the other subtests, with results expressed in terms of mean percentage correct responses (out of either 5 or 10 items). The last two columns show means by gender.

Table 5. Endline Results on EGRA Subtests with a Fluency Component, by Intervention Group

Subtest	Intervention group	Mean (Endline) ¹	Mean (Baseline)	Standard Error	Standard deviation	Boys	Girls
Letter naming	Intervention	21.5	1.5	0.9	29.5	21.3	21.7
	Control	1.5	1.6	0.3	2.3	1.6	1.5
Syllable reading	Intervention	14.3	0.5	1.9	27.4	14.4	14.2
	Control	0.6	0.6	0.2	1.6	0.6	0.6
Familiar word reading	Intervention	8.9	0.2	1.5	17.1	9.2	8.7
	Control	0.2	0.4	0.1	0.9	0.4	0.1
Nonsense word reading	Intervention	5.4	0.1	0.9	11.5	5.6	5.3
	Control	0.2	0.2	0.1	0.8	0.3	0.1
Oral reading fluency	Intervention	7.4	0.2	1.4	16.4	7.1	7.0
	Control	0.2	0.3	0.0	1.0	0.3	0.0

¹ Correct items per minute. All differences between intervention and control group are statistically significant (p-value < .05). Baseline values in column 4 are provided for convenience, but all other values refer to endline only.

As the data show, although scores at baseline were nearly equal, at endline the children in the control schools were still not reading at all. This is the case for the majority of the children in the schools as indicated by the low standard deviation. Because there was so little variation, it also follows that the standard error was very low in the control schools. On the other hand, there was much more variation among the children in the intervention population, and thus a higher standard deviation and standard error.

The intervention group still demonstrated low performance after two years of schooling, but achieved much higher scores in letter, syllable, and word-reading tasks than their counterparts in the control schools, indicating that the foundations of literacy are being acquired through the program interventions in these schools. However, achievement was not uniform across all children; some were more affected by the intervention than others, as shown by the persistence of some zero scores. The data, disaggregated by gender and presented in the last two columns of **Table 5**, confirm that the variation is not due to gender, and that the intervention affected boys and girls equally. (Any differences shown in the table are too small to reflect actual differences in ability).

Table 6 presents results from the EGRA instrument that are untimed tasks; rather than a measurement of fluency, the scores are presented as the *percent correct out of total items possible*.

Table 6. Results on EGRA Subtests Measured by Percent Correct, by Intervention Group, at Endline

Subtest	Intervention Group	Mean (Endline) ¹	Mean (Baseline)	Standard error	Standard deviation	Boys	Girls
Syllable segmentation (10 items)	Intervention	63%	27%	0.9	5.4	62%	63%
	Control	29%	45%	0.2	2.8	27%	31%
Initial sound identification (10 items)	Intervention	26%	5%	0.5	4.7	27%	26%
	Control	2%	4%	0.0	0.6	2%	2%
Reading comprehension (5 items)	Intervention	5%	0%	0.0	0.7	6%	5%
	Control	0%	0%	0.0	0.0	0%	0%
Listening comprehension (5 items)	Intervention	43%	28%	0.01	2.1	45%	41%
	Control	41%	33%	0.01	0.1	44%	37%

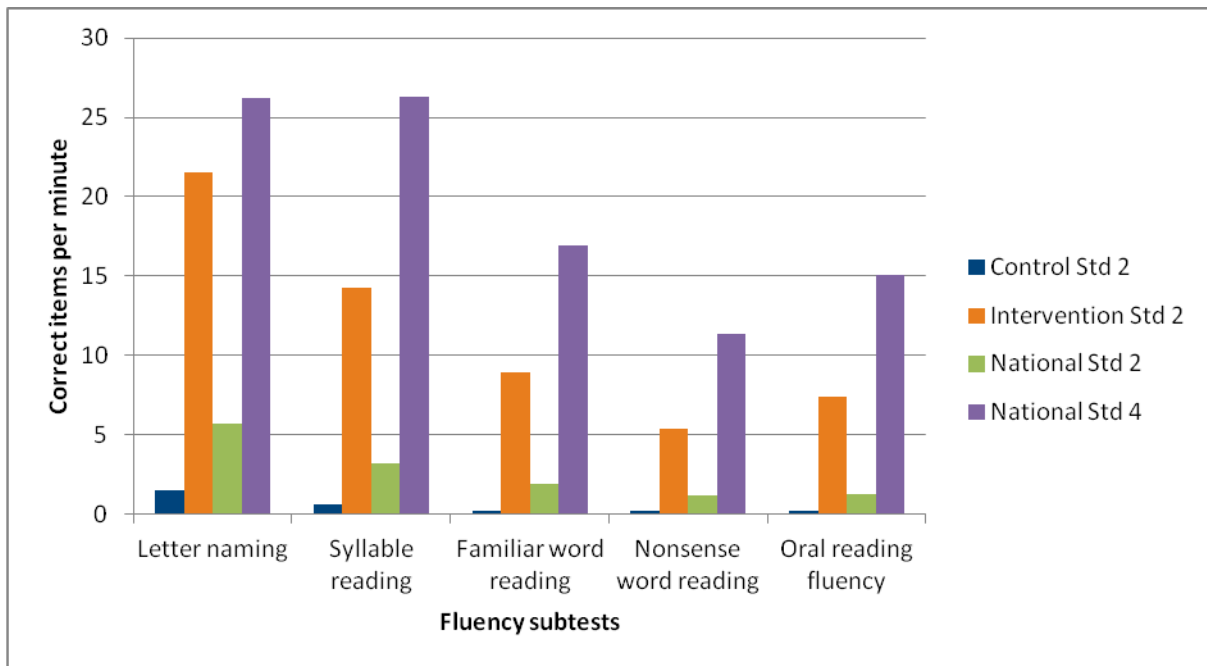
¹ Percentage of correct items out of total possible. All differences between intervention and control group are statistically significant (p -value $\leq .05$), with the exception of Listening comprehension. Baseline values in column 4 are provided for convenience, but all other values refer to endline only.

Similar to the results of the timed tests presented in **Table 5**, the results for the Intervention group were significantly higher than for the control group. Children in the control schools scored, on average, close to zero on all but the syllable segmentation task, where some children were able to identify the syllables of about 3 words in 10. There was very little difference between boys and girls on any of the subtests in both groups. Although some percentages may be higher—for example, 45% vs. 41% on listening comprehension—this actually translates into very similar performance on this exercise consisting of 5 questions.

To put these scores further into perspective, **Figure 4** below shows the intervention sample in comparison to the national study sample (see separate report for full results). After one year of MTPDS intervention, scores of children beginning Standard 2 were about half what they were for children beginning Standard 4, who had 3 years of national public schooling. Therefore, not only are the intervention schools performing better than the control schools,

they have also proven to be more effective in comparison to the entire national study sample averages. This would eliminate any doubt about whether the effect sizes for the intervention sample are due to sampling bias, if control schools were at an unfair advantage for some reason such as language or socio-economic status (even though baseline reading scores were the same).

Figure 4: 2012 Endline Comparison of Intervention Sample and National Sample



It is surprising that the control sample was also performing lower than the mean of the rest of the national sample; this is a statistically significant difference as well. The control schools were not completely isolated from the MPTDS interventions that were delivered at national scale—they also participated in the national CPD workshops. Apart from the difference in class sizes, we have not found anything in the data that would explain why these control schools might be expected to perform lower than the national sample and lower than the intervention sample. Note that the actual difference is minimal though—only 1 to 4 items, so the most important finding is that the intervention schools have results that are much higher than the study control group and the entire national averages.

2012 Endline Zero Scores

The averages above include a high percentage of students who may not have given a single correct response on a given subtest, either because they could not read, or because they were unwilling to read. Thus it is also useful to view the results across the subtests according to how large this group of non-readers was. **Figure 5**, below, presents the zero scores, by group and by subtask. Once again, the national average is also presented as a reference.

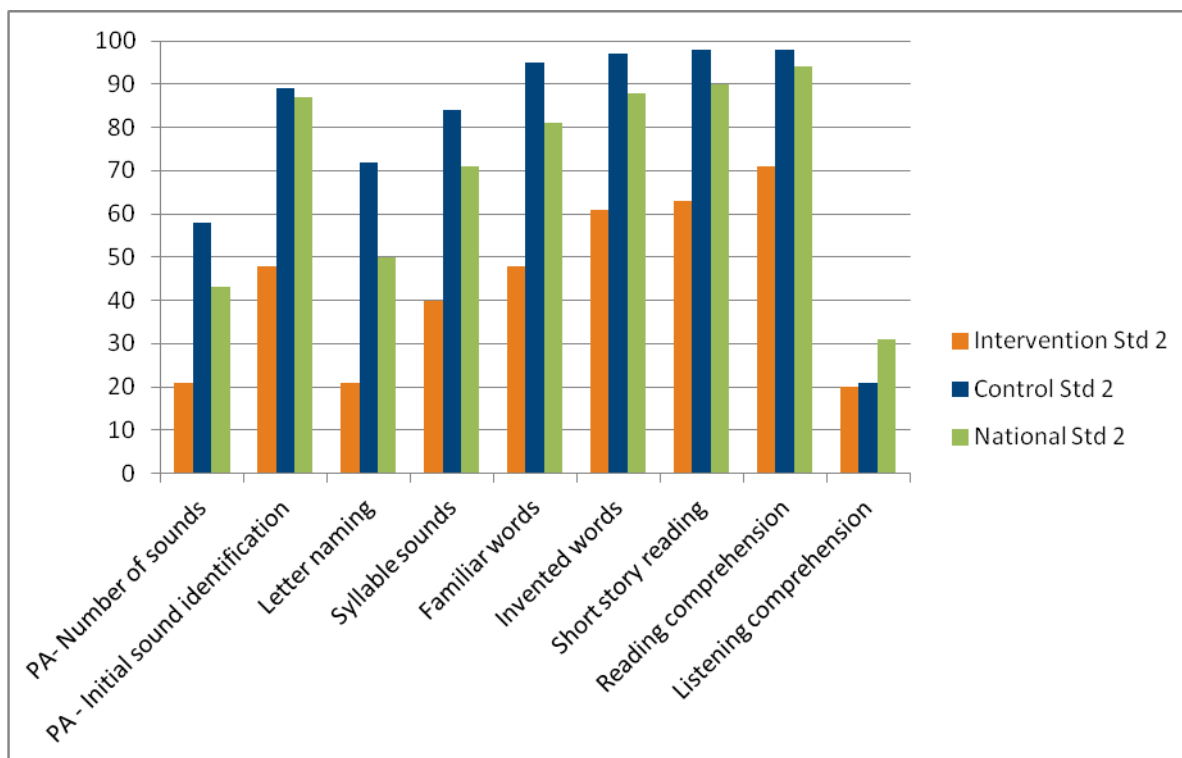
Figure 5: Percentage of Zero Scores by Group (2012 Endline)

Figure 5 shows that the proportion of students unable to read any items correctly was much higher in control schools and in the national sample than in intervention schools. Again, a smaller bar is desirable in this case. Thus the same conclusions can be drawn from **Figure 5** as from the tables above—children in control schools who did not receive any MTPDS-supported reading intervention did not learn reading skills. Some of the children were able to distinguish sounds in words, and even match a letter symbol to its conventional name, but these children were a small minority, and these skills did not translate into improved word reading ability.

By contrast, the USAID MTPDS intervention has succeeded in reducing the proportion of children who show no demonstrable reading ability based on this instrument. However, the percentages of zero scores were still very high, with more than 50% of children unable to read any words of a simple, short story. Therefore, although it is encouraging that the intervention is having a positive effect, there were still many children left behind.

It is important to note in **Figure 5**, that the listening comprehension scores for both the intervention and control groups were exactly the same. Because this is a listening task and not a reading task, it is included primarily as a measure of oral language ability in the

Interpreting Item Means

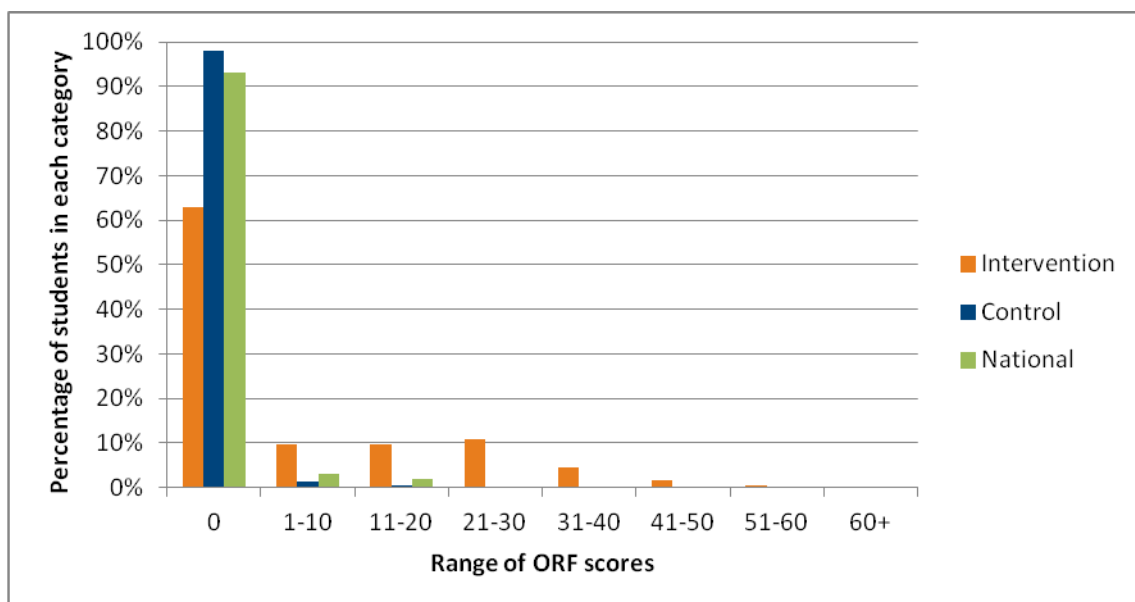
It is important to remember that mean scores presented here mask high and low scores among the group, and that the mean would be slightly different if a different sample of children had been chosen. The standard deviation shows the degree to which individual scores differ from the mean, and the standard error is an estimation of how close to the actual population mean our sample measurement is, or how much it would change if we added or removed children. So as the standard deviation increases, we see that there are children with much higher and lower abilities than the mean allows us to discern. As the standard error increases, we see whether the mean would change or not if our sample were slightly different.

language of instruction—in this case, Chichewa. The result suggests that children, regardless of school type, may have had limited ability to process Chichewa language, even orally, or were simply not accustomed to being asked questions that they needed to respond to (as opposed to recitation). The student questionnaires (see section 4) indicate that 87% of children in the sample spoke the same language at home and at school¹⁰, and the head teacher questionnaires confirm that Chichewa was the main local language and language of instruction. The fact that the ability was exactly equivalent between the control and intervention schools reassures us that the children’s prior language ability is not a factor in the differences that are seen between the control and intervention groups in the reading measurements.

Oral Reading Fluency and Comprehension

Because the ultimate goal of reading is comprehension, and because research shows there is a strong correlation between reading fluency and comprehension, it is worth looking in a bit more detail at these two subtasks. The other subtasks measure underlying skills leading to fluency—phonemic awareness, decoding, etc. Thus oral reading fluency (ORF) is often used as the best “composite” indicator of ability to decode. As mentioned above, the average score on ORF for children in the intervention sample was 7.4 correct words per minute; however, this mean is influenced by the 63% of students who didn’t read anything at all. Of the children who did read, the mean increases to 19.1 correct words per minute for the intervention group, and 8.0 for the control group. **Figure 6** below shows the distribution of scores by range of fluency. Less than 1% of children in all categories of the sample scored higher than 50 correct words per minute, so the bars do not appear in the graph although there were children in those categories.

Figure 6: 2012 Endline ORF Scores by Range



¹⁰ Although this measurement needs to be interpreted with caution—Teachers are allowed to teach in local mother tongues for the language of instruction although the curriculum is entirely in Chichewa and students are required to learn to read Chichewa, regardless of their mother tongue. So when a child reports that they speak the same language at home and school, it may mean teachers are teaching children to read in Chichewa using another mother tongue.

We can't be sure that scores in the range of 50 correct words per minute are not an anomaly. However, the proportions of children in the ranges from 1 to 40 words per minute were larger and show that children who demonstrated reading fluency did so at much higher levels than indicated by the overall mean of 7.4 correct words per minute (see Table 5). This is only the case for intervention schools; the few students in control schools who demonstrated any reading ability at all were only in the lowest range bracket.

The figure above is also useful for interpreting the comprehension scores, since we are able to see how many words the children actually read in the story they were given. The reading comprehension score is based on a total of 5 questions asked about a 59 word short story. The number of questions a child actually attempts to answer depends on the distance that the child reads in the text. In other words, a child will not be asked all 5 questions if he or she does not read the entire passage in the minute allotted for the task. Only 3 children in this sample, all from the intervention schools, were asked all 5 questions. Thus the mean percentages shown in the tables and figures above may under represent actual capacity of the children because they show the total correct answers out of the total possible of 5.

Table 7, on the other hand, shows the number of children who attempted 1, 2, 3, 4, or 5 questions, and the percent correct out of the total number attempted. This table includes only the intervention group, since very few readers in the control schools read enough text to even attempt the comprehension questions. So, for example, a child who was asked 3 questions has 4 possible scores: 0% (0 of 3), 33% (1 of 3), 67% (2 of 3) or 100% (3 of 3).

Table 7. Comprehension Scores (intervention group only)

Items attempted	N	Percent correct							
		0%	25%	33%	50%	60%	67%	75%	100%
1	288	66.0%							34.0%
2	143	8.4%			22.4%				69.2%
3	103	1.9%		7.8%			25.2%		65.0%
4	26	0.0%	3.8%		3.8%			11.5%	80.8%
5	3	0.0%				33.3%			66.7%

* Based on unweighted scores. Because baseline comprehension scores were zero for nearly all children and few children were asked any questions, no comparison between baseline and endline has not been presented for this analysis.

This presentation allows us to see more accurately the distribution of ability among the different categories of readers. What is particularly interesting is that, for any number of items attempted above 1, the largest proportion of children got all questions right (see the 100% correct column, indicating all questions asked were correctly answered. Bold items indicate that this figure is the highest proportion for this category). Therefore the actual ability of children to answer questions correctly about what they read was not as dismal as the mean scores above imply.

The situation looks very different for the control schools (see **Table 8**). First of all, no children even attempted more than 2 questions, and only 15 children attempted one or two

questions. As many children answered the questions correctly as answered the questions incorrectly.

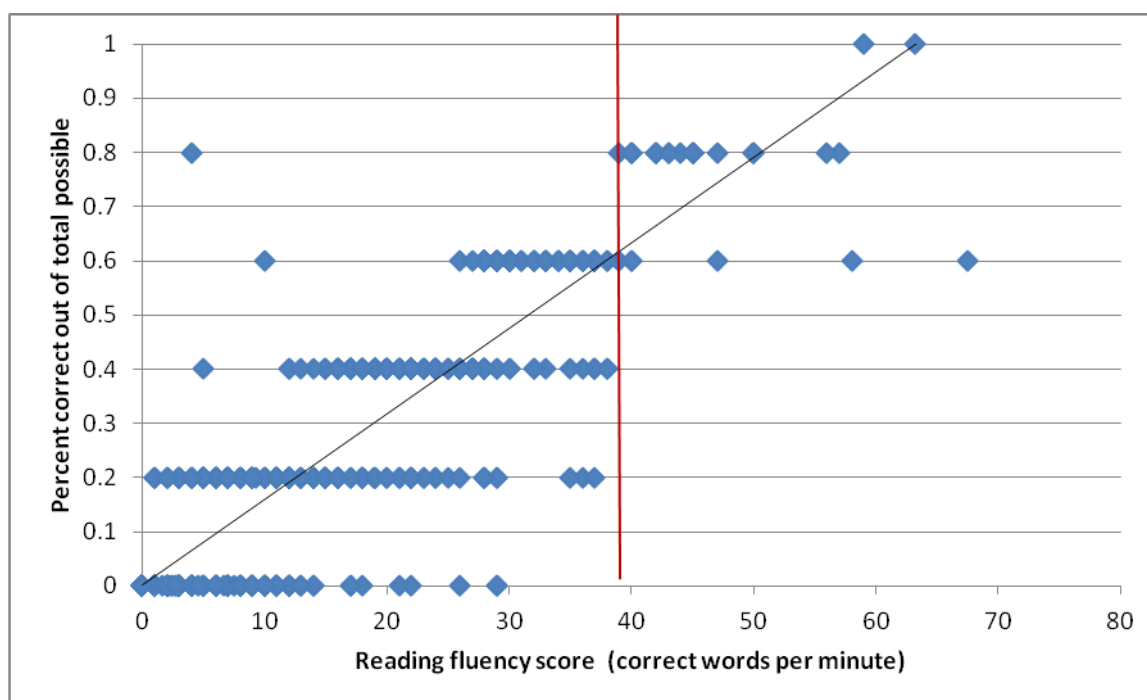
Table 8. Comprehension Scores (control group only)

Attempted	N	0%	25%	33%	50%	60%	67%	75%	100%
1	13	54%							46%
2	2								100%

* Based on unweighted scores

Whether you consider the measure as number of questions answered correctly (correct) out of questions attempted or number of questions correct out of questions possible, one thing remains true—higher reading fluency is associated with higher comprehension scores. The final figure in this section, **Figure 7** below, shows the distribution of comprehension scores in relation to the number of questions answered correctly out of the total of 5 possible. The trend line clearly favors higher comprehension scores for children with more fluency. The children who achieved 80% comprehension or better were largely reading above 40 correct words per minute, which supports prior research that suggests that a minimum of 40 – 60 words per minute (depending on language and context) is necessary to begin reading with comprehension¹¹.

Figure 7: Fluency vs. Comprehension (intervention group only)



* Based on unweighted scores

¹¹ Chichewa is an agglutinative language; in other words, words are created by joining together small units of meaning (morphemes) into longer words. Research in Kenya in Kiswahili—a related language family—indicates that one Kiswahili word equals about 1.5 English words. If this is true for Chichewa relative reading ability might be closer to 30 correct words per minute.

Descriptive Statistics

Learner Background

While most of the EGRA consists of questions and tasks that directly measure learners' ability to read, a short interview at the end of the assessment aims to find out some basic information about the child's home environment. The questions have been largely

The instruments used in this study include the EGRA instrument to measure reading ability, as well as questionnaires that gather contextual information about the home and school environment from students, teachers, and head teachers. This information will be referred to when interpreting the reading performance measured by the EGRA instrument, and it provides a general picture of the learning environment in Malawian schools today. There are limitations to the accuracy of the descriptive data and the extent to which control and intervention schools can be compared based only on this information. With that in mind, the descriptive data show that the characteristics of children in the sample were largely similar. Differences among teachers were detected, especially in terms of teaching attitudes and behaviors.

standardized across countries and provide reliable clues about factors outside of the school environment that can influence learning to read. **Table 9** summarizes learners' responses to these questions (unweighted population). Some of the questions are proxies for socioeconomic status (SES),¹² and others are factors shown in research to correlate with early learning. This information is used in two ways: first, to check if characteristics between control and treatment schools might be significantly different, indicating a sampling bias, and second, to determine whether certain characteristics are correlated with reading results. The ability of the study to draw these correlations depends on finding both variability in the characteristics of the sample and variations in the scores of the children. Baseline characteristics are also shown for convenience purposes, but caution must be exerted when making comparisons between baseline and endline; changes in characteristics do not necessarily indicate any actual evolution in abilities, behaviors, or policies—only the effect of drawing different samples. Additionally, some of the questions might have changed between baseline and endline (added, rephrased), resulting in “not available” or n/a.

Table 9. Learner Characteristics (% of sample with this characteristic or average value)

Item	Endline		Baseline	
	Control (n=504)	Intervention (n=1401)	Control (n=480)	Intervention (n=976)
Female (%)	49%	50%	50%	49%
Age (Average value)	9.9	10.2	9.1	9.1
SES: Has 0-3 items	87%	84%	89%	84%
SES: Has 4-6 items	11%	13%	10%	14%
SES: Has 7-9 items	2%	3%	1%	2%
Uses same language at home and school	94%	87%	96%	87%

¹² The questionnaire asks whether the child has at home the 9 following items: radio, telephone, electricity, television, refrigerator, interior toilet, bicycle, motorcycle, and car/truck/mill/other equivalent. For analysis, these are reduced to a value of how many total items the child has, regardless of which ones they are, and this becomes a proxy for wealth.

Item	Endline		Baseline	
	Control (n=504)	Intervention (n=1401)	Control (n=480)	Intervention (n=976)
Went to nursery/kindergarten	55%	67%	55%	61%
Was absent last year for more than 1 week	25%	24%	32%	32%
Was in Std 2 last year (Is repeating)	9%	11%	6%	9%
Has school reading textbook	7%	12%	20%	18%
Has other books at home	18%	18%	28%	23%
Mother has some post-primary education	26%	26%	16%	12%
Father has some post-primary education	39%	35%	19%	17%

* Based on unweighted data

From the descriptive data above, it appears that the characteristics of the two samples are largely the same: children come from resource-poor environments, both economically and academically. A minority of parents are educated beyond primary school,¹³ and there are few reading materials at home. The only large difference between the intervention and control group is the number of children who report having gone to nursery school, which is larger in the intervention group. Regression analysis results in large p-values for this variable, showing that in this case, prior attendance in nursery school does not have a relationship to the impact we see in this study.

The number of children with textbooks is also higher in the intervention group, but not by a large margin. On the one hand, it would be normal to expect that children in the intervention sample possess textbooks more often than the control group, since distribution of textbooks was part of the program; however, the learners tested are in Standard 2, while books were distributed in Standard 1, and children were not allowed to keep them year after year. Nevertheless, there may be some effect of this overlap being reflected in these percentages. Finally, it is important to note that 11% of the children in the Intervention sample report that they were in Standard 2 the previous year, thus they were repeating Standard 2 in the assessment year. This is important because we are attempting to measure learning that took place in Standard 1 and included specific, direct inputs. These 11% of children may not have had that input at all, and thus could cause the means to underrepresent the effect of the intervention.

Teacher Interview

Part of the research included interviews with teachers to learn more about their background, attitudes, and practices. However, because the teachers interviewed were in Standard 2 and had only been working with the sampled children for one month at the time of the assessment, the results of the interview will not reflect the influence of the teacher on

¹³ Note that the figures for the mother and father's education levels are reported by the child, and so not necessarily accurate. Furthermore, children mostly answered that they did not know how much education their parents have (45-60% of the time), so the figures reflect only about half of the sample who gave any answer at all.

the reading scores measured by this round of data collection. For that reason, **Table 10** only presents a few key descriptive characteristics about the teachers who responded to the questionnaire. This table must be read as only indicative of the overall school climate in the sampled schools and not as factors having a direct influence on the skills measured in this report.

Table 10. Teacher Characteristics (% of sample with this characteristic or average value)

Item	Endline		Baseline	
	Control	Intervention	Control (n=11)	Intervention (n=27)
Female (%)	53%	30%	55%	41%
Age (average value)	36	41	43	39
Class size (average value)	153	106	173	105
Is a trained teacher (has received pre-service training)	80%	97%	91%	100%
MSCE (4 years)/JCE (2 years) qualified	80%/20%	64%/33%	n/a	n/a
Average years in the teaching profession	12	17	18	14
Has sufficient teaching materials	13%	20%	27%	15%
Teacher has meetings with parents	40%	70%	91%	52%
Has received feedback from head teacher about how to teach reading in Chichewa	60%	76%	n/a	n/a
Has received coaching visits from PEAs	60%	90%	n/a	n/a
Average MTPDS coaching visits	0	5.8	n/a	n/a
Pupils sounded out unfamiliar words every day	47% (avg=3.5)	73% (avg=4.4)	18% (avg=3)	52% (avg=3.8)
Pupils read aloud to teacher or others every day	33% (avg=3.5)	63% (avg=4.3)	73% (avg=4.7)	70% (avg=4.4)
Pupils were assigned reading to do every day	67% (avg=3.9)	80% (avg=4.5)	27% (avg=3.1)	41% (avg=3.0)
Never checks homework	47%	10%	27%	30%
Believes a child should be able to identify letter names in Std 1	40%	93%	27%	41%
Believes a child should be able to sound out unfamiliar words in Std 1	20%	77%	0%	15%
Believes a child should be able to understand stories they read in Std 1	0%	43%	9%	0%

Some of the key differences that stand out between control and intervention schools,

according to **Table 10**, are the types of teaching resources, behaviors, and attitudes noted by teachers who were interviewed. Class sizes were also notably larger in control schools. Further analysis of the 2012 intervention sample data (presented in the next section) reveals, however, that class size is not correlated with reading outcomes for this data set. This may partially be due to the fact that all reading outcomes are very low. In the national study, however, class size was shown to have an effect on reading outcomes. Therefore there is a possibility that control schools were at a disadvantage because of class size, and this may further explain why the control schools performed lower than even the national sample averages. However, since the intervention schools also performed well above the national averages as well as above the control schools, it is unlikely that the gains seen in this report are due only to class size differences.

Teachers in intervention schools were more likely to be trained and to teach more than just Standard 2. They held meetings with parents more often, assigned and reviewed homework, and conducted activities in the class that reflect the skills being assessed (sounding out words, reading aloud). They also appeared better resourced and received more visits from PEAs. Particularly interesting is the difference in expectations among the teachers in intervention schools compared to control schools; teachers in the intervention schools believed children should be capable of much more in Standard 1 than did teachers in control schools. Bearing in mind results are self-reported and not otherwise verified, the findings need to be viewed with some caution, and the exact influence of the MTPDS intervention on these attitudes and behaviors cannot be discerned from this questionnaire alone.

It is also useful to see what types of materials teachers were using in their classrooms. **Table 11** provides the results of the survey questions that asked teachers whether each type of reading material was used in the classroom. Both control and intervention schools had and used the Primary Curriculum and Assessment Reform (PCAR) materials and Read Malawi supplementary readers equivalently, but control schools were also still using the old curriculum textbooks more than intervention schools did. This may be explained by the fact that the intervention schools then had the MTPDS-provided readers and teacher’s guides, which replaced the old curriculum materials for them.

What is also worth noting is that teachers in the control as well as intervention are exposed to *Tikwere!* Interactive Radio Instruction and Read Malawi programmes which are also timetabled. However, treatment schools have an addition hour of literacy per day for Standard 1 which has come with the reading intervention program.

Table 11. Materials Used in Class

Group	PCAR Chichewa textbook from MIE	PCAR Chichewa teachers’ guide from MIE	Old curriculum textbooks	Other books	<i>Nditha Kuwerenga</i> reader	<i>Maziko a Kuwerenga</i> teacher’s guide	Read Malawi supp. readers
Control	93.3%	93.3%	66.7%	6.7%	0.0%	0.0%	20.0%
Intervention	93.3%	93.3%	46.7%	0.0%	83.3%	76.7%	26.7%

The fact that some intervention schools were not using the new teachers’ guides and readers is an important finding, and may require looking into further to determine why.

However, we need to recall that the assessment was conducted with Standard 2 teachers for the intervention that was designed for Standard 1. We also know that many teachers are re-assigned from lower to higher grades, and that there is a lot of teacher turnover. Finally, the program has been expanded to provide a consignment of reading materials for teachers in Standards 1-4 that constitutes teacher's guide and a child's reader. These will act as resource books for teachers as they contain 5 key components of reading. Thus it is possible that schools are still working out teaching assignments and materials distribution between Standards and classrooms. On the other hand, we can see that since no control schools were using the MTPDS *Nditha Kuwerenga* readers, there has been no spillover effect from the intervention schools to the control schools in this area.

Head Teacher Interview

In each school visited, head teachers were also interviewed to get information about head teachers, characteristics of the school and pedagogical support provided to the teachers. As above, distribution of some of the answers to key questions from this survey is provided in **Table 12**, below, by control and intervention group. The answers pertain only to the respondents (head teachers or deputy head teachers) or to their school.

Table 12. Head Teacher and School Characteristics (% of sample with this characteristic, or average value)

Item	Endline		Baseline	
	Control	Intervention	Control (n=13)	Intervention (n=30)
Female (%)	13%	3%	25%	7%
Years in position as head teacher or deputy	7.0	3.6	9.2	7.2
Highest level of education is MSCE/JCE	94%/6%	84%/16%	n/a	n/a
Teaching hours per week (average)	11.5	22.8	15	15
Never provides instructional support to teachers	25%	16%	30%*	30%*
Has received training in school management	n/a	n/a	77%	70%
Has received training in implementing an early grade reading program	56%	91%	61%	27%
Has supported teachers in the pedagogy of teaching reading	69%	84%	77%	57%
Is satisfied with the performance in reading at the school	6%	53%	31%	13%
Teachers lesson plans are reviewed on a daily basis by the head teacher	n/a	n/a	92%	70%
Never observes teachers in their classrooms	25%	16%	0%	23%
There is clean water on premises	63%	84%	85%	77%
There is electricity	6%	6%	8%	0%

Item	Endline		Baseline	
	Control	Intervention	Control (n=13)	Intervention (n=30)
School has a library** (according to teacher/head teacher)	20%/25%	23%/31%	27%/15%	11%/27%

* “Never” was assumed to be those who did not provide any number.

** No standard definition of a library was used, so that may explain the variation in responses between teachers and head teachers.

The proportion of female head teachers is low, particularly in intervention schools, and much lower than the proportion of female teachers (see Table 10). This may be explained by the rural nature of the schools in the sample, where it is difficult to retain female teachers who have a tendency to marry and move to urban centers. However, there may also be gender bias in promotion practices. Head teachers in intervention schools were slightly less educated than in control schools who had a larger proportion of head teachers with the MSCE (four-year diploma) than the JCE (two year diploma). Note that when teachers were asked whether there was a library at the school, they answered affirmatively less often than when head teachers were asked. However, in both cases, the intervention groups had a slightly higher frequency of libraries at the school than control schools. A notable difference between control and intervention schools is the number of head teachers who have received special training in implementation of an early grade reading program. We would of course expect that all head teachers in the implementation schools had received this training, but there may be cases where the head teacher was replaced or absent from the training for some reason, which explains why the number is less than 100%. There may also be confusion in the way the question was asked and what, specifically, it was referring to. Almost as important, however, is the number of head teachers in control schools who say they have received training in early grade reading programs. There were some follow-up questions pertaining to who organized the program and who invited them, but the answers are not specific enough to allow us to determine exactly what program this would have been or what the content covered. Similarly, while there were more head teachers who supported their teachers in teaching early reading in Chichewa, the majority of control schools report doing the same. Again, the questions alone don’t provide enough information on what the content of this support is, or even if “the pedagogy of reading” means the same thing in control schools as in intervention schools.

Another important difference is that head teachers in intervention schools were absent much more than in control schools. Again, the reason for these absences was not specified, other than that it was for “official school business.” MTPDS may want to investigate whether any elements of the program were calling head teachers away more frequently, or even if there were any special events in the month preceding data collection that might account for higher absences.

Finally, the difference in reported satisfaction by head teachers is of interest. It may be that head teachers from the intervention group may be noticing the changes in reading ability and therefore feel more satisfied with progress. Nevertheless, since performance remains low and not meeting established benchmarks, more training and awareness-raising may be

needed to ensure that head teachers are encouraged to continue to take measures to improve results.

In order to give any judgment about whether performance is satisfactory or not, head teachers should be in a position to know what the level of achievement is. The next table, **Table 13**, provides information on how head teachers were assessing progress of their students. Regardless of method, it appears that head teachers in the intervention sample were using more methods more often.

Table 13. How Do You Know Your Children Are Progressing?

Group	Classroom observation	Monitor test results	Evaluates orally	Reviews homework	Teacher reports	Other
Control	63%	56%	31%	31%	44%	63%
Intervention	78%	78%	47%	50%	72%	78%

Other ways that head teachers indicated that they monitored student progress were: through attendance records, daily assessment, continuous assessment, or Standard 8 selection. Thus we see that head teachers use a variety of methods to check learners' progress. However, the key point is how effectively they use the results to improve the teaching and learning of reading in their schools and how frequently they monitor progress of learners, but this instrument did not detect variations at that level.

Conclusions and Recommendations

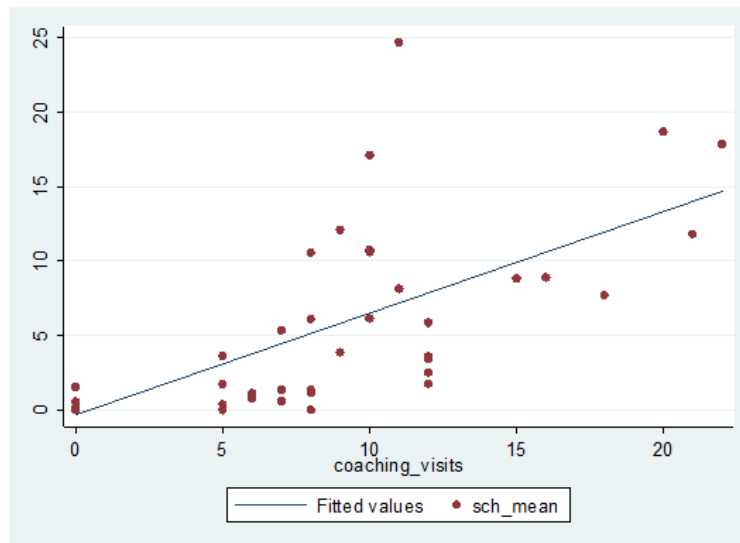
Factors that affect reading outcomes

A report on outcomes of the early grade reading assessment generally includes a section identifying and explaining factors that are associated with learner achievement. This information is helpful because educators and policy makers can use it to prioritize actions. However, there are always limitations to this type of analysis. First, analyses cannot be taken to mean that the relationships are necessarily causal or exhaustive—certain factors may simply vary along with learner achievement, and other factors exist that were not measured in this study. Additionally, as mentioned earlier, because this assessment was carried out at the beginning of Standard 2, we cannot associate teacher characteristics with reading outcomes that would have been primarily the result of another teacher's efforts which is a major limitation. Also, finding factors associated with achievement depends on finding natural variations among the sample. In this study, the characteristics of the groups sampled varied little, except for the existence and intensity of the MTPDS reading intervention.

Given those limitations, it is not surprising that there were very few correlations that stood out as statistically significant in this study. The only variable that was statistically significant using linear regression was the number of coaching visits. **Figure 8** shows that coaching visits have a strong positive relationship with a child's reading ability (as coaching visits increase, so do student scores). Analysis shows that each additional coaching visit results in an average increase of 0.65 words per minute for each child in the class (p -value = 0.001). While this seems small, coaching visits are intended to be done in sets of 10 or more, potentially influencing a child's reading level by as much as 6.5 words per minute over the course of a

school year. But more importantly, coaching visits can greatly decrease the proportion of children who score zero words per minute, thus not reading at all. The students in schools which received coaching visits were 24% more likely to successfully read some of the passage for each coaching visit their school received (odds ratio = 1.245, p-value 0.009). It is also suspected that these gains are cumulative, and that the training that a teacher gets one year will be passed on to the children they teach for many years to come. Additional coaching in these subsequent years could improve these benefits even more.

Figure 8: Relationship Between Coaching Visits and Oral Reading Fluency (Intervention sample only, endline)



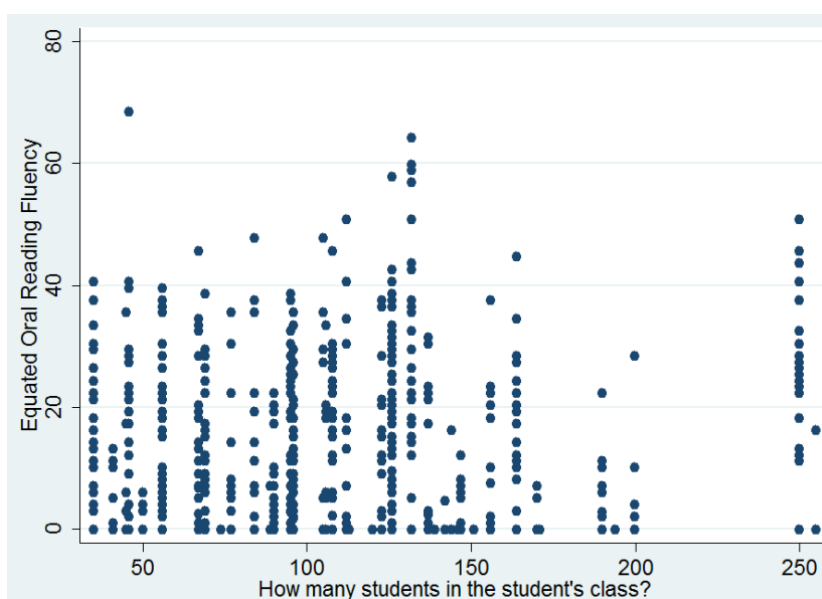
Again, coaching is a variable that we can measure and analyze relatively easily in terms of frequency, though we only know from program intervention documentation and qualitative evidence what goes on during coaching visits and how this influences teaching practice. Through coaching teachers develop confidence in teaching literacy lessons effectively and the coach serves to ensure mastery of skills by the teacher. In addition the demonstration sessions act as models to teachers. One important aspect of coaching is that of assessment of learning; in the course of teaching teachers assess learners level of literacy and consequently come up with ways to support learners not performing well. Therefore on the one hand, the coaching visits serve as refresher training to improve teaching behaviors and implementation of the lessons. On the other hand, coaching visits may also serve as an accountability tool, so that teachers who know that a coach will be coming to visit and check their progress may be more inclined to stay on track with the program and show improvement. Either way, coaching seems to be instrumental to ensuring systematic instruction.

However, being able to measure and analyze the frequency of coaching visits does not mean it is the main or only factor responsible for the growth in children's scores. The reading intervention, which includes coaching as one factor, is responsible for the growth. Coaching visits were specifically related to implementation of paced, direct instruction in reading skills so the influence of one and the other cannot necessarily be separated (i.e., we cannot assume that introducing coaches but not new instructional methods and materials to control schools would have made a difference). Whether it is the content, the teaching method, the

increased time on task, or a combination of the above cannot be discerned through this study methodology. However, the important conclusion is that clear, deliberate, appropriately sequenced instruction in fundamental pre-reading skills will bring children closer to the goal of reading than the current methods that are widespread in Malawian schools.

All other variables from the student exit interview, teacher and head teacher questionnaires were analyzed using regression with and without covariates of treatment and year (baseline or endline), and no other statistically significant results were found. This is frustrating in some ways—we would hope to see that having textbooks in school, or books in the home (p-value = 0.232, standard error=2.1) makes a difference, for example. However, there just isn't enough variation in the sample to draw these conclusions. On the other hand, we can also be relieved that things such as socio-economic status, and class size are not necessarily determinants of a child's reading ability (though the same caveat about lack of large variation across the sample still applies). See, for example, Figure 9, showing effectively no relationship between reading and class size—whether class size is small or large, there is still a range of ability from zero to above 40 correct words per minute found among the students.

Figure 9: Relationship Between Class Size and Oral Reading Fluency (Intervention sample only, endline)



Summary and recommendations

Through the *Maziko a Kuwerenga* program, children improved their basic reading skills, in particular letter and syllable-level tasks. In fact, for letter naming and syllable segmentation the intervention schools are performing close to the Standard 1 benchmarks set by the MoEST, as show in **Table 6**, below (see National study report for further details on benchmarks.)

Table 6. Intervention outcomes compared to national suggested benchmarks

Subtest	Measure	Suggested benchmark Std. 1	Intervention mean (endline)
Letter naming	correct letters per minute (clpm)	24+	21.5
Syllable segmentation	% correct	70%	63%
Initial sound identification	% correct	80%	26%
Syllable reading	correct syllables per minute	30	14.3
Familiar word reading	correct words per minute	20	8.9
Nonsense word reading	correct words per minute	15	5.4
Oral reading fluency	correct words per minute	20	7.4
Reading comprehension	% correct	40	5%
Listening comprehension	% correct	60%	43%

Evidence from EGRA studies worldwide, as well as the national study completed in Malawi in 2012, confirm that there is a strong predictive relationship between these early skills and later reading fluency. Therefore, these children have a much better chance at succeeding in reading and in school, later on. However, the intensity of inputs still is not sufficient to have raised overall means to benchmark levels. This is partially due to the persistence of a large number of zero scores across the subtests, which lowers the means substantially. As reported earlier, the mean oral reading fluency increases to 19 cwpm—just one less than the benchmark—when children with zero scores are removed from the analysis.

The instruments used in this study did not capture all possible factors related to the consistency or intensity of inputs among intervention schools, however, qualitative monitoring by the project suggests some explanations:

- The quality of coaching varies among PEAs, and more training and tools are needed to ensure high quality and productive coaching across schools.
- Teachers still do not have strategies for using books, where available, to enhance teaching and learning of reading.
- Some teachers, despite training, still struggle with the sounds of the Chichewa alphabet. More training and coaching is needed for teachers to master the skills.

International studies also confirm the relationship between reading fluency and comprehension. It is not surprising that comprehension scores are so low for this group of Standard 2 learners, since reading fluency was also very low. However, as children demonstrated more reading fluency, they were also increasingly likely to answer the comprehension questions correctly (see Table 7). The children in this sample who read with 80% comprehension or better also read with close to or above 40 correct words per minute. Therefore although comprehension strategies, and particularly listening comprehension and speaking ability, are important at all ages, an instructional focus at the Standard 1 level must be on the fundamental early skills of phonological awareness and alphabetic principle, coupled with practice applying these skills to decoding of new words. Regular practice in a

print-rich environment will increase the automatic recognition (i.e., fluency) of words translating into increased fluency.

Although this study did not find a significant relationship between class size and reading outcomes, the national study found that class size has a strong negative relationship with a child's reading ability. "[E]ach additional child in a classroom decreases the overall average of the class by 0.02 words per minute (coefficient = -0.025, p-value = 0.005). On the scale of classrooms in Malawi, reducing class size down to 50 students could increase scores by as much as 3 words per minute in some schools. Similarly, the probability of a child not being able to read at all increases 1% for every additional classmate the child has (odds ratio = 1.0055, p-value = 0.012)." (MTPDS, 2013). Additionally, according to the national study, a child with a school textbook is more than twice as likely to read more than 40 words per minute by Standard 4. Therefore, it is still possible that the inconsistency in scores is affected by the challenge of large class sizes and lack of textbooks for every child.

Therefore many of the basic recommendations made in the MTPDS baseline report are still valid and need to be followed through: reducing class sizes, increasing time spent learning basic literacy skills in P1, teaching sounds of all letters of the alphabet, revising early reading texts to focus on decoding and word-building skills, and training teachers to use these strategies in the classroom in an equitable manner. This study shows what is possible when these strategies are introduced in Standard 1 even where constraints to teacher capacity, retention, class sizes and local resource mobilization are present. An important follow up study would be to find out how this cohort of children fares when they are in Standard 3 and 4. We would expect that they will use these pre-reading skills as tools to accelerate their literacy acquisition and will be fluent readers sooner than their peers, but this also depends on having some kind of sustained instruction with appropriate grade-level reading materials to support that goal. Therefore a follow-on program to develop Standard 2-4 curriculum that build on the sequence and scope of this initial Standard 1 program and expand teacher training and materials provision nationally is recommended.