



USAID
FROM THE AMERICAN PEOPLE

MALAWI

ABE/LINK

**MALAWI TEACHER PROFESSIONAL
DEVELOPMENT SUPPORT**

**MALAWI NATIONAL EARLY GRADE
READING ASSESSMENT SURVEY
FINAL ASSESSMENT - NOVEMBER 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 activity by Creative Associates International, RTI International, and Seward Inc.

Malawi Teacher Professional Development Support (MTPDS)

Malawi National Early Grade Reading Assessment Survey Final Assessment – November 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, Revised May 21, 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(s) and do not necessarily reflect the views of USAID or the United States Government.

Table of Contents

Acknowledgments.....	vi
Foreword.....	vii
Abbreviations.....	viii
Executive Summary.....	9
Background.....	9
Findings.....	10
Conclusions.....	11
Introduction.....	13
Background.....	13
Purpose and Research Design.....	14
Methodology.....	16
Instruments.....	16
Grade and Language.....	17
Sampling.....	18
Data Collection.....	19
Limitations.....	19
Descriptive Statistics.....	20
Learner Background Questionnaire Results.....	20
Teacher Background Questionnaire Results.....	22
Head Teacher Background Questionnaire Results.....	24
Results by Subtest, EGRA 2012.....	25
Overall Results by Subtest.....	25
Phonological Awareness Subtests.....	28
Oral Reading Fluency Scores.....	31
Reading and Listening Comprehension.....	32
Comparison of Baseline to Endline.....	35
Sample Populations.....	35
Comparison of 2010, 2011, and 2012 EGRA Data.....	36
Evolution in Scores by Subtest and Distribution of Scores.....	39
Benchmarks and Targets.....	41
Results Disaggregated by Gender and Region.....	43
Results by Gender.....	43
Regional Results.....	44
Factors Predicting Achievement, EGRA 2012.....	48
Student Factors.....	48
Students who read 40 correct words per minute or more.....	49

Teacher and School-Level Factors	52
Conclusions and Recommendations	53
Recommendations for Action	55
Annex 1. Statistically Significant Regressions for the Entire National Sample	57
Annex 2. Statistically Significant Regressions on the Subpopulation of 40+ cwpm	58
Annex 3. EGRA 2012 Instrument.....	62
Annex 4. Benchmarks and Targets for EGRA, 2012–2017	64
Annex 5. Technical Notes on National Study Sampling	65

List of Tables

Table 1: Summary of Three-Year Research Design Outputs.....	15
Table 2: Cronbach’s Alpha for EGRA 2012	17
Table 3: 2012 Study Sample, by Division and Standard.....	18
Table 4: Learner Characteristics (% of sample with this characteristic, or average value)	20
Table 5: Teacher Characteristics (% of sample with this characteristic or average value).....	22
Table 6: Materials Used in Class	23
Table 7: Head Teacher and School Characteristics	24
Table 8: How Do You Know Your Students Are Progressing?	25
Table 9: Results on EGRA Subtests, by Standard	25
Table 10: Endline Scores Excluding Zero Scores on Oral Reading Fluency	31
Table 11: Distribution of Comprehension Scores (Standard 2 and Standard 4).....	33
Table 12: Changes in Descriptive Statistics over Time.....	35
Table 13: Subtest Results for All Three Years, and Effect Sizes	36
Table 14:Percentage of Learners Meeting Benchmarks and Performance Targets from 2010 to 2012	41
Table 15: Standard 2 Results Disaggregated by Gender.....	43
Table 16: Standard 4 Results Disaggregated by Gender.....	44
Table 17: Fluency Subtest Results, by Division, Standard 2.....	45
Table 18: Subtest Results, by Division, Standard 4	45
Table 19: Changes in Regional Scores over Time.....	47
Table 20: Linear Regression: Grade vs. Oral Reading Fluency	49
Table 21: Linear Regression: Gender vs. Oral Reading Fluency	49
Table 22: Linear Regression: Repeating vs. Oral Reading Fluency.....	49
Table 23: Subtest Results for All Three Years, and Effect Sizes	51
Table 24: Linear Regression : Class Size vs. Oral Reading Fluency.....	53

List of Figures

Figure 1:	Scores on Measures of Fluency.....	27
Figure 2:	Scores on Other Measures.....	27
Figure 3:	Comparison of Means with Zero Scores Removed.....	28
Figure 4:	Scores on Syllable Segmentation Task.....	29
Figure 5:	Distribution of Initial Sound Scores.....	30
Figure 6:	Endline ORF Scores by Range.....	31
Figure 7:	Reading and Listening Comprehension Scores by Standard.....	33
Figure 8:	Fluency vs. Comprehension (Standards 2 and 4 combined).....	34
Figure 9:	Letter Naming and ORF.....	35
Figure 10:	Invented Word Reading and ORF.....	35
Figure 11:	Syllable Reading and ORF.....	35
Figure 12:	Familiar Words and ORF.....	35
Figure 13:	Evolution of Fluency Measures from Baseline to Endline.....	38
Figure 14:	Baseline ORF Scores by Range.....	40
Figure 15:	Endline ORF Scores by Range.....	40
Figure 16:	Oral Reading Fluency Scores by Range from Baseline to Endline.....	41
Figure 17:	Standard2 Fluency Scores by Region.....	46
Figure 18:	Standard4 Fluency Scores by Region.....	46
Figure 19:	Subset of Students Who Read with 40 correct words per minute Fluency or Above.....	50
Figure 20:	Reading Comprehension Scores by Subgroups of Achievement.....	50

Acknowledgments

The United States Agency for International Development (USAID) Malawi Teacher Professional Development Support (MTPDS) project thanks the Ministry of Education, Science and Technology (MoEST) for its participation in piloting, training, data collection, data analysis, and report preparation, during the 2012 Early Grade Reading Assessment (EGRA) 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 USAID for 2010 and 2011.

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 students 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 2012 EGRA final assessment report documents the performance of 3,000 Standard 2 and 4 learners, from a random selection of 150 schools nationwide. The results of this nationally representative sample present a challenge to everyone with an interest in primary education in Malawi. It should be read along with the final EGRA assessment report of the MPTDS reading intervention that took place in Salima and Ntchisi, and has been scaled up as an example of how to improve reading outcomes in Standard 1.

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

Abbreviations

ABE/LINK	Assistance to Basic Education/Linkages to Education and Health initiative
CERT	Centre for Education, Research and Training
CfBT	Consortium for British Teachers
CLS	Centre for Language Studies
CPD	continuous professional development
clpm	correct letters per minute
cspm	correct syllables per minute
cwmp	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
IRI	Interactive Radio Instruction
IRR	Inter-Rater Reliability
M&E	monitoring and evaluation
MANEB	Malawi National Examinations Board
MIE	Malawi Institute of Education
MoEST	Ministry of Education, Science and Technology
MTPDS	Malawi Teacher Professional Development Support project
NPC	National Primary Curriculum
ODL	Open and Distance Learning
PCAR	Primary Curriculum and Assessment Reform
PEA	Primary Education Advisor
PSIP	Public Sector Investment Program
PTA	Parent–Teacher Association
RCT	randomized controlled trial
RTI	RTI International (trade name of Research Triangle Institute)
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
TTC	teacher training college
USAID	United States Agency for International Development

Executive Summary

Background

The United States Agency for International Development (USAID) Malawi Teacher Professional Development Support (MTPDS) project is a three-year activity 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. This report presents the results from a nationally representative study of primary school learners' early grade reading skills in the Chichewa language, undertaken near the beginning of the school year, in November 2012. It is the third in a series of national samples designed to identify overall skills and gaps in order to inform national-level policies and strategies and determine the influence of policies and projects to address reading over time.

This study followed an initial baseline study, conducted in November 2010, and a second assessment conducted in November 2011, using the Early Grade Reading Assessment (EGRA) methodology. The reading skills tested are: initial sound identification, syllable segmentation, letter naming, syllable reading, familiar word reading, nonsense word reading, oral reading fluency, reading comprehension, and listening comprehension. Each of these subtests measures important component skills in early reading that are predictive of later performance in literacy.

The series of national surveys aims to inform the Ministry of Education, Science and Technology (MoEST), development partners, and other relevant stakeholders of the current status of learner performance in early grade reading in Malawi and to contribute to policy dialogue as well as general public awareness and support for improving education. The results are also expected to help MoEST and other stakeholders determine the impact of major systemic reforms and target specific weaknesses identified by the diagnostic tests through reading interventions. The USAID MTPDS project used the findings from these assessments to measure progress of learner performance over the duration of project implementation as well as a tool to inform program inputs. As USAID played a key role in MoEST's efforts to improve early grade reading during this period, any gains registered over the three years can be seen to reflect the support provided by USAID in these efforts. Finally, through repeated participation of MoEST representatives in all aspects of planning and implementing the assessments, the process aims to build significant capacity for continuing the practice of early grade reading assessments in the future, led by MoEST.

All three assessments—2010, 2011, and 2012—were conducted at the beginning of the school year for Standards (grades) 2 and 4, thus measuring what students had learned after a full one and three years of learning, respectively. In 2010, the baseline consisted of a broad national sample allowing generalization across the population, but it was not large enough for comparison of specific sub-groups. In 2011, the sample was increased in order to maximize the precision for reporting on differences among subgroups, such as grade, gender, and division. In 2012, the sample was drawn using the complete list of schools from the 2011 education management information system (EMIS) database, which is the most recent available. It was drawn using a random number generator. The sample was stratified by division and by district in proportion to population as in baseline. Where a single sex school was selected, a matching school of the opposite sex was selected from the same

district. The data were collected from October 30 to November 23, 2012, in all divisions of Malawi.

Findings

According to the sample data from 2012 (using weighted means), it is clear that Malawian students are not learning how to read in the first year of primary school and are learning only very little in the two years that follow. The average student beginning Standard 4 is reading only 15 words per minute of a short story, with 13% comprehension (See Table 9 for complete details). A child in Standard 2 is only beginning to learn letter names and syllable reading, but 90% couldn't read a single word of a short story. At present, none of the Standard 1 benchmarks for 2012 have been met by students starting Standard 2. Even students starting Standard 4 are only just meeting the Standard 1 benchmarks in most areas.

For a large proportion of students the assessment was discontinued due to their inability to answer any questions correctly, therefore the overall means are reduced significantly because of these zero scores. When we only consider students who actually demonstrated some measurable reading skills (score greater than zero), the means double and even triple in some cases for Standard 2 students, and increase for Standard 4. Yet the scores still remain below the grade-level benchmarks. The average inter-grade gain in oral reading fluency is approximately 7 correct words per minute.

Comparison between the 2010 baseline and 2012 endline scores show slight improvement among Standard 4 students sampled, but very little progress among Standard 2 students (see Table 13 for details). At both baseline and endline, zero scores represent the largest proportion of Standard 2 and Standard 4 student scores. The subtests that showed a change with a moderate or strong effect size between 2010 and 2012 were Standard 2 letter naming (0.31), Standard 4 syllable reading (0.34), Standard 4 nonsense word reading (0.33), and Standard 4 reading comprehension (0.33). In 2012, scores on most subtests increased, but the practical significance (the real value of the change) is small—though not as small as in 2011. In other words, a change from 11 to 17 familiar words per minute is an increase, but 17 words read in 60 seconds is still, practically, very low for a Standard 4 learner.

For the phonological awareness subtests—syllable segmentation and initial sound identification—the effect size is small to medium and negative, showing a performance decrease (from baseline—there was a slight increase between midterm and endline). The largest effect size, however, was negative, for the initial sound identification exercise, in which Standard 4 performance actually decreased. The reasons for this are described later in this report.

There is no significant difference between the performance of girls and boys. The range of scores across districts is still very low in Standard 2—fewer than 10 correct items per minute. The Shire Highlands region stands out as having the highest score on the letter naming task, but is similar to the national average on all other subtasks. On the other hand, the Northern region outperforms the national average on all tasks, albeit not by much. Shire Highlands, Central Western, and Northern regions had the highest performance and the most significant gains in performance since baseline, whereas the Southern Eastern region had the lowest performance and no gains since baseline.

Descriptive data from the sample suggest that the average Standard 2 child is close to 9 years old, living in a poor household, and has no access to reading materials at home; however, the child probably went to nursery or kindergarten and speaks the same language

at home and school. The profile is similar for a Standard 4 child from our sample, except that he or she is older and likely to have missed less school than the Standard 2 child, and more likely to have a school textbook. Classrooms are extremely overcrowded, with many kids above or below the normal age range. More Standard 2 teachers, according to this data, report being trained than Standard 4 teachers. The PCAR materials from MIE remain the most commonly used teaching materials, which nearly all teachers had, regardless of Standard.

Conclusions

It is encouraging that overall endline results for Standard 4 are higher than at baseline and midterm, and it may be the case that national-level school reform efforts, coupled with a variety of smaller initiatives and projects to promote reading, including USAID's MTPDS, are having an effect. However, the reason these efforts are only affecting students in the higher grade deserves attention. We should recall that the Standard 4 cohort assessed at endline was the same cohort (though not necessarily the same students) assessed at baseline. Therefore it seems that efforts to improve school quality are accumulating over the years, and thus having an effect at the higher grades. However, the teaching of basic skills in the first year of primary school remains insufficient and even with these measurable gains, the end result is that students are not able to read with comprehension.

Regression analyses uncovered very few factors associated with higher scores except for grade level, class size, and repeating. For every additional grade that the student is in, we expect to see an increase in oral reading fluency (ORF) of 6.9 words, and a child who is currently repeating a grade read, on average, 3.1 words per minute slower than their classmates who are in the current grade for the first time. Larger class sizes have a strong negative correlation with a child's reading ability. Each 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. Looking only at a subset of the sample—students who read at a rate of 40 words per minute or more—we see that having the school textbook means a child is twice as likely to be in this group. A few school and teacher-level factors were found to be significant, but are harder to explain:

- If the teacher sounds out unfamiliar words two or more days per week: 5.2 words per minute advantage;
- If students read aloud to the teacher two or more days per week: -13.9 words per minute disadvantage; and
- If the PTA meets every 2-3 months or more often: 4.0 words per minute advantage.

A separate randomized control study carried out under USAID's MTPDS aimed to show how much of an effect could be made on reading scores when a systematic focus on early literacy skills is introduced into Standard 1. This study included a small sample of schools in Salima and Ntchisi who received the *Maziko a Kuwerenga* (Foundations of Reading), reading program, and compared results after one year to control schools that did not have the intervention. The results showed that while overall performance remains low, large absolute and relative gains in reading performance were achieved in the intervention schools that did not occur in comparison schools. Students in control schools demonstrated few, if any, measurable pre-reading skills, including naming the letters of the alphabet. Although scores for students in the intervention schools are still very low, they are close to

the level of achievement demonstrated by Standard 4 learners described in the present national study report.

From this we can conclude that by addressing fundamental reading skills such as phonemic 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 from this national study and the smaller intervention study confirms that students who succeed on basic skills such as letter naming and familiar word reading have higher ORF and comprehension scores. Therefore it is important to focus on these basic skills early in Standard 1. Although carefully scripted lessons, teacher training, coaching, and providing text books are having a positive effect, and the gains are large relative to control schools, which showed no reading skills at baseline or endline, they are not large enough to classify students as “readers” or allow them to read to learn. Additional efforts need to be made to improve classroom and home environments to promote reading and multiply the effects of the school reading programs in order to meet established benchmarks. We also need more information on the extent to which the teachers in the schools were actually using the inputs provided to know whether the gains measured are a minimum or maximum to be expected.

This study shows improvement from 2010 to 2012, so it is likely that as the focus on reading becomes more established and teachers begin integrating new teaching strategies more systematically, these gains will increase.

Introduction

Background

The United States Agency for International Development (USAID MTPDS project is a three-year) activity, supporting the professional development of teachers in Malawi and implementation of the National Primary

This report presents the results from a nationally representative study of primary school learners' early grade reading skills in the Chichewa language, undertaken near the beginning of the school year, in November 2012, as part of the USAID Malawi Teacher Professional Development Support (MTPDS) project. It is the third in a series of national samples designed to identify overall skills and gaps in order to inform national-level policies and strategies. This section provides some background on the project, as well as the methodology (sampling, instruments, etc.) used to collect data.

Curriculum (NPC) with the goal of improving early grade reading and performance of learners. This report presents the results of the final 2012 MTPDS Early Grade Reading Assessment (EGRA) survey, conducted near the beginning of the school year, in November 2012. It is the third in a series of annual, national diagnostic tests supported by USAID and administered in collaboration with the Malawi Ministry of Education, Science and Technology. The reading skills tested are: initial sound identification, syllable segmentation, letter naming, syllable reading, familiar word reading, nonsense word reading, oral reading fluency, reading comprehension, and listening comprehension. Each of these subtests measures important component skills in early reading that are predictive of later performance in literacy.

The national sample was randomly selected to be representative of the entire country, and thus generalizable to the larger population, and as such, potentially reflecting changes in system-wide educational policies and practices. While it is important to recognize that it can take many years to change educational achievement in a measurable way on a national scale, some important developments have taken place between 2010 and 2012 that can be expected to positively influence general reading outcomes if implemented and sustained in a meaningful way:

- Between the 2010 and 2012 EGRA assessments, USAID's MTPDS project in partnership with the MoEST implemented three two-day continuous professional development (CPD) trainings in literacy for all teachers in Standards 1–4.
 - Literacy Module 1 – January 2011
 - Literacy Module 2 – May 2012
 - Literacy Module 3 – August 2012
- The Tikwere Interactive Radio Instruction (IRI) program continued to broadcast daily radio programs with literacy content based on the new Primary Curriculum books into classrooms nationwide.
- On October 27, 2011, the MoEST and other education stakeholders were convened as a body called the National EGRA Coordination Committee. They agreed on an initial set of literacy benchmarks and targets to serve as key indicators of progress in developing the quality of education services (See Annex 4). These indicators are expected to assist the Ministry in tracking literacy achievement and to provide a way for teachers and parents to chart progress against approved standards.
- The USAID Read Malawi program continued, which was geared towards improving literacy level of learners in selected schools across the country. Under this program a

variety of supplementary readers with stories targeting Standards 1 to 4 have been produced and distributed to over 900 schools. The stories were written in English and Chichewa by select teachers and college lecturers.

- There have been small scale interventions by Save the Children and World Vision International using Literacy Boost and community participation to improve reading outcomes. They also supported a National Literacy Conference in October 2012 which was attended by a diverse group of professional from private and public sector involved in implementing literacy programmes in Malawi.
- Roll out of school improvement grants to 12 districts under the Primary School Improvement Plan (PSIP). These programmes were created to empower the school management committees to assist the schools in coming up with activities too support education.
- Continued influx of additional student teachers into school under the ODL program.

In addition to these broader reforms and activities at the national level, the USAID MTPDS program implemented an early grade reading program called *Maziko a Kuwerenga* in selected districts. The programme was designed in collaboration with MoEST aiming to improve Chichewa reading by targeting the “big five” reading competencies; phonological awareness, alphabetic principal, fluency, vocabulary, and comprehension. The programme began in Salima and Ntchisi districts in September 2011 reaching 268 schools and approximately 46,000 Standard 1 learners. The *Maziko a Kuwerenga* programme consists of:

- Supply of the *Nditha Kuwerenga* (“I can read”) reader for each child in Standard 1; and structured lesson plans for literacy teaching to all Standard 1 teachers;
- Eight days of CPD training on early reading skills and on the use of the materials provided;
- School- based coaching and support to teachers in implementing methods; and
- Community mobilization in support of literacy and monitoring teaching and learning in their schools.

The program was subsequently expanded to five more education districts (Mzimba North, Ntcheu, Zomba Rural, Blantyre Rural and Thyolo) in March 2012 and now reaches 1,042 schools and over 160,000 Standard 1 learners.

While these were all positive developments, it is also important to note that during the study period, availability of textbooks may have actually declined, as there was no supply during this period while the MoEST awaited funding for the printing of new revised lower primary Chichewa textbooks.

Purpose and Research Design

This study followed an initial baseline study, conducted in November 2010, and a midterm assessment conducted in November 2011, using equivalent Malawi EGRA instruments. The series of national surveys aims to inform MoEST, development partners, and other relevant stakeholders of the current status of learner performance in early grade reading in Malawi and to contribute to policy dialogue as well as general public awareness and support for improving education. The results are also expected to help the Ministry and other stakeholders determine the impact of major systemic reforms and target specific weaknesses identified by the diagnostic tests through reading interventions. The MTPDS

project also used the findings from these assessments as a tool to inform program inputs, but the results cannot be taken to be a measurement of impact of the MTPDS program. There are far too many other concurrent factors to claim that the inputs of this one program have had an impact at the national level in only three years (even less considering the time taken for implementation at national scale). Finally, through repeated participation of MoEST representatives in all aspects of planning and implementing the assessments, the process aims to build significant capacity for continuing the practice of early grade reading assessments in the future, led by MoEST.

In addition to these three national surveys, MTPDS carried out separate baseline and follow-up assessments which were specifically designed to measure the effects of the USAID MPTDS reading intervention *Maziko a Kuwerenga* in Salima and Ntchisi compared to the control districts of Dedza and Mwanza. The results of these intervention studies are available in a separate report, with highlights only presented in the conclusions of this report. **Table 1** provides an overview of the three-year research design.

Table 1: Summary of Three-Year Research Design Outputs

Date	Study details	Study write-up
National study baseline : 2010	<ul style="list-style-type: none"> 2460 students from 99 schools in 6 divisions (34 districts) Beginning of Standards: 2, 4 November 2010 	“Malawi Early Grade Reading Assessment: National Baseline Report,” March 2011. Project report submitted to USAID. Prepared by J. Mejia.
Intervention study 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 document only.
National study mid-term: 2011	<ul style="list-style-type: none"> 3,019 students from 150 schools in 34 districts Beginning of Standards: 2, 4 November 14 to December 1, 2011 	“Malawi National Early Grade Reading Midterm Assessment 2011,” July 2012. Project report submitted to USAID. Prepared by E. Miksic and S. Harvey.
Intervention study midterm	<ul style="list-style-type: none"> Intervention: 210 learners from 20 schools in Salima and Ntchisi End of Standard 1 May 2012 	“Early Grade Reading Assessment (EGRA): Snapshot of Progress in Learner Achievement in Salima and Ntchisi Districts – May 2012” (Unpublished report prepared by E. Miksic)
National study final: 2012	<ul style="list-style-type: none"> 5240 students in 202 schools in 34 districts Beginning of standards: 2 and 4 October 30 – November 23, 2012 	Present report: “Malawi National Reading Assessment Survey: Final Assessment – November 2012,” April 2013, MTPDS. Project report submitted to USAID. Prepared by S. Pouezevara.

Date	Study details	Study write-up
Intervention study final: 2012	<ul style="list-style-type: none"> • Intervention: 1332 learners from 33 schools in Salima and Ntchisi • Control: 480 learners from 16 schools in Dedza and Mwanza • Beginning of Standard 2 • November 2012 	<p>“Malawi Reading Intervention: Early Grade Reading Assessment (EGRA) Final Assessment – November 2012”. April 2013, MTPDS Project report submitted to USAID. Prepared by S. Pouezevara.</p>

Methodology

Instruments

The EGRA is administered orally to individual learners. It takes approximately 15 minutes to administer per learner, and in this study, it was combined with a questionnaire measuring a variety of learner background variables aimed at identifying factors that are consistently correlated with performance. Five of the subtests are timed and hence can be said to have a fluency component. These are letter naming, syllable reading, familiar word reading, nonsense word reading, and oral reading fluency. The results on these subtests are expressed as a measure of “correct items read per minute.” On the other hand, four of the tests are not timed and therefore do not have a fluency component. These are syllable segmentation, initial sound identification, reading comprehension, and listening comprehension. The results for untimed components are expressed as a “percentage of correct items out of total possible.”

The EGRA instrument for Malawi’s baseline, midterm, and final assessments was originally developed for the 2010 baseline assessment during workshops with the MoEST and other reading experts (for details, see the Malawi 2010 EGRA baseline report [MTPDS, March 2011]). The instruments, including several different versions of reading passages, were piloted and then analyzed using *Rasch analysis*. From one assessment to another, only the listening and reading comprehension passages were changed significantly; otherwise, items remained the same but the order was randomized to retain equivalency in terms of difficulty, but minimize the possibility that students memorized the order of items in advance of data collection based on availability of previous versions of the assessment. Where new subtests were created (oral reading fluency, reading comprehension, and listening comprehension), the test items from all three years were administered to a sample of 240 students, split evenly between Standards 2 and 4 for endline equating, and 60 students split evenly for midterm equating. A method of statistical analysis called *Means Equating* was used to assign a value to each passage based on its observed difficulty in comparison to the baseline. The factors for the two follow up studies were as follows. Numbers less than 1 indicate it is easier than baseline.

- **Midterm Oral Reading** = .934, Reading Comprehension = .676, Listening Comprehension = .912
- **Endline Standard 2:** Oral Reading = 1.049, Reading Comprehension = .505, Listening Comprehension = 1.184
- **Endline Standard 4:** Oral Reading = 1.015, Reading Comprehension = .603, Listening Comprehension = 1.008

This means that all relevant scores from each of those time periods were multiplied by the corresponding value to form a new score—an *equated score*. All scores reported here are equated scores unless otherwise mentioned.

For the first time in Malawi, the 2012 national survey used an electronic data collection approach for EGRA using the Tangerine™ software developed by RTI International. Tangerine was piloted among a sample of schools receiving the USAID MTPDS reading intervention in May 2012 in Salima and Ntchisi, which allowed the project to be confident in its feasibility and convenience for a national-scale data collection.

In addition to the reading assessment subtests, each child was asked a series of questions pertaining to the home and school environment. Additionally, teachers of Standards 2 and 4 who were present on the day of data collection, as well as the head teacher, were interviewed at each school. A total of 375 teachers and 148 head teachers were interviewed (at all schools, combined). These interviews consisted of questions about training, experience, and reading instruction. Interviews were used as part of the analysis of predictive factors of reading performance.

Cronbach’s Alpha is used to determine reliability of the instrument and how the different subtests work together to measure the underlying construct—in this case, early reading skills. The midterm assessment report noted that the instrument had an overall reliability score of 0.87, an acceptable level for this type of instrument. Since the reading and comprehension passages were changed, an updated analysis was done resulting in the figures provided in **Table 2** below. Individual correlations and alphas changed very little, and the overall reliability increased to 0.89, indicating that the changes made to certain subtests have not reduced the instrument’s reliability.

Table 2: Cronbach’s Alpha for EGRA 2012

Item	No. of obs.	Item-Test correlation	Item-Rest correlation	Covariance	Alpha
Letter naming	5224	0.89	0.83	49.23	0.88
Syllable segmentation	5209	0.96	0.93	45.65	0.87
Initial sound identification	5206	0.96	0.95	54.02	0.86
Syllable reading	5202	0.95	0.94	60.79	0.87
Familiar word reading	5199	0.95	0.94	55.20	0.86
Nonsense word reading	5199	0.92	0.91	75.28	0.90
Oral reading fluency	5220	0.44	0.40	74.36	0.90
Reading comprehension	5029	0.51	0.49	74.52	0.90
Listening comprehension	5199	0.96	0.94	55.08	0.86
Overall test				64.36	0.8942

Grade and Language

All three assessments—2010, 2011, and 2012—included students at the beginning of Standards 2 and 4, thus measuring what students have learned after a full one and three years of learning, respectively. In Standard 1, typically, many early foundational skills are

learned, and it is a common expectation that students will be able to read by the end of Standard 2. Standard 3 is a time to develop skills in comprehension. As such, students in Malawi should be expected to read fluently and with comprehension by the end of Standard 3. Given the three-year design of the series of studies, the cohort of students who were tested as Standard 2 students in 2010 were in Standard 4 in 2012; thus the 2012 results reflect cumulative gains for students in the system for three years.¹

According to the curriculum policy and practice in Malawi, teachers are expected to use the learners' local, familiar language as the medium of instruction for all subjects in Standards 1–4. In the majority of schools, this language is Chichewa; however, in certain regions of the country, other home languages predominate. In such places, learners are still expected to learn to read in Chichewa. English is also taught as a separate subject starting in Standard 1. Because the curriculum prescribes that the learning outcomes in reading are to be achieved in Chichewa, the EGRA subtests are using Chichewa items for assessment. However, enumerators were required to translate the instructions for completing the assessment into the learners' home language to ensure that they understand the task.

Sampling

Malawi's six divisions encompass 34 education districts. In 2010, the baseline consisted of a broad national sample allowing generalization across the population, but it was not large enough for comparison of specific sub-groups. In 2011, the sample was increased in order to maximize the precision for reporting on differences among subgroups, such as grade, gender, and division. In 2012, the sample was drawn using the complete list of schools from the 2011 EMIS database, which is the most recent available. It was done using a random number generator. The sample was stratified by division and by district in proportion to population as in baseline (**Table 3**). Where a single sex school was selected a matching school of the opposite sex was selected from the same district.

Table 3: 2012 Study Sample, by Division and Standard

Division	Schools	Learners, Standard 2		Learners, Standard 4		Total learners per division
		Boys	Girls	Boys	Girls	
Central Eastern	59	784	793	289	287	2153
Central Western	46	319	310	185	186	1000
Northern	23	123	117	113	114	467
Shire Highlands	22	106	108	106	104	424
South Eastern	25	125	130	125	124	504
South Western	27	236	237	110	109	692
Total	202	3388		1852		Learner Total: 5240

¹ While keeping in mind that the studies are based on random samples of schools as well as students. No deliberate effort was made to re-test the same actual students, and the chances of this are unlikely; however, the group average from 2012 Standard 4 students can be considered to represent the same cohort of students who were tested in 2010 as Standard 2 students.

For cost and efficiency reasons, it was not possible to draw a simple random sample of learners from across the country, since in most cases doing so would have resulted in the testing of one individual learner per school. Instead, to enable inferences to be made about the performance of the entire nation, not just the learners were sampled, but *sampling weights* were attached to the data. That is, because the learners were grouped within schools, and schools within divisions, the design did not give every individual learner an equal chance of selection, necessitating the use of a statistical process to determine the probability of selection of each person in the sample—the sampling weight. Based on the total population and learners in the sample, a weight was calculated for each level of selection (schools by division, learners by grade and gender). Due to the Reading Intervention concurrently being studied in the treatment districts of Ntchisi & Salima, and the control districts of Dedza & Mwanza, these districts are oversampled in this national study. Therefore, each school from these districts is given a lower weight to account for their higher probability of selection². The analysis was conducted using sampling weights in Stata; the weights increased the power of the estimates based on data from the individual learners to make them representative of the estimated population within each group.

Data Collection

The data were collected from October 30th to November 23rd, 2012 in all education divisions of Malawi. Prior to data collection there was a training of trainers organized by USAID's MTPDS and the participants to this training were drawn from MoEST departments. This was in preparation for training of enumerators by officers from various institutions of MoEST which included DTED, DIAS, Basic Education, Planning, Domasi College of Education, MANEB, and several Teacher Training Colleges (TTCs). MTPDS staff provided support in coordinating the exercise. The supervisors of data collection teams were selected from various relevant MoEST departments and institutions, including those listed above. Primary Education Advisors (PEAs) served as enumerators, assessing individual learners on their Chichewa reading ability and carrying out the associated learner interviews. PEA's did not collect data in their own zones/clusters in order to minimize potential bias. Assessors were trained by MoEST master trainers who were themselves trained during an October 2012 workshop facilitated by a consultant, Paula Gaines, and Odala Banda, MTPDS EGRA Manager.

Limitations

All research has limitations to the reliability and generalizability of the results, and this case is no different, although we have tried to mitigate the effect of any known limitations. First, although the instrument methodology and items did not change from the 2010 baseline to the 2012 endline, the format of administration changed from recording answers on paper to recording answers directly into a digital database using tablet computers. The precise influence of this change is uncertain. However, since 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) it is expected that the only major

² More detailed information is provided in Annex 5.

difference is that the 2012 endline results are more accurate since there was less chance that data entry errors took place when coding data from paper into the database.

Second, different enumerators were used to administer the test from baseline to endline, and although training procedures aim to ensure that all assessors were judging students the same way, there are always some differences in the way different enumerators hear or tolerate reading “errors” from the child. In order to instill the sense of ownership by MoEST of the assessment results and to develop the capacity of MoEST officials, MTPDS shifted from using private enumerators in 2010 to using MoEST staff (especially PEAs) as enumerators for the 2011 and 2012 EGRA assessments. In order to mitigate the potential bias of using MoEST officers, MTPDS put measures in place to avert the possibility of PEAs conducting assessments in schools within their own jurisdiction.

Descriptive Statistics

Learner Background Questionnaire Results

As part of all three national EGRAs, students were also asked questions about their socioeconomic status, family situation, and school background. **Table 4** summarizes learners’ responses to these questions.³ Some questions are used as proxies for socioeconomic status⁴, and others are factors shown in research to correlate with learner learning.

Descriptive data from the sample suggest that the average Standard 2 child is close to 9 years old, living in a poor household, and has no access to reading materials at home; however, the child probably went to nursery or kindergarten and speaks the same language at home and school. The profile is similar for a Standard 4 child from our sample, except that he or she is older, likely to have missed less school than the Standard 2 child, and more likely to have a school textbook. Classrooms are extremely overcrowded, with many kids above or below the normal age range. More Standard 2 teachers, according to this data, report being trained than Standard 4 teachers. The PCAR materials from MIE remain the most commonly used teaching materials, which nearly all teachers had, regardless of Standard.

Table 4: Learner Characteristics (% of sample with this characteristic, or average value)

Item	Standard 2 (n=1180)	Standard 4 (n=1158)
Female (%)	52%	50%
Age (Average value)	8.6	10.8
Has 0-3 items	66%	65%
Has 4-6 items	17%	23%
Has 7-10 items	5%	6%
Uses same language at home and school	83%	84%
Went to nursery/kindergarten	68%	69%
Was absent last year for more than 1 week	27%	19%

³ Not based on weighted data, but actual sample proportions.

⁴ 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	Standard 2 (n=1180)	Standard 4 (n=1158)
Is not repeating a grade this year	89%	90%
Has a school textbook or supplementary reader	9%	34%
Has other books at home	17%	29%
Mother has some post-primary education	39%	42%
Father has some post-primary education	48%	58%

These data provide us with a picture of the environment in which learners in our sample are situated, particularly when combined with information from teacher and head teacher interviews. From the descriptive data provided above, we can see that a Standard 2 child in our sample was probably close to 9 years old, living in a poor household with few luxury items like electricity, a television, or a car, and little access to reading materials at home; however, the child probably went to nursery or kindergarten and spoke the same language at home that was taught in school. Only half of fathers, and less than half of mothers have some post-primary education.⁵ The picture is largely the same for a Standard 4 child, except that he or she was older and likely to have missed less school than the Standard 2 child. Standard 4 children also have a more textbooks than Standard 2 children.

Of particular interest is that the average ages are above what would be expected of a child who started school at the official entry age of 6 years old in Standard 1—7 years old in Standard 2 and 9 years old in Standard 4. These averages are also a product of a wide range of ages. Although we know that the student’s self-reported age is not always entirely accurate, the range of reported ages in the Standard 2 sample was 5 to 15, and in the Standard 4 sample 7 to 16. This would indicate that students are either starting late or repeating grades often; either way, the job for the teacher becomes more complex when dealing with a range of learner types. The question “What grade were you in last year?” identified approximately 10% of students in each group who were repeating Standard 2 or Standard 4 when the assessment was administered.

The results of this exit interview can also be used to correlate performance in reading with factors outside the school environment. Past EGRAs in other countries have shown, for example, that going to pre-school, having books in the home, and speaking the language of instruction in the home are predictors of reading achievement. The ability of the Malawi study to actually show this is limited in this particular sample by the low variability of responses; in other words, since the majority of students have the same characteristics—high poverty, lack of books at home, low education level of both parents—we have less chance of using these characteristics to explain differences in reading ability. Later in the report we will examine how student, teacher and school characteristics are associated with reading scores and other limitations to these correlations.

⁵ Note that the description of parents’ education is from the point of view of the child, so cannot be expected to be fully accurate. Also, 24% of Standard 4 students and 38% of Standard 2 students said they did not know, thus further reducing the sample number.

Teacher Background Questionnaire Results

Another element of the national study was to determine what characteristics of the teacher and classroom might have an effect on reading scores. **Table 5** summarizes teachers' responses to these questions, and so the responses pertain only to the respondents. It is important to note that these questions were asked of teachers at the beginning of the year, whereas the assessment of learning reflects learning that largely took place the year before, and most likely with a different teacher. Therefore the results from this teacher questionnaire are mostly useful for getting an overall picture of the teaching strengths and limitations of teachers in the Malawian system today and addressing those factors that have historically shown in other studies to be strongly correlated with student outcomes, for example: teacher training, limited absences, and teaching behaviors that promote reading skills.

Table 5: Teacher Characteristics (% of sample with this characteristic or average value)

Item	Standard 2 (n=172)	Standard 4 (n=156)
Female (%)	65%	59%
Age (Average value)	38.5	34.2
Teaches more than one grade level	9%	7%
Average class size	122	100
Education level (JCE/MSCE)	28%/72%	23%/87%
Is a trained teacher	96%	52%
Has sufficient learning materials	13%	16%
Has meetings with parents	56%	67%
Teaching practice in reading has been observed by a head teacher	75%	70%
Has received feedback on teaching practice	80%	78%
Has received coaching visits from PEA	82%	83%
Learners sounded out unfamiliar words every day	52% (avg=3.8)	37% (avg=3.5)
Learners read aloud to teacher or others every day	72% (avg=4.5)	48% (avg=3.8)
Learners were assigned reading to do every day	42% (avg=3.3)	32% (avg=3.3)
Never checks homework	9%	7%
Believes a learner should be able to recognize letters and say letter names in Standard 1	79%	82%
Believes a learner should be able to sound out unfamiliar words by the end of Standard 2	91%	67%
Believes a learner should be able to understand stories they read by the end of Standard 2	93%	92%

Most teachers in this sample of Standard 2 and Standard 4 teachers were in their 30s and had an MSCE degree. As a result of pre-service student teacher deployment practices, Standard 2 teachers were more often trained teachers than Standard 4 teachers, who were

still undergoing training (i.e., through the ODL program) or had just started teaching. A “trained teacher” for the purpose of this study is one who has been given a certificate to teach. However, there are also untrained teachers in the schools who are either assistant or volunteer teachers provided by the community to make up for the shortage of teachers. The majority of teachers reported being supported by head teachers and PEAs, but the content, quality, and frequency of this support is not measured through this questionnaire. Some PEA visits are for administrative purposes rather than professional development purposes.

Only about half of teachers indicated using daily practice in basic reading skills, such as those measured by this test, in their classroom. The exceptions are Standard 2 ‘sounding unfamiliar words’ and Standard 2 ‘reading aloud’, where more than half of the teachers reported engaging learners in these tasks every day. In the case of ‘reading aloud’, 72% of teachers say the children read aloud every day, but whether this is reading or recitation is uncertain. Their expectations of performance were largely in line with established benchmarks: most teachers believed students should be able to identify letters by or during Standard 1 and should be able to sound out unfamiliar words in Standard 1 or Standard 2.

The most striking finding from this survey is the high enrollment figures provided by teachers for the number of students in their class. The average class size was 122 in Standard 2 and 100 in Standard 4, but classes ranged from 13 to 300 students. As mentioned above, these large classrooms were filled with students of different ages and abilities, and insufficient learning materials, making it very difficult, if not impossible, to provide adequate support to each child and follow up with parents (only half of teachers reported holding meetings with parents).

Table 6 below provides an overview of the materials that are being used in classrooms. The PCAR materials from MIE remain the most commonly used teaching materials, which nearly all teachers had. Other books were more prevalent in the higher grades, but some teachers reported having the USAID MTPDS or USAID Read Malawi materials as well. About 14% of teachers in Standard 2 and 7% in Standard 4 report using the MTPDS reading materials (*Nditha Kuwerenga* and *Maziko a Kuwerenga*). This reflects the fact that the intervention schools were included in the sample; however, the percentages are disproportional to what we would expect them to be based on sampling. Therefore some schools outside of the intervention schools seem to have access to some of the MTPDS materials. This may be because at the literacy CPD trainings teachers were given modules which express skills that are in *Maziko a Kuwerenga* reading intervention. These books are mentioned as references during trainings and are not necessarily given to teachers for use in teaching in their classes, but they may have interpreted the question as pertaining to their experience with the materials and not whether they actually have and use them in the class.

Table 6: Materials Used in Class

Group	PCAR Chichewa textbook from MIE	PCAR Chichewa teachers’ guide from MIE	Old curriculum textbooks	Other books	Nditha Kuwerenga reader	Maziko a Kuwerenga teacher’s guide	Read Malawi supp. readers
Std 2	98%	97%	66%	7%	14%	13%	17%
Std 4	97%	97%	62%	20%	7%	5%	10%

Head Teacher Background Questionnaire Results

Finally, there were also questions asked of 140 head teachers or deputy head teachers about themselves, or pertaining to conditions of school infrastructure and teacher support. **Table 7** summarizes the responses to these questions. As with all student, teacher and head teacher questionnaires, we have to be cautious of attaching too much meaning to the responses, since they are self-reported and not otherwise verified. A self-report of, for example, “checking teachers’ lesson plans” does not guarantee that the lesson plans were of high quality or that the head teacher read with a critical eye and made suggestions. The head teachers were most often male and had served in that capacity for 7 years on average. A majority reported having training in school management and early grade reading, but the content of these trainings is not captured in this instrument. All answers are pertaining to the respondent (the head teacher or deputy head teacher) or to the school in which they work, and so are applicable to both Standard 2 and Standard 4 students in the sample. Teacher characteristics are provided in Table 5 above.

Table 7: Head Teacher and School Characteristics

Item	% of sample with this characteristic, or average value
Female (%)	30%
Years in position as head teacher	6.83 (Min 2 months, Max 23 years)
Highest education level	JCE: 13% MSCE: 87%
Has received training in school management	86%
Has received training in implementing an early grade reading program	62%
Has supported teachers in the pedagogy of teaching reading	82%
Is satisfied with the performance in reading at the school	35%
Teachers lesson plans are reviewed on a daily basis by the head teacher	81%
Never observes teachers in their classroom	8% (average is 2x per term)
There is water on the premises	74%
The school has electricity	19%
School has a library**	26% (reported by teachers), 31% reported by head teachers

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

The training in early grade reading may include—but is not limited to—the MTPDS-supported CPD training program. Although they reported providing support to teachers (observations, checking lesson plans), they were also largely unsatisfied with reading outcomes in the school. No doubt the conditions evident in schools—few libraries, crowded classrooms—are difficult to overcome even with strategies they may have been given during training programs. **Table 8** provides an indication of how head teachers were able to determine their satisfaction through monitoring learning progression.

Table 8: How Do You Know Your Students Are Progressing?

Group	Classroom observation	Monitor test results	Evaluates orally	Reviews homework	Teacher reports	Other
Head teachers	62%	61%	32%	32%	60%	0%

Classroom observation and monitoring test results or relying on teacher reports were the most common methods of progress monitoring.

All of these characteristics are generally used to find characteristics that correlate with student learning outcomes. In the case of this study, there were very few characteristics strongly correlated with student outcomes. These issues will be discussed in the section below on correlation. Next we provide the results for 2012, and then the results over time, from baseline to endline.

Results by Subtest, EGRA 2012

Overall Results by Subtest

Table 9 summarizes the results of all the subtests of the 2012 assessment, for Standards 2 and 4. The mean scores show the results for the entire population, as either a number of correct words per minute (in the case of the first five subtests), or percent of questions answered correctly (in the case of the last

Malawian students are not learning how to read in the first year of primary school and are learning only very little in the two years that follow. The average student beginning Standard 4 is reading only 15 words per minute of a short story with 13% comprehension. A child in Standard 2 is only beginning to learn letter names and syllable reading, but 90% couldn't read a single word of a short story. At present, none of the Standard 1 benchmarks for 2012 have been met by students starting Standard 2. Even students starting Standard 4 are only just meeting the Standard 1 benchmarks in most areas.

four subtests.). According to the standard rules of EGRA administration, a test will be discontinued if the child is unable to respond correctly to even one of the first 5-10 items; these students are given a "zero score" on that subtest. The proportion of zero scores is presented in the last column of **Table 9**. Giving no response at all is considered not responding correctly; thus it is important to recognize that some students obtain zero scores because they just don't say anything at all, but we cannot know if they are not responding because they *cannot*, or because they are *unwilling* to do so.

Table 9: Results on EGRA Subtests, by Standard

Subtest	Standard/ Grade	Mean	Standard error	Percent zero scores
Letter naming (100 items, correct per minute)	Standard 2	5.7	0.5	49.5%
	Standard 4	26.7	1.3	11.5%
Syllable reading (50 items, correct per minute)	Standard 2	3.2	0.4	70.8%
	Standard 4	26.3	1.8	20.2%
Familiar word reading (50 items, correct per minute)	Standard 2	1.9	0.3	81.4%
	Standard 4	16.9	1.2	28.9%
Nonsense word reading	Standard 2	1.2	0.2	87.7%

Subtest	Standard/ Grade	Mean	Standard error	Percent zero scores
(50 items, correct per minute)	Standard 4	11.4	1.0	34.3%
Oral reading fluency* (59 items, correct per minute)	Standard 2	1.3	0.26	90.2%
	Standard 4	15.4	1.2	39.1%
Syllable segmentation (10 items - % correct)	Standard 2	35.2%	1.4%	43.4%
	Standard 4	58.6%	1.5%	20.3%
Initial sound identification (10 items - % correct)	Standard 2	3.1%	0.3%	86.6%
	Standard 4	6.7%	0.7%	72.9%
Reading comprehension* (5 questions - % correct)	Standard 2	0.9%	0.2%	94.4%
	Standard 4	13.3%	1.0%	50.5%
Listening comprehension* (5 questions - % correct)	Standard 2	32.6%	1.2%	30.7%
	Standard 4	50.0%	1.2%	9.3%

* Before equating

According to the data in Table 9, the average student beginning Standard 4 is reading only 15 words per minute of a short story, with 13% comprehension (Oral reading fluency and Reading comprehension measurements). A child in Standard 2 is only beginning to learn letter names and syllable reading, but 90% couldn't read a single word of a short story.

Note that for syllable segmentation and initial sound identification, the percentage is based on 10 items; therefore one correct item out of 10 would result in a score of 10%. Thus the average for initial sound identification is less than 1 correct word out of 10 for both Standards 2 and 4. This reflects a high percentage of zero scores, as indicated in the last column (86.6% and 72.9% for Standard 2 and 4, respectively, on this subtest). See more detailed discussion of these phonological awareness subtests below.

The data from **Table 9** are presented visually in **Figures 1** and **2** below. The graphs show that students clearly do improve their reading skills from Standard 2 to 4, and by quite a large margin on all subtests, but especially on letter naming and syllable reading tasks. However, the final score remains very low for learners who have finished three years of primary schooling, and far too low for them to be considered "readers." The markers above each graph show the target scores (benchmarks) for Standards 1 and 3.

Figure 1: Scores on Measures of Fluency

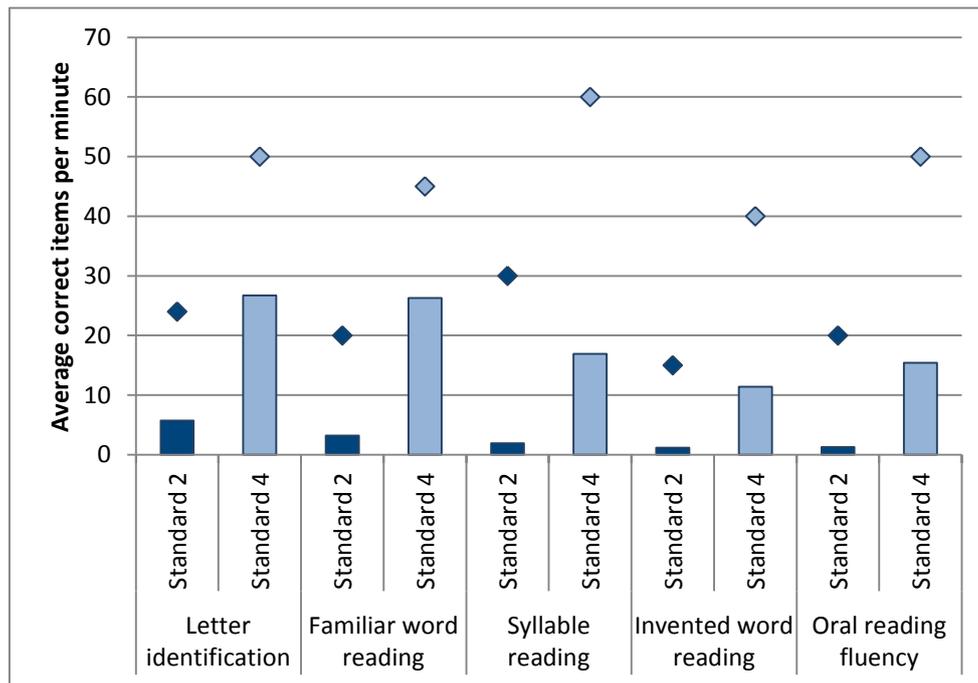
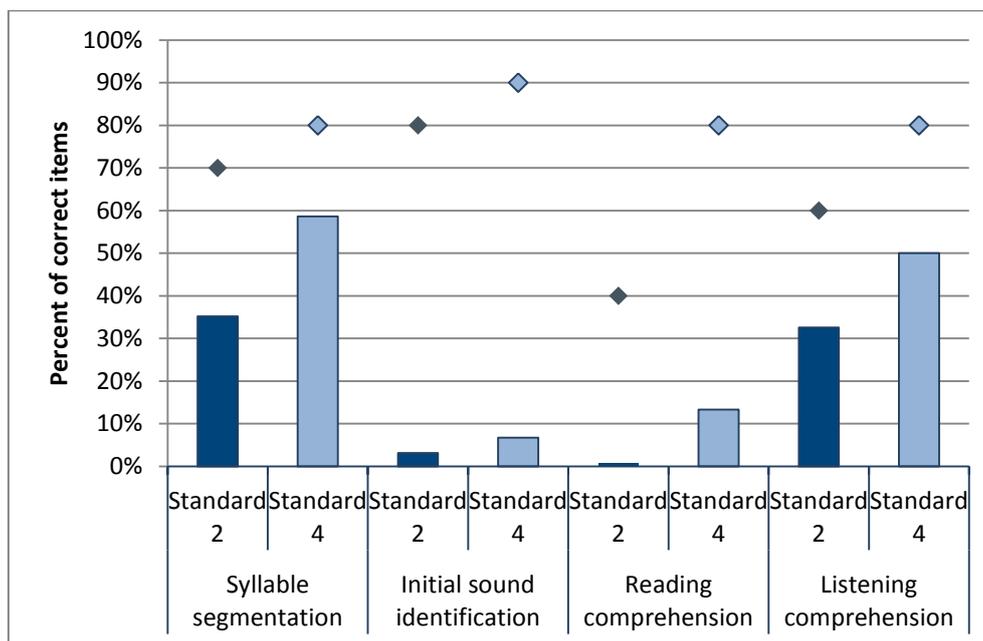


Figure 2: Scores on Other Measures

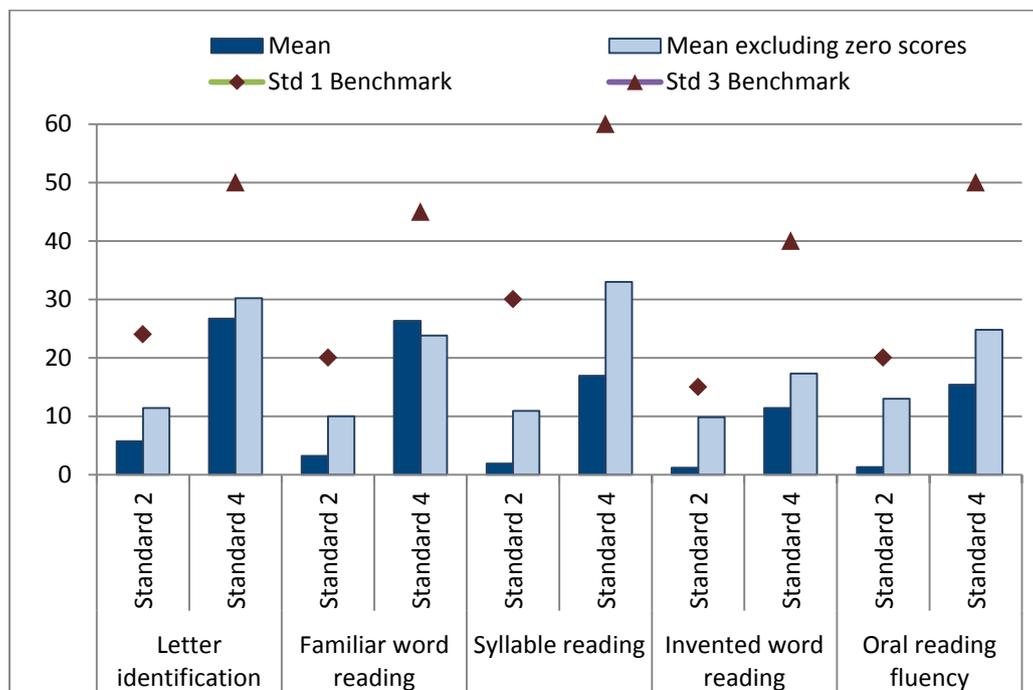


Benchmarks and targets were established by the MoEST in 2011 for performance from 2012 to 2017 for all of the subtests measured.⁶ These benchmarks can be found in Annex 4. At present, none of the Standard 1 benchmarks for 2012 have been met by students starting Standard 2. Even students starting Standard 4 were only just meeting the Standard 1 benchmarks in letter identification and familiar word reading, but not at all in other subtests.

⁶Benchmarks and targets agreed upon at the MoEST-convended EGRA Coordinating Committee Meeting, Crossroads Hotel, Lilongwe, October 27, 2011.

The above scores provide means including zero scores. Zero scores can be interpreted either as students who could not read (they attempted, but got all answers wrong), or were unwilling to read (they did not even attempt to read). Because of the high proportion of zero scores, the means under-represent the ability of students who actually attempted enough of the test to have some demonstrated reading ability. **Figure 3** below compares the means on fluency subtests to the mean when zero scores are removed.

Figure 3: Comparison of Means with Zero Scores Removed



When we only consider students who actually demonstrated some measurable reading skills (score greater than zero), the means double and even triple in some cases for Standard 2 students, and increase for Standard 4. Yet the scores still remain well below the grade-level benchmarks.

The section that follows shows how these overall scores compare to baseline. First, we will take a closer look at a few select subtests: phonological awareness (measured by both initial sound identification and syllable reading), ORF, and comprehension.

Phonological Awareness Subtests

Phonological awareness is the awareness that words are made up of sounds. Learners who have acquired phonological awareness can hear, identify, and manipulate sounds in words (for example, knowing how to replace the “c” in cat with a “b” to make a new word). This awareness is essential understanding that words can be separated into sounds and subsequently, to the ability to identify individual sounds in words. Mastering these skills leads to improved ability to decode or sound out words. In Malawi, EGRA assessed two different phonological awareness skills.

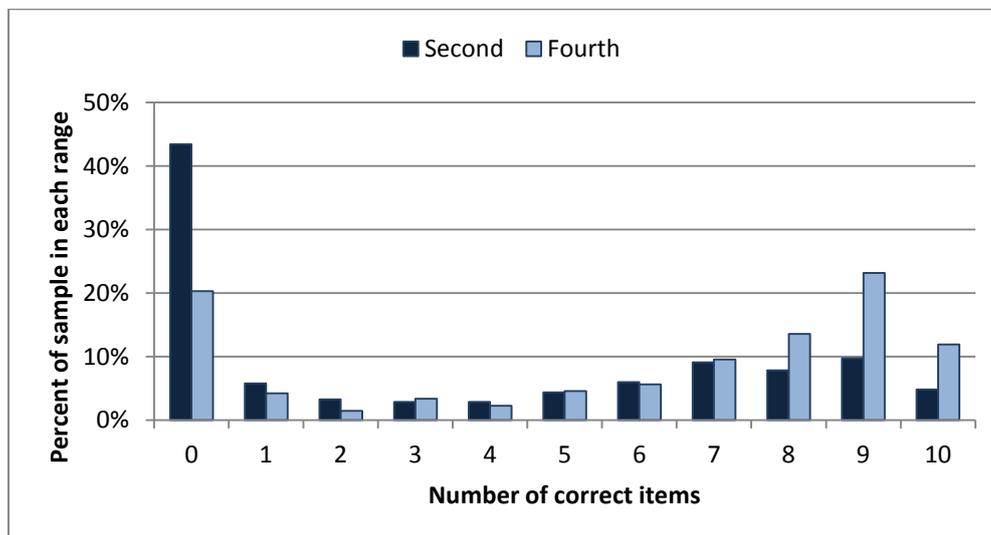
- Syllable Segmentation:** The first of two measures of phonological awareness. The subtest measures learners’ ability to hear a word and break it up into syllables. This is one of the first skills needed to understand how to decode or read new words. The child is asked to say the syllables in the word, not count them.

- **Initial Sound Identification:** The second measure of phonological awareness. This subtest measures learners’ ability to hear and isolate the first sound in a word.

Because these are important pre-reading skills, and because the scores were so different (high for syllable segmentation and low for initial sound identification), it is worth looking in a bit more detail at the results. The **initial sound identification** scores are especially of concern, because there has been no improvement at all from baseline to endline and scores are equally low for Standard 2 and Standard 4. This may indicate that there is inconsistency in test administration, that students don’t understand what is being asked of them because it is not a skill that is explicitly taught, or that this is a skill that students are actually not acquiring.

The **syllable segmentation** task assesses learners’ ability to hear a word and separate it into syllables. This skill is one of the most basic phonological awareness skills. In Chichewa this skill is especially important because it is an agglutinative language⁷ that requires that grammatical markers such as gender or verb tense be added on to words. Being able to separate these markers and find the root word is important to reading fluency. Many languages such as Chichewa are taught with a method that focuses on learning syllables instead of individual letter sounds. This makes syllable segmentation a skill that is more easily acquired. **Figure 4** below shows that although zero scores still represent the highest proportion of learners, of the students who demonstrate some ability, most are grouped at the successful end of the spectrum—8 to 10 items out of 10 correct. Furthermore, students showed positive evolution in their ability from Standard 2 to Standard 4 on this measure. This indicates that learners are gaining some understanding that words can be broken into smaller parts. However, by Standard 4 we would expect learners to have mastered this skill, thus some more explicit instruction could be useful.

Figure 4: Scores on Syllable Segmentation Task

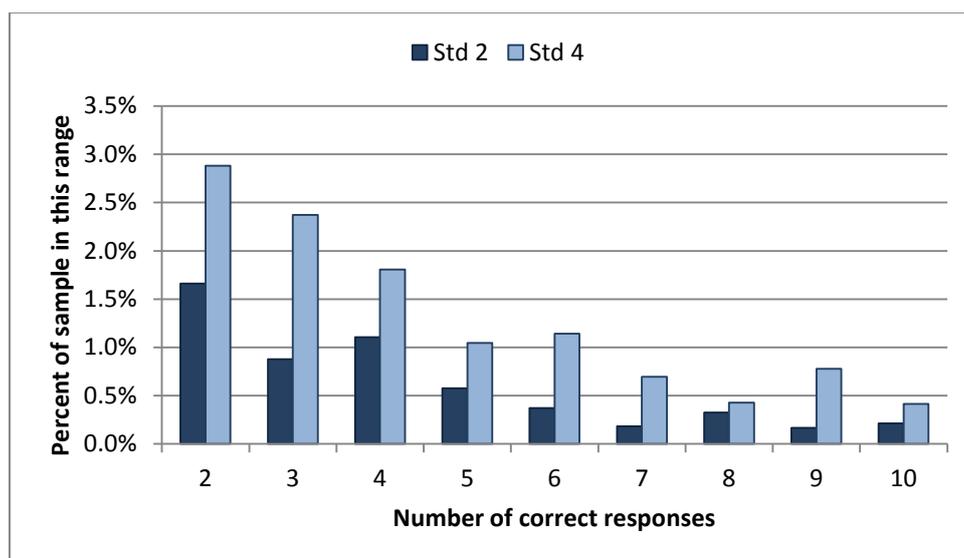


The **initial sound identification** task is a more difficult and advanced phonological awareness skill. Learners often gain an understanding of syllable segmentation quite easily, however

⁷ An agglutinative language is one where words are formed by combining morphemes (small units of meaning) to change a word’s meaning to signify tense, possession, plurality, etc.

hearing and identifying individual sounds often requires specific instruction. The students were asked to repeat the first sound they heard in the following words: kala, dona, khala, atate, bala, mana, gada, wada, nola, gwada. So, for example, an assessor asks the child, “Tell me the first sound you hear in the word ‘kala,’” and the child should respond /k/. This skill is an important first step in being able to hear and separate all the individual sounds in words, which helps learners to be able to understand that words are made up of individual sounds, supporting ability to decode words and ability to spell words correctly when writing. Because this is a skill that generally needs to be explicitly taught, practiced learners who do not receive this instruction tend to have difficulty on this assessment. Most students only answered one item correctly. It is important to note that this task followed the syllable segmentation task, and since all words were made up of two syllables, it may be that students were responding with syllable sounds instead of phonemes (i.e., “ka” in the case of kala, when the correct answer should have been /k/). **Figure 5** below shows the distribution of scores (total number of initial sounds identified correctly) for the Standard 2 and 4 sample combined, and the percent of the total sample represented by each category. (This graph does not include the students who scored zero, otherwise it would skew the scale too much, or the students who got only one correct answer, which could be the result of chance.)

Figure 5: Distribution of Initial Sound Scores



It indicates that learners in both Standard 2 and 4 performed very low, and there is no improvement across grades. Generally we would expect to see improvement from Standard 2 to Standard 4; however, it is likely that learners had no instruction in this specific skill and therefore were unfamiliar with what the task was asking as well as lacking in higher level phonological awareness skills.

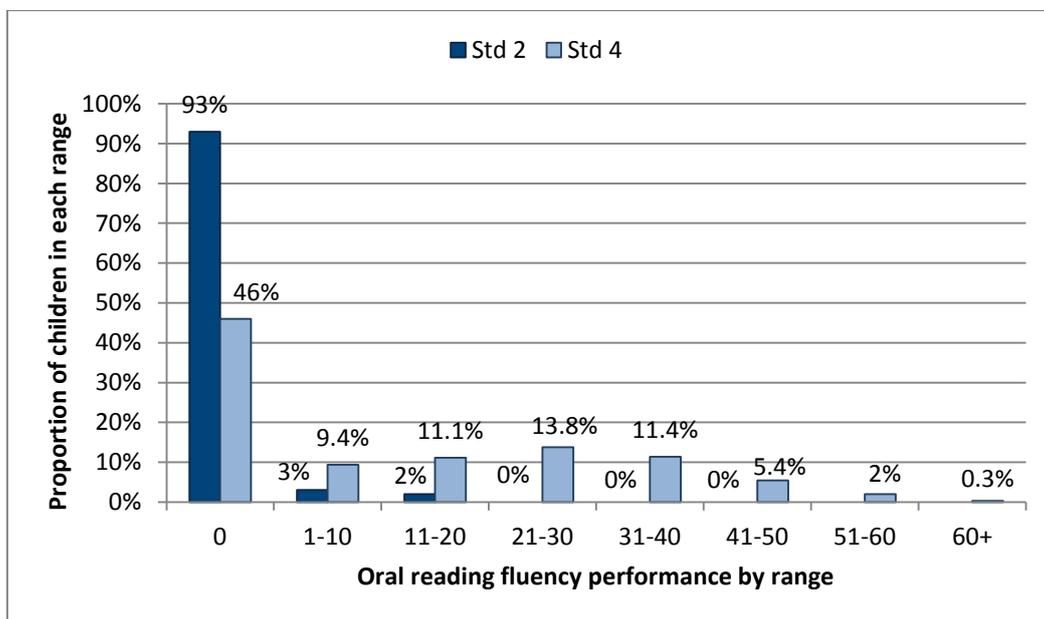
The difference in results from syllable segmentation and the initial sounds identification measures is telling of what type of instruction is likely to be taking place in the classroom. Teachers are likely to teach Chichewa with a focus on syllables. During initial adaptation of the EGRA instrument it was made clear that vowels were taught to learners as individual sounds, while consonants were not taught as individual sounds but as part of syllables. This focus on syllables instead of individual sounds is likely why we see very low scores and no improvement on the initial sound identification task. Since this skill is one of the lower level phonological awareness skills necessary that supports word reading and writing, it may be

the case that learners were not proficient enough in the higher level skills such as sound segmentation—which may be part of the reason the word reading scores are so low.

Oral Reading Fluency Scores

Measuring reading fluency is one way to capture, in one simple indicator, a general measure of early learning skills, because the rate of fluency reflects whether or not students have mastered *automaticity* in the underlying skills—letter sound correspondences, syllable reading, decoding, familiar word reading. Fluency has also been shown to be linked to comprehension; students who read slowly, with hesitation, will not register the meaning of entire phrases due to limited short-term memory capacity. **Figure 6** shows ORF scores for Standards 2 and 4, by range of fluency. Unlike the means provided above, this shows the distribution of scores across abilities.

Figure 6: Endline ORF Scores by Range



The **inter-grade gain** in ORF may be estimated by taking the difference between average ORF in Standards 2 and 4, and dividing the difference by 2. This gives us an inter-grade gain of 7.0. This is also verified through the use of regression techniques, which show a statistically significant difference in scores by Standard, resulting in a 7 word advantage per Standard (see Student Factors, for more details). This is lower than what is typically found—10 to 15 words per minute—in other EGRA studies in similar contexts.

Table 10 below shows the mean and standard error when zero scores are removed. This makes an important difference, particularly in the case of Standard 2 students. Using the mean of the total sample gives the impression that students are not learning at all; however, looking only at students who do demonstrate some reading ability (the rows “less zero scores”, we see that students starting Standard 2 are more than halfway to the Standard 1 benchmark of 20 correct words per minute. Note, however, that it is a very small proportion of the students.

Table 10: Endline Scores Excluding Zero Scores on Oral Reading Fluency

Group	Sample	Mean	Standard error	95% Confidence Interval
-------	--------	------	----------------	-------------------------

		N =			Upper	Lower
Total sample	Std 2	3360	1.3	0.3	0.8	1.8
	Std 4	1839	15.4	1.2	13	17.7
Less zero scores	Std 2	586	13.2	1.1	11	15.3
	Std 4	1305	25.2	0	23.3	27.1

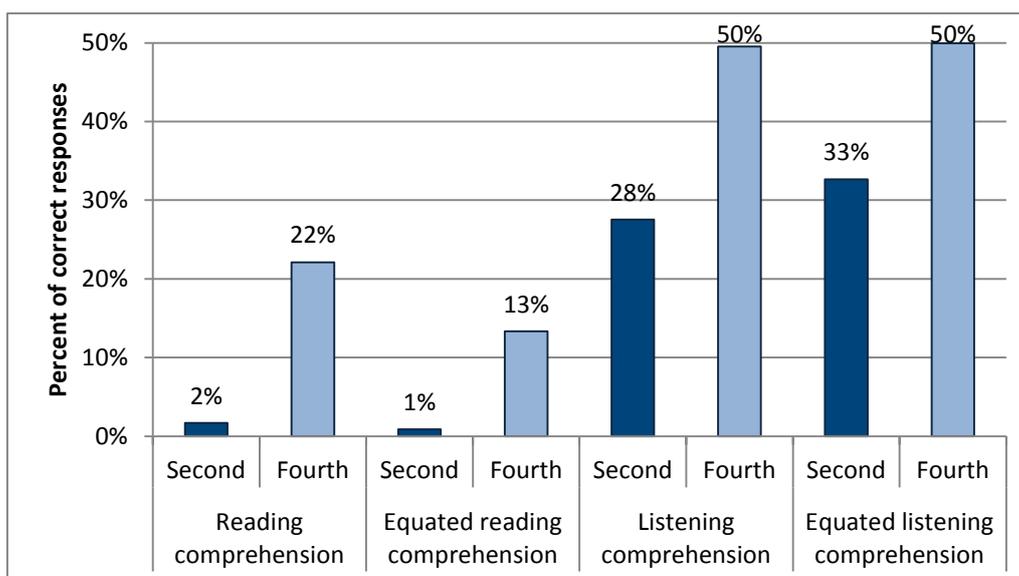
Reading and Listening Comprehension

The ultimate goal of reading is comprehension, or making meaning from text. The EGRA listening comprehension task is designed to understand students' ability to comprehend orally without having to overcome issues of decoding. Part of comprehension is understanding the meanings of the words, thus if students do not have enough vocabulary in the language, they will not be able to comprehend. The reading comprehension task is designed to determine whether students can answer questions about a short story they just read. Numerous studies have shown that there is a strong relationship between reading fluency and comprehension across languages. The ability to make meaning out of a string of words is related to knowing the meaning of each individual word, and also the ability to decode the words fast enough to process them before the limits of short-term memory are reached.⁸ This fluency threshold is usually situated between 40 and 60 words per minute for a beginning reader; adults and fluent readers are expected to have a fluency rate above 150 words per minute, thus the threshold is necessary, but not sufficient for comprehension.

Figure 7 below shows average reading and listening comprehension scores for both standards. Both equated and unequated scores are presented. The equated scores are slightly lower, reflecting the fact that the reading passage at endline was easier than the passage used at baseline, so the value of a correct response is adjusted downward. Both reading and comprehension scores are derived out of the number of questions answered correctly out of a total of 5 questions. The actual number of questions answered varies; for Standard 2 students, the average number of reading comprehension questions answered was less than one, and almost three questions for listening comprehension. In Standard 4, students answered on average one question for reading and 2.5 questions for listening.

⁸ Abadzi and Crouch

Figure 7: Reading and Listening Comprehension Scores by Standard



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 child read in the text. In other words, a child will not be asked all 5 questions if he or she did not read the entire passage in the minute allotted for the task—only 27 students in this sample of Standard 2 and 4 students were asked all 5 questions. Thus the mean percentages shown in the tables and figures above may under represent actual capacity of the students because they show the total correct out of the total possible of 5. **Table 11**, on the other hand, shows the number of students who attempted 1, 2, 3, 4, or 5 questions, and the percent correct out of the total number attempted. 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). This table includes both Standard 2 and Standard 4 readers and shows that as students were asked more questions (because they have read more text and are more fluent), the tendency to answer all of them correctly also increased.

Table 11: Distribution of Comprehension Scores (Standard 2 and Standard 4)

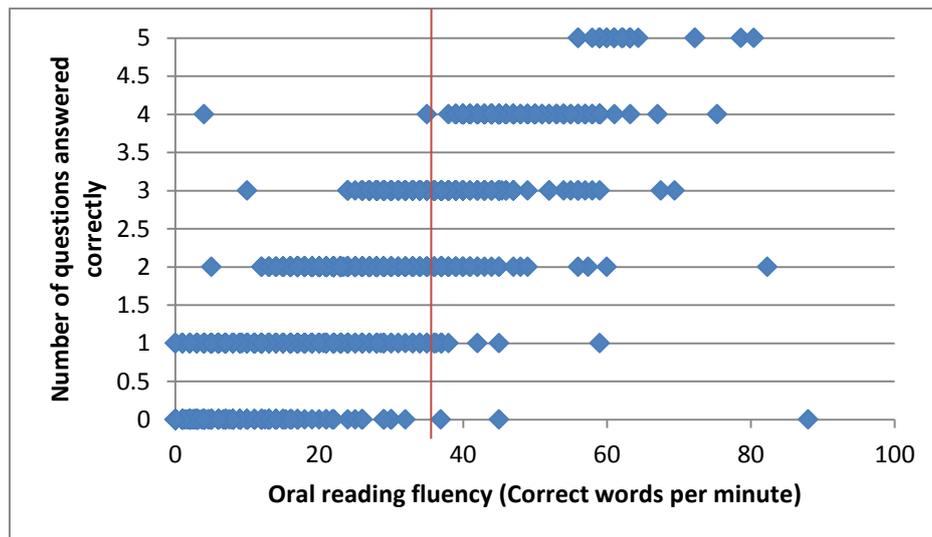
Q's asked	N	Mean ORF	Percent of questions answered correctly											
			0	20	25	33	40	50	60	67	75	80	100	
1	986	4	73%										27%	
2	512	20	8%	28%						64%				
3	516	32	1%	7%			29%				63%			
4	264	44	0%	2%		8%			23%			67%		
5	50	62	0%	2%		4%			8%			30%		56%

* Based on unweighted scores

Thus, whether you consider the measure as number of questions correct out of questions attempted (asked) or number of questions correct out of total questions possible (even if not asked), one thing remains true—higher reading fluency is associated with a higher probability of answering all questions correctly. **Table 11** above and **Figure 8** below both

show the distribution of comprehension scores. **Figure 8** shows the number of questions answered correctly out of the total of 5 possible, in relation to oral reading fluency as an outcome. The trend clearly favors higher comprehension scores for students with more fluency. The students who achieved 80% comprehension or better were largely reading close to 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 comprehension.

Figure 8: Fluency vs. Comprehension (Standards 2 and 4 combined)



Relationship between Subtests

Like reading and comprehension, the other subtests have a predictive relationship between them—some stronger than others. The following figures (**9** through **12**) show the relationships between performance on the fluency subtasks and performance on ORF. The familiar and invented words subtests (**Figures 10** and **12**) show the tightest linear correlations; in other words, the better a child scored on these subtests, the higher their reading fluency score was. This is especially the case with familiar word reading, although only six of the familiar words re-appeared in the oral reading passage. This is additional evidence of the importance of systematically teaching the foundational skills as a prerequisite to promoting connected text reading fluency and comprehension. The content in the primary school curriculum for literacy does not spell out these skills.

There is also a very strong correspondence between the subtests at the lower-levels: knowing letters enabled a child to read syllables (correlation coefficient = 0.815), reading syllables enabled one to read familiar words (coefficient = 0.910), and knowing words enabled one to read passages with fluency (0.943).

Therefore instructional strategies that build on these foundational skills in a regular, systematic way are likely to create fluent readers faster than approaches that do not build on letter and syllable-level knowledge as a prerequisite.

Figure 9: Letter Naming and ORF

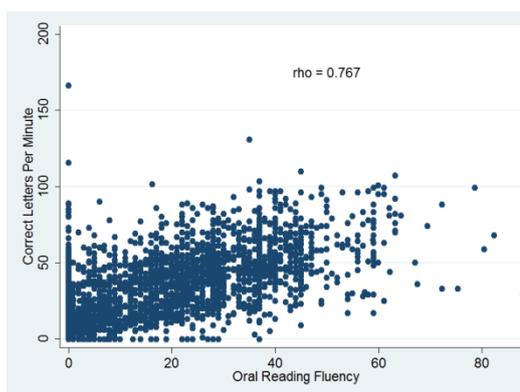


Figure 10: Invented Word Reading and ORF

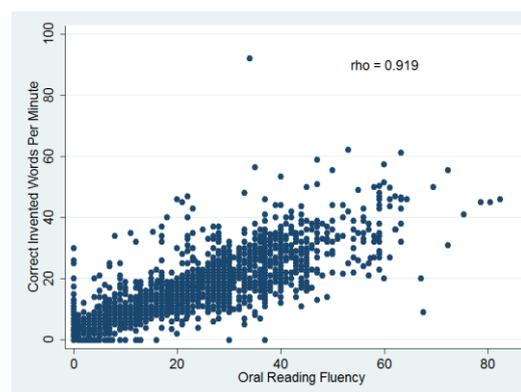


Figure 11: Syllable Reading and ORF

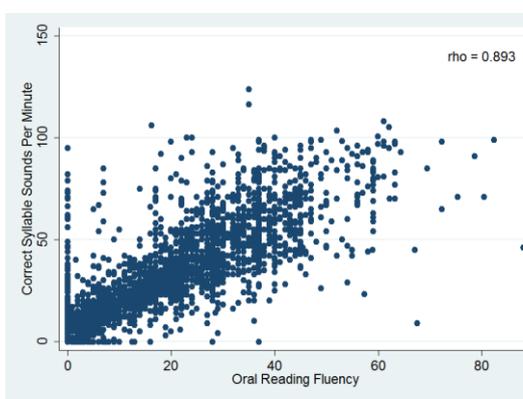
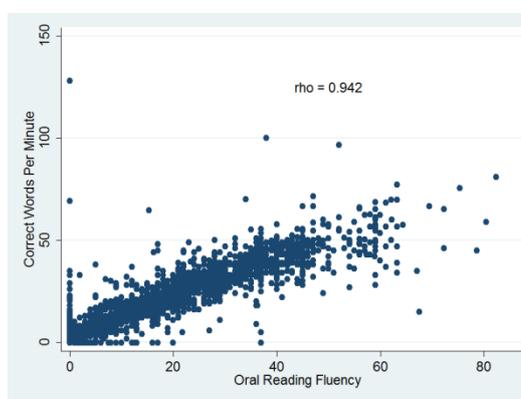


Figure 12: Familiar Words and ORF



Comparison of Baseline to Endline

Sample Populations

As indicated in **Table 1** above (Research Design), the sample size increased from baseline to midterm and endline, allowing us to be more confident in the generalizability of findings and changes over time. In terms of the characteristics of the samples, the descriptive data (non-weighted) also indicate that the populations were largely similar (see **Table 12**). Students at midterm were also from a low socioeconomic background; 59% of learners reported attending kindergarten or nursery school at midterm as opposed to 63% at endline—a negligible difference. Also, the percentage of learners who reported having textbooks was low, and decreasing over time. This may be because books have not been replenished in schools and because there is high enrollment.

Results of the endline assessment show slight improvement among Standard 4 students sampled, but very little progress among Standard 2 students. The overall scores of students who completed Standard 3 are still below the Standard 1 benchmarks established by the MoEST after the midline assessment in 2011, except in letter naming. At baseline and endline, zero scores represent the largest proportion of Standard 2 and Standard 4 student scores.

Table 12: Changes in Descriptive Statistics over Time

Characteristic	2010 Baseline	2011 Midterm	2012 Endline	Trend
Child attended kindergarten or nursery	58%	59%	68%	Increase

Characteristic	2010 Baseline	2011 Midterm	2012 Endline	Trend
Has textbook	31%	30%	20%	Decrease
Has 1-3 items at home	61%	67%	65%	Neither
Has 4-6 items at home	27%	21%	19%	Decrease
Has 7 or more items at home	5%	4%	5%	Neither
Avg Std 2/Std 4 class size	141/111	n/a	129/100	Decrease
Child speaks Chichewa at home	74%	72%	84%	Increase
Child is repeating the current Standard	7%	7%	10%	Increase
Teacher is trained	92%	75%	83%	Decrease
School has a library	39%	27%	26%	Decrease

Comparison of 2010, 2011, and 2012 EGRA Data

One of MTPDS's objectives is to look at changes in reading results over time. These points were EGRA studies administered in November of each year, consisting of the baseline (Nov. 2010), midterm (Nov. 2011), and endline (Nov. 2012). The purpose is to see whether any changes at the systemic level or large-scale reading program inputs such as MTPDS, Tikwere! and Read Malawi have influenced the teaching and learning of reading in the early grades.⁹ Additionally, the 2011 and 2012 results serve to validate 2010 results, to the extent that they reveal similar patterns, especially since sample size was larger in 2011 and 2012. As a reminder, most subtests were adapted simply by changing the order of the items presented. Those that were changed significantly—the reading comprehension and listening comprehension subtests—were analyzed and equated. The figures presented in **Table 13** below are equated scores, so the table can be read with the assumption that all assessments were of equivalent difficulty across all three years. To discuss the percent change in learners reaching the benchmark, effect sizes are classified as low (0.0 to 0.19), moderate (0.2 to 0.39), or strong (0.4 and up).

Table 13: Subtest Results for All Three Years, and Effect Sizes

Subtest	Standard	Mean* 2010	Mean* 2011	Mean* 2012	Effect Size**
Letter naming (100 items, correct per minute)	2	2.3	3.1	5.7	0.31***
	4	21.3	24.0	26.7	0.29
Syllable segmentation	2	46.0%	27.5%	35.2%	-0.27

⁹ Given the wide variety of different interventions taking place in Malawi, this study cannot be read as a program evaluation of MTPDS. For a report on an isolated study of reading conducted by MTPDS in addition to the program's national-level policy, advocacy and curriculum work as well as national-scale CPD inputs, please see separate report "EGRA Reading Intervention Study, 2012."

Subtest	Standard	Mean* 2010	Mean* 2011	Mean* 2012	Effect Size**
(10 items - % correct)	4	66.4%	52.7%	58.6%	-0.25
Initial sound identification (10 items - % correct)	2	5.0%	3.9%	3.1%	-0.15
	4	14.9%	7.9%	6.7%	-0.54***
Syllable reading (50 items, correct per minute)	2	1.3	0.8	3.2	0.19
	4	18.9	20.2	26.3	0.34***
Familiar word reading (50 items, correct per minute)	2	0.8	0.5	1.9	0.16
	4	11.3	14.0	16.9	0.38***
Nonsense word reading (50 items, correct per minute)	2	0.5	0.3	1.2	0.17
	4	7.7	8.5	11.4	0.33***
Oral reading fluency (61, 54, & 59 items, correct per minute)	2	0.8	0.4	1.3	0.08
	4	11.3	13.9	15.4	0.28
Reading comprehension (5 questions - % correct)	2	0.81%	0.25%	0.86%	0.01
	4	8.76%	10.57%	13.34%	0.33***
Listening comprehension (5 questions - % correct)	2	31.13%	46.26%	32.63%	0.05
	4	50.70%	63.72%	49.96%	-0.03

* The mean is either a measure of “correct items read per minute” (cpm) or “percent correct,” depending on the subtest.

** Baseline to endline

*** Denotes a moderate or strong effect size. An effect size of .2 or .3 is considered “moderate.” Scores greater than .3 are ‘strong’.

Table 13 above shows the mean results on the EGRA assessments from 2010, 2011, and 2012 as well as the effect size of the differences between baseline and endline. For most of the subtests, little to no change was found between 2010 and 2011 (effect sizes were 0.15 or lower—see midterm report for details). The MTPDS midterm EGRA report concluded that:

“The results of the 2011 midterm study were broadly similar to those of the EGRA baseline study conducted in 2010. The same pattern of strengths and weaknesses was found across subtests between Standards 2 and 4.”

The subtests that showed a change with a medium or large effect size between 2010 and 2011 were syllable segmentation, initial sound identification, and listening comprehension. In 2012, the pattern has changed, mostly for the better, across all subtests. There are positive changes in scores, but the practical significance (the effect sizes) is small—though not as small as in 2011. The largest positive effect size is seen among Standard 4 students in familiar word reading, and the other key skills subtests showed a moderate effect size in Standard 4. All effect sizes are larger for Standard 4 students, with the exception of listening comprehension and letter naming. The only Standard 2 subtest with a moderate effect size was letter naming, which increased from 2.3 to 5.7 letters identified in one minute from baseline to endline, respectively.

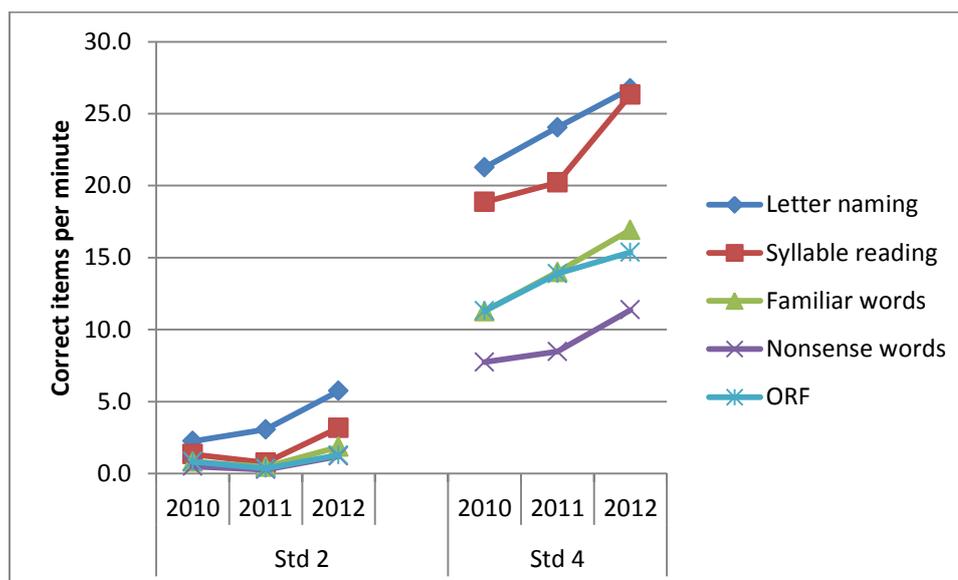
As in 2011, for the phonological awareness subtests—syllable segmentation and initial sound identification—the effect size is small to medium and *negative*, showing a performance decrease (from baseline—there was a slight increase between midterm and

endline). The real change in scores is only 1 out of 10 items for syllable segmentation and less than 1 out of 10 items for initial sound identification. For subtests with only 10 items it is difficult to discern an actual change in ability when the average score changes by so little. The most important conclusion is not that students' skills in these areas were decreasing—it is that they were not improving and they were extremely low for both Standard 2 and 4 students. The initial sound scores are particularly alarming because both Standard 2 and 4 students were only able to answer one item correctly on average. This is most likely because this skill is not being explicitly taught to students, and instead they are learning to read through whole word memorization or syllables. Additional details on these two subtests were provided in the section “Results by Subtest” (see Figures 4 and 5).

The strong positive change in listening comprehension that was identified between baseline and midterm was not found again between midterm and endline or in the comparison between baseline and endline—in fact there was a decrease in achievement between baseline and endline, thus decreasing the overall effect size from midterm to endline.

Figure 13 shows the data presented in **Table 13** above in a visual format for the fluency subtests for Standard 2 and Standard 4, respectively. We can see that for both groups, gains were largely unnoticeable between baseline and endline, but larger between midterm and endline. However, for Standard 2 students, although this pattern is present, the difference in actual gains is too small, and the starting point too low, to really infer any actual evolution in students' skills, with the exception of letter naming. Scores for Standard 4 are also higher at endline than at midterm or baseline, but with much larger gains—anywhere between 3.5 and 7.5 words per minute. The effect size from **Table 13** confirms that this is of notable significance. The 2011 midterm report estimated that the average inter-grade gain was about 7.25 words per minute, and the 2012 data confirm a gain of approximately 7 words per minute between grades; therefore the magnitude of change seen in Standard 4 is much more likely to reflect actual positive evolution in achievement than the Standard 2 gains.

Figure 13: Evolution of Fluency Measures from Baseline to Endline



This is encouraging news, and it may be the case that national-level efforts such as the ones described in the section above are having an effect. However, the reason these efforts are only affecting students in the higher grade, and why they are not larger, deserves attention. There are several possible explanations for this. First, the conditions in lower primary

(Standards 1 and 2) classrooms may be very different than in Standard 3 and 4 classrooms; for example, classroom infrastructure—where there are classrooms at all. This study found that student-to-teacher ratio is much higher in Standards 1 and 2. Teachers in the upper lower primary have been observed to be more committed to teaching learners to read. These classes most of the time are handled by trainee teachers who are dedicated to teaching as they are under intense monitoring and supervision by college lecturers and mentors in the districts. We should recall that the Standard 4 cohort assessed at endline was the same cohort (though not necessarily the same students) as that assessed at baseline for Standard 2. Therefore it seems that efforts to improve school quality are accumulating over the years, and thus having an effect at the higher grades. Additionally, learners in Standards 3 and 4 have been exposed to a variety of reading materials over the years unlike learners in Standards 1 and 2. The demographic information provided in Table 4, above, confirm that Standard 4 children are much more likely to have a school textbook or reader than Standard 2.

Nevertheless, the teaching of basic skills in the first year of primary school remains inefficient and the result even by Standard 4—15.4 words per minute reading fluency—is inadequate. Some of the national programs described in the introductory section (i.e., Read Malawi) are providing books and teaching materials to schools, but the level of these books is aimed at emerging readers, and not basic skills development. More importantly, teachers need to know how to use books, and students need to spend time using them.

Finally, we have to emphasize that the inter-grade gain in ORF is still very small at 7 words per minute, so this may be just the effect of time and not the effect of any teaching or school conditions. More research must be done to find out what inter-grade gains are possible in this context through more sustained, deliberate reading instruction and also, possibly, comparisons with other countries.

Evolution in Scores by Subtest and Distribution of Scores

Another way to look at the evolution in scores between baseline and endline is by only looking at the ORF measurement, a timed subtest in which students are given a minute in which to read a short story (61 words in the baseline passage, 59 at endline). This subtest requires the students to draw on all of the skills measured in the other components—letter, sound, and syllable knowledge; identification of familiar or sight words; and decoding of unfamiliar words. For this reason, it is a very good indicator of reading ability and progress over time. The following graphs show the distribution of fluency scores measured in Standard 2 and Standard 4 at baseline (**Figure 14**) and endline (**Figure 15**). The scores are grouped according to the range of correct words per minute, in intervals of 10.

Figure 14: Baseline ORF Scores by Range

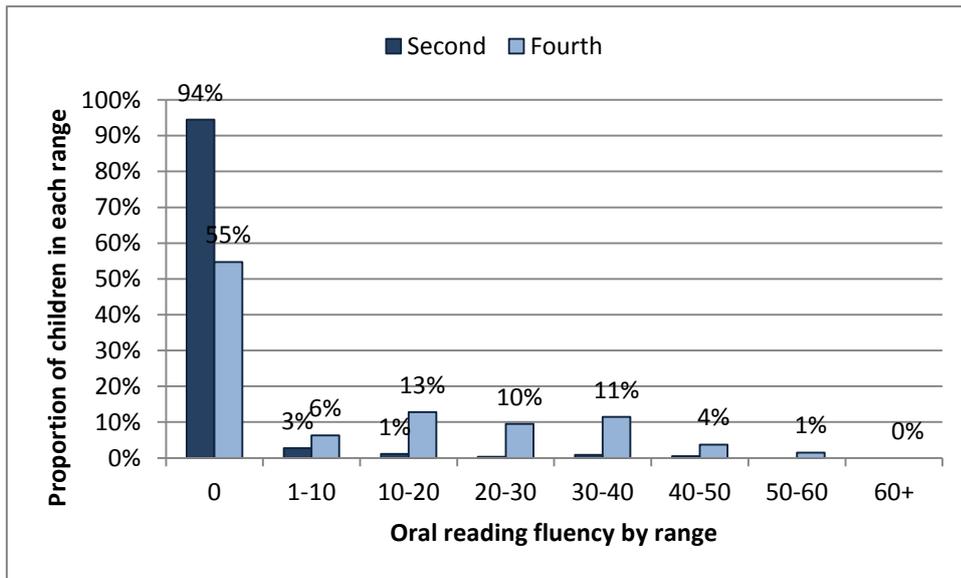
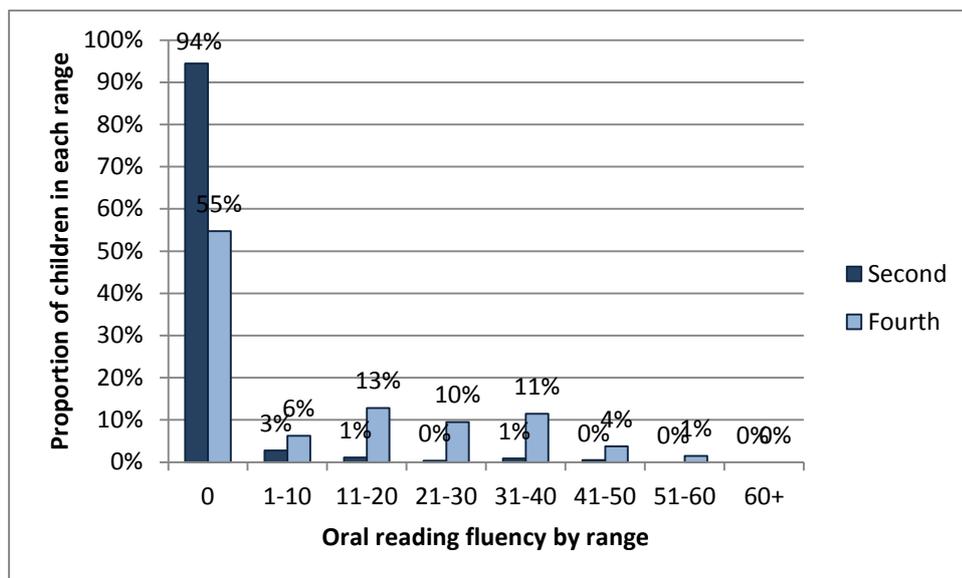
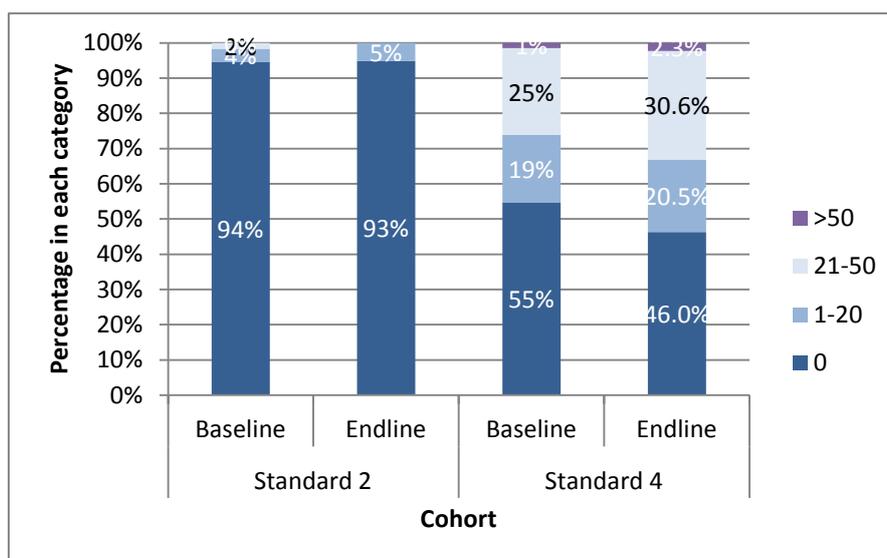


Figure 15: Endline ORF Scores by Range



These graphs also present zero scores as a range. At both baseline and endline, this is the category with the largest share of students for Standards 2 and 4 alike. At baseline, the Standard 4 students who did show any measurable reading ability were distributed primarily across the median ranges of ability (11-40 correct words per minute), with very few students at the high end. However, in 2012, the distribution of scores follows a much more standard bell curve, with more students at both the low and high ends of the spectrum, and the highest proportion grouped around the mean. This is most likely due to the larger sample size at endline. Additionally, the decrease in zero scores shows in the redistribution of means in the 1-10 range and higher. **Figure 16**, below, represents the same data with the ranges collapsed roughly according to established benchmarks (see next section). This presentation shows clearly that the majority of students were those who could not read even one word of the reading passage presented, except in Standard 4 in 2012. For this group there is a noticeable increase in the share of students who read at all, and especially those who read more than 20 correct words per minute.

Figure 16: Oral Reading Fluency Scores by Range from Baseline to Endline



Benchmarks and Targets

In 2011, the Malawi MoEST’s National EGRA Coordination Committee, in collaboration with MTPDS, established a set of benchmarks and annual performance targets for Standards 2 and 4. In 2011, more than 10% of learners reached the benchmarks for letter naming (Standard 4), syllable segmentation (Standards 2 and 4), and listening comprehension (Standards 2 and 4). For the other subtests, the percentages reaching the benchmark were considerably lower. Less than 10% of learners reached the level expected in subtests that required decoding, which included syllable reading, familiar word reading, nonsense word reading, and ORF (with comprehension). **Table 14**, below, shows the results for 2012 according to benchmarks and targets for each Standard. Again in 2012, syllable segmentation was the only area where both Standard 2 and Standard 4 students met the performance target (20% of students scored 70% correct or better in Standard 2, and 30% of students scored 80% correct or better in Standard 4). Standard 4 students also surpassed the 2012 performance target for letter naming, and the share of Standard 2 students meeting this benchmark nearly tripled, but this was not sufficient to reach the target of 10% scoring 24 correct letter sounds per minute or better. The only other area where more than 10% of students achieved the benchmark is Standard 4 syllable reading—12.2% of students read 60 correct syllables per minute; however, the target for 2012 was 20%.¹⁰

Table 14: Percentage of Learners Meeting Benchmarks and Performance Targets from 2010 to 2012

Subtest	Standard	Benchmark	2012 Perf. target	% Reaching benchmark 2010	% Reaching benchmark 2011	% Reaching benchmark 2012
Letter naming	2	24+ clpm	10%	2.3%	2.1%	7.1%

¹⁰ As a reminder, this study sampled children at the *beginning* of Standard 2 and 4, but as a measurement of what they learned in Standard 1 and Standard 3. Therefore, the results are compared against the Standard 1 and 3 benchmarks.

Subtest	Standard	Benchmark	2012 Perf. target	% Reaching benchmark 2010	% Reaching benchmark 2011	% Reaching benchmark 2012	
	4	50 clpm	10%	11.5%	14.3%	15.4%	✓
Syllable segmentation	2	70% correct	20%	41.9%	24.7%	31.5%	✓
	4	80% correct	30%	52.0%	36.0%	48.6%	✓
Initial sound identification	2	80% correct	5%	0.1%	0.7%	0.7%	
	4	90% correct	5%	0.0%	0.8%	1.2%	
Syllable reading	2	30 cspm	10%	1.8%	0.4%	3.0%	
	4	60 cspm	20%	10.0%	7.3%	12.2%	
Familiar word reading	2	20 cwpm	10%	1.8%	0.6%	2.6%	
	4	45 cwpm	10%	3.4%	5.2%	6.5%	
Nonsense word reading	2	15 cwpm	5%	1.4%	0.5%	2.5%	
	4	40 cwpm	5%	0.6%	1.2%	2.9%	
Oral reading fluency	2	20 cwpm	10%	1.7%	0.4%	1.8%	
	4	50 cwpm	10%	2.7%	2.1%	3.5%	
Reading comprehension	2	40% correct	5%	1.4%	0.0%	0.1%	
	4	80% correct	5%	0.2%	0.0%	0.0%	
Listening comprehension	2	60% correct	30%	20.3%	24.3%	18.7%	
	4	80% correct	30%	24.1%	21.7%	24.6%	

clpm = correct letters per minute; cspm = correct syllables per minute; cwpm = correct words per minute.

As indicated earlier in the section describing overall results and specific subtests, when we do not consider zero scores, more students are closer to the benchmarks, and the picture improves somewhat: the Standard 4 means are above the Standard 2 benchmarks, but they are still far from the Standard 3 benchmark. In particular, for the initial sound identification task the low scores and low percentage of students meeting the benchmark can be partially explained by the high number of zero scores and the fact that many students may have answered with syllables instead of sounds. Yet it is still important to recognize that these students with zero scores were part of the system and are being left behind. The reasons for this need to be examined and addressed.

Results Disaggregated by Gender and Region

Results by Gender

In recent years there have been efforts in Malawi to promote girls education, such as the Girls Attainment of Basic Literacy Education (GABLE SMC-EQ), Ambassador’s Girls Scholarship Programme, Promotion of Girls Education and various community initiatives in support of girls education like mother groups and counselors, among others. These

have been established to ensure that girls remain in school. According to EMIS data more girls enroll in the early grades but when they come of age most of them drop out gradually. This is due to casual labour, long distance travel to school, early marriages, family responsibility as some of them could be taking care of their siblings in case of orphans. The dropout rate becomes high in the upper classes. A major contributing factor is that of cultural practices that deter girls from going further with education.

There is no significant difference between the performance of girls and boys. The range of scores across districts is still very low in Standard 2—fewer than 10 correct items per minute. The Shire Highlands region stands out as having the highest score on the letter naming task, but is similar to the national average on all other subtasks. On the other hand, the Northern region outperforms the national average on all tasks, albeit not by much. Shire Highlands, Central Western, and Northern regions had the highest performance and the most significant gains in performance since baseline, whereas the Southern Eastern region had the lowest performance and no gains since baseline.

Table 15 shows the difference in scores among boys and girls in the Standard 2 sample. Given that the overall scores are very low, with little variation, it is not surprising that there were only minor differences between boys and girls in the sample. These are not statistically significant differences, indicating that any differences in scores—positive or negative—could be due to random chance, and being a boy or a girl was not a predictor of achievement.

Table 15: Standard 2 Results Disaggregated by Gender

Subtest	Sex	Mean	Standard error	Difference	p-value	% zero scores
Letter naming (100 items, correct per minute)	Boys	6.0	0.7	0.6	0.560	48.0%
	Girls	5.5	0.7	-0.6		50.9%
Syllable reading (50 items, correct per minute)	Boys	3.4	0.5	0.4	0.372	71.3%
	Girls	3.0	0.4	-0.4		70.4%
Familiar word reading (50 items, correct per minute)	Boys	1.9	0.4	0.0	0.947	79.7%
	Girls	1.8	0.5	0.0		83.0%
Nonsense word reading (50 items, correct per minute)	Boys	1.3	0.3	0.1	0.707	85.9%
	Girls	1.2	0.3	-0.1		89.4%
Oral reading fluency (59 items, correct per minute)	Boys	1.3	0.4	0.1	0.721	89.1%
	Girls	1.2	0.3	-0.1		91.3%
Syllable segmentation (10 items - % correct)	Boys	35.4%	1.9%	0.0	0.901	42.3%
	Girls	35.1%	2.0%	0.0		44.5%
Initial sound identification (10 items - % correct)	Boys	3.1%	0.4%	0.0	0.865	86.5%
	Girls	3.1%	0.4%	0.0		86.7%

Subtest	Sex	Mean	Standard error	Difference	p-value	% zero scores
Reading comprehension (5 questions - % correct)	Boys	0.9%	0.2%	0.0	0.466	93.7%
	Girls	0.8%	0.2%	0.0		95.0%
Listening comprehension (5 questions - % correct)	Boys	33.5%	1.6%	0.0	0.255	30.2%
	Girls	31.8%	1.3%	0.0		31.1%

Table 16 shows the difference in scores among boys and girls in the Standard 4 sample. Although boys had lower scores, the actual difference is only 1-3 items per minute, or percentage points in the case of untimed subtask. These are not important differences, nor are they statistically significant, except in the case of listening comprehension. Again, the actual difference is too minor to signify anything inequitable about reading achievement.

Table 16: Standard 4 Results Disaggregated by Gender

Subtest	Sex	Mean	Standard error	Difference	p-value	% zero scores
Letter naming (100 items, correct per minute)	Boys	25.2	1.3	-3.0	0.092	11.9%
	Girls	28.2	1.7	3.0		11.1%
Syllable reading (50 items, correct per minute)	Boys	24.7	1.6	-3.2	0.148	22.0%
	Girls	27.9	2.5	3.2		18.5%
Familiar word reading (50 items, correct per minute)	Boys	15.8	1.3	-2.2	0.149	28.6%
	Girls	18.0	1.6	2.2		29.1%
Nonsense word reading (50 items, correct per minute)	Boys	10.3	0.9	-2.1	0.148	34.3%
	Girls	12.4	1.5	2.1		34.2%
Oral reading fluency (59 items, correct per minute)	Boys	14.2	1.2	-2.3	0.069	39.9%
	Girls	16.5	1.4	2.3		38.3%
Syllable segmentation (10 items - % correct)	Boys	60%	0.0	0.0	0.347	20.4%
	Girls	60%	0.0	0.0		20.2%
Initial sound identification (10 items - % correct)	Boys	10%	0.0	0.0	0.210	75.7%
	Girls	10%	0.0	0.0		70.2%
Reading comprehension (5 questions - % correct)	Boys	12.0%	0.0	0.0	0.013	11.9%
	Girls	14.7%	0.0	0.0		11.1%
Listening comprehension (5 questions - % correct)	Boys	53.1%	0.0	0.1	0.002	22.0%
	Girls	46.9%	0.0	-0.1		18.5%

* Statistically significant at <.05 level

Regional Results

The sample size used in 2012 allows us to make some comparisons by region, since each region had a sample size of more than 400 students. In addition to national CPD workshops delivered to all Standard 1-4 teachers, the MTPDS reading intervention program in Standard

1 covered seven districts: Salima, Ntchisi (Central Eastern Division); Ntcheu (Central West Education Division); Mzimba North (Northern Education Division); Zomba Rural (South East Education Division); Blantyre Rural (South West Education Division), and Thyolo (Shire Highlands Education Division), and now reaches 1,310 schools.

Although schools in these regions were selected for the random national sample, and schools that received the intervention were included in the analysis, results were weighted proportionately to their increased chance of being selected. The districts of Mzimba North, Ntcheu, Zomba Rural, Blantyre Rural, and Thyolo have only been participating for one year, beginning in March 2012, whereas Salima and Ntchisi have been in the program for two years. **Table 17** provides the results by Division for Standard 2, and **Table 18** provides the results for Standard 4. The national average is also presented in the last row as a convenient reference. As a reminder, this national assessment is not an impact evaluation of MTPDS and is not expected to detect changes due only to these activities; however, to explain any changes from baseline to endline it is useful to know what type of national-level changes may have taken place during the three-year period.

Table 17: Fluency Subtest Results, by Division, Standard 2

	Letter naming	Syllable reading	Familiar word reading	Nonsense word reading	ORF
Central Eastern	7.6	5.2	3.4	2.2	3.0
Central Western	4.2	2.6	1.5	0.9	1.1
Northern	6.7	4.4	2.8	1.7	1.5
Shire Highland	8.6	2.6	1.3	0.9	0.7
South Eastern	4.3	2.0	0.9	0.5	0.6
South Western	5.2	2.7	1.6	1.2	1.1
National Average	5.7	3.2	1.9	1.2	1.3

Table 18: Subtest Results, by Division, Standard 4

	Letter naming	Syllable reading	Familiar word reading	Nonsense word reading	ORF
Central Eastern	26.8	27.8	19.3	12.1	17.9
Central Western	29.4	35.2	22.0	16.2	20.1
Northern	17.1	16.9	12.1	7.1	10.3
Shire Highland	37.8	27.5	18.9	12.3	17.0
South Eastern	29.8	29.9	19.8	12.9	18.8
South Western	24.4	21.5	11.9	8.2	10.8
National Average	26.7	26.3	16.9	11.4	15.4

Figures 17 and 18 that follow present these results graphically, which allows one to see at a glance both the differences between grades and the differences compared to the national average.

Figure 17: Standard 2 Fluency Scores by Region

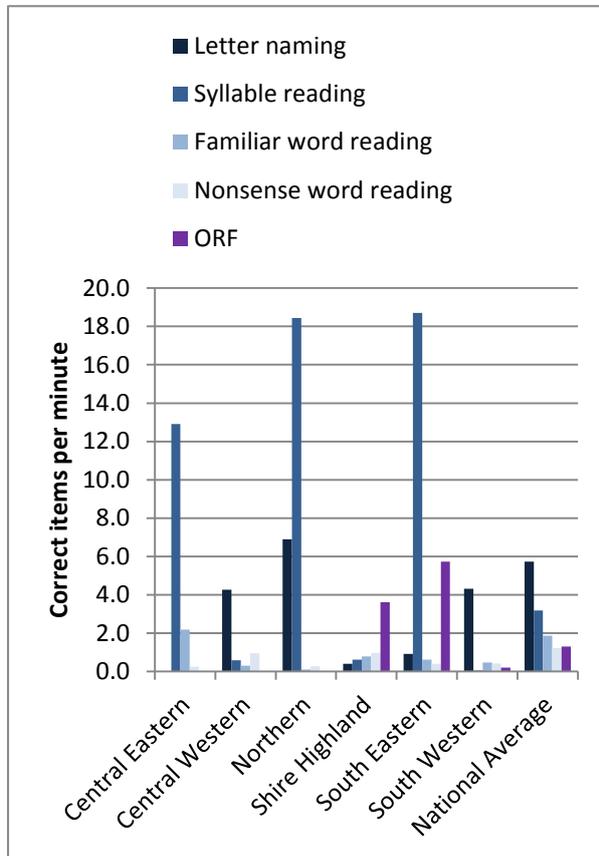
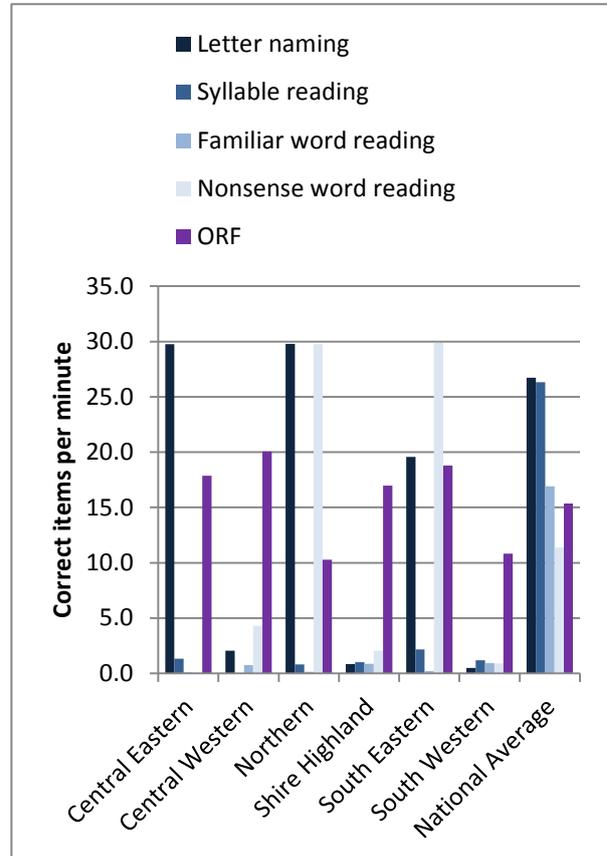


Figure 18: Standard 4 Fluency Scores by Region



At midterm, Standard 2 results on fluency were low across all divisions, and only Standard 4 results varied by region—the Shire Highlands and the South Eastern divisions fared the best. The Central Eastern and Western divisions were in the middle, while the Northern division fared least well. The low performance of the Northern division, when compared to other districts, was statistically significant.

In 2012, the picture is slightly different. The range of scores was still very low in Standard 2—fewer than 10 correct items per minute on all tasks. The Shire Highlands region stands out as having the highest score on the letter naming task, but is similar to the national average on all other subtasks. On the other hand, the Northern region outperformed the national average on all tasks, albeit not by much. The South Eastern region fared least well. However, only the Shire Highlands letter names score is statistically significant, and none of the real differences is large enough to be meaningful, so it is unlikely that there are instructional or demographic differences underlying these variations¹¹.

More variation is seen across Standard 4 classrooms, as well as larger ranges in achievement. Again, Shire Highlands performed the highest on letter naming, but very close

¹¹ An annex on demographics by region will be sent as an annex to this report at a later time.

to the national average on the rest of the subtasks. The advantage seen by Northern students in Standard 2 apparently did not translate to achievement in Standard 4, since they performed well below the national average on all subtasks. Regardless of raw scores, all three of these regions also showed the most progress from baseline to endline, with very strong effect sizes. **Table 19**, below, shows only the regions, grade and subtests that resulted in a strong effect size (greater than .4), positive or negative. For the remainder of the regions, the figures presented above provide details on performance across all fluency subtests; since there was no effect size detected over time, one can assume that the baseline and midterm scores were similar to what was reported above.

Table 19: Changes in Regional Scores over Time

Division	Grade	Subtest	Baseline	Midterm	Endline	% Increase	Effect Size
Central Western	Std 2	Letter Names	0.9	1.8	4.2	3.7	0.497
		Syllable Sounds	0.4	0.5	2.6	6.1	0.415
	Std 4	Letter Names	13.0	19.0	29.4	1.3	0.911
		Syllable Sounds	13.3	18.9	35.2	1.6	0.988
		Familiar Words	8.3	13.7	22.0	1.7	0.997
	Std 4	Invented Words	6.0	7.8	16.2	1.7	0.838
		Reading Fluency	7.2	14.2	20.1	1.8	0.952
Northern	Std 2	Letter Names	0.9	1.7	6.7	6.7	0.626
		Syllable Sounds	0.1	1.0	4.4	39.1	0.528
		Familiar Words	0.1	0.7	2.8	35.1	0.420
	Std 4	Invented Words	0.1	0.3	1.7	29.1	0.437
		Letter Names	10.4	12.1	17.1	0.6	0.528
		Syllable Sounds	7.7	11.3	16.9	1.2	0.572
		Familiar Words	3.5	8.0	12.1	2.5	0.664
	Std 4	Invented Words	3.0	4.9	7.1	1.4	0.476
Reading Fluency		4.3	8.3	10.3	1.4	0.511	
Shire Highlands	Std 2	Letter Names	4.1	4.3	8.6	1.1	0.441
		Syllable Sounds	16.0	28.7	27.5	0.7	0.541
		Familiar Words	10.0	18.0	18.9	0.9	0.565
	Std 4	Invented Words	5.1	12.1	12.3	1.4	0.690
		Reading Fluency	9.6	17.1	17.0	0.8	0.469
Southern Eastern	Std 2	Syllable Sounds	5.0	1.3	2.0	-0.6	-0.406
		Familiar Words	3.0	0.7	0.9	-0.7	-0.544
		Invented Words	2.0	0.5	0.5	-0.8	-0.612
	Std 2	Reading Fluency	3.4	0.5	0.6	-0.8	-0.715

It may be interesting to look more closely at factors in the Shire Highlands region that might explain the consistently higher scores on letter naming in that region. According to project

monitoring, what has been observed in this division is that the district education management is quite supportive and has been on the forefront coming up with ways and means of assisting teaching and learning. PEAs are also providing adequate monitoring of the schools to ensure effective teaching and learning. These are only anecdotal observations, however, and not supported by the data from this study. Apart from that there might a number of influences such as teacher's commitment and attitude, influence of other programs, or class sizes. These influences may need to be looked into in more detail in a follow on to this report.

Factors Predicting Achievement, EGRA 2012

This section presents factors and learner characteristics that were found to be associated with learner achievement. This information is helpful because educators and policy makers can use it to prioritize actions. However, the results should not be taken to mean that the relationships are necessarily causal or exhaustive. Certain factors may vary along with learner achievement,

This section presents factors and learner characteristics that were found to be associated with learner achievement. The 2012 study revealed that out of the information included in the student context questionnaire, nothing stood out as being statistically significant for the entire sample, except for grade level and repeating. However, when looking at a subsample of children who read more than 40 correct words per minute, class size, repeating and having a textbook all increase the chances that a child will be in this group.

and other factors exist that were not measured in this study. Additionally, much of the information comes from an individual's self-reporting, which may not always be accurate, especially for certain questions, such as asking a child to report on the schooling level of his or her parents. There is also a very low distribution of scores, especially in Standard 2, as well as a lack of variance in learner characteristics, which limits the ability to run regression analyses.

It is also important to note that the EGRA study gathered data about students' reading ability at the beginning of Standards 2 and 4, in order to make inferences about the skills acquired in Standards 1 and 3, respectively. The student questionnaire asked relevant questions about the students and teachers at the beginning of Standards 2 and 4, so the answers cannot always be directly associated with the time, place, and methods that contributed most to the learning outcomes.

Student Factors

At midterm, the student factors that were associated with ORF were: whether the child had the school reading text (8.7 cwpm advantage); if the child had other books at home (4.5 cwpm advantage); whether home language was the same as the language of instruction at school (2.3 cwpm advantage); and if the father had graduated from secondary school (4.2 cwpm advantage). These patterns were not found again in the 2012 data. The endline study revealed that out of the information included in the student context questionnaire, nothing stood out as being statistically significant for the entire sample, except for grade level and repeating, therefore no additional analysis of factors has been presented. Other factors that may influence performance, such as time on task, or absenteeism, were not measured by this instrument, so these results should not be considered an exhaustive list of explanations of performance.

Tables 20 and **21** below are linear regressions for the entire sample of Standard 2 and 4 combined. For every additional grade that the student is in, we expect to see an increase in ORF of 6.9 words per minute, and that increase is statistically significant (See **Table 21**). Thus the gain between Standard 2 and Standard 4 would be 13.8 correct words per minute. For boys, the inter-grade gain was 6.9 (they are the constant represented in **Table 21**).

Table 20: Linear Regression: Grade vs. Oral Reading Fluency

Grade	Coefficient	Standard error	t	p> t	95% Confidence interval	
Grade	6.9	0.6	11.74	0	5.774	8.104
Constant *	-12.6	1.2	-10.09	0	-15.072	-10.144

*"Constant" refers to the y-intercept, or the average number of correct words per minute for a child at grade "0."

Table 21: Linear Regression: Gender vs. Oral Reading Fluency

Gender	Coefficient	Standard error	t	p> t	95% Confidence interval	
Female	0.7	0.6	1.23	0.22	-0.422	1.822
Constant *	6.9	0.5	13.66	0	5.914	7.910

*"Constant" refers to the y-intercept or, in this case, boys.

Table 22 shows the relationship between repeating and the scores on the oral reading fluency subtest. It tells us that students who are repeating a grade read, on average, 3.1 words per minute slower than their classmates who are in the current grade for the first time. More discussion of the effect of repeating is included in the next section.

Table 22: Linear Regression: Repeating vs. Oral Reading Fluency

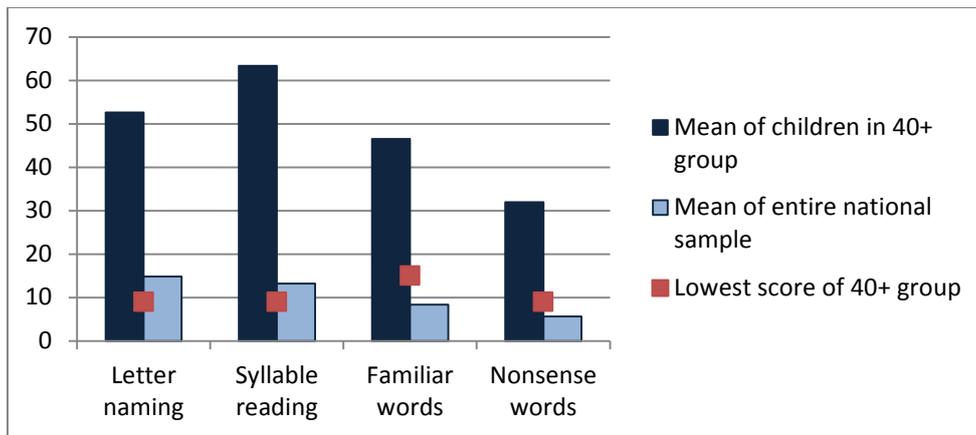
Repeat (Y/N)	Coefficient	Standard error	t	p > t	95% Confidence interval	
Repeating current grade	-3.1	0.8	-3.76	0	-4.659	-1.455
Constant	7.6	0.5	14.56	0	6.564	8.622

"Constant" refers to the y-intercept or, in this case, a child who has not repeated a grade.

Students who read 40 correct words per minute or more

Another way to try to find out under what conditions students succeed is by looking more closely at the performance and characteristics of a subset of students who demonstrate the highest performance among those in the sample. **Figure 19**, below, plots the scores of a subset of students who read with a fluency of 40 correct words per minute or more on the ORF passage. (We will refer to this group as "the 40+ group" from now on.) The students in this subset were mostly Standard 4 students (average age is 11 years old), but there were a few Standard 2 students included; 54% were girls. For each of the other fluency subtests, the table shows: the average score for the 40+ group (a dark bar); the minimum score for any child in the 40+ group (as a point); and the average of the entire Standard 4 sample (a light bar).

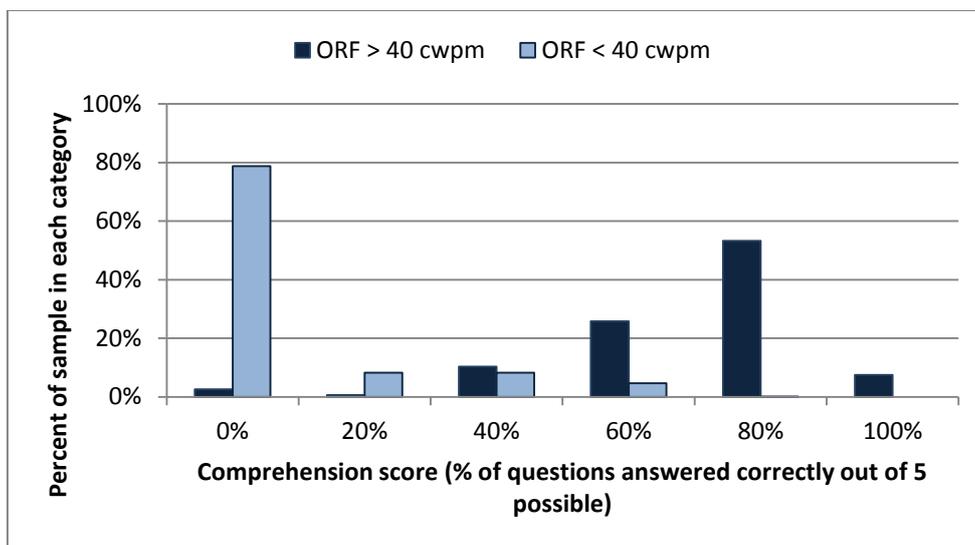
Figure 19: Subset of Students Who Read with 40 correct words per minute Fluency or Above



We can see clearly that students who were starting to read with fluency (more than 40 correct words per minute) also mastered the underlying components of reading above and beyond the majority of their peers. For example, these students could identify more than 50 letter names correctly in one minute—so it is likely that they knew all of the letters of the alphabet (name and sound). They could also read syllables at a rate of one syllable per second, or with very little hesitation. The minimum score recorded for the 40+ group is higher than the average of all of the students in the sample for familiar and unfamiliar word reading. This provides further evidence that students need to reach certain minimum standards on the basic skill sets in order to pass into the stage of reading as opposed to decoding.

Because the ultimate goal of reading is comprehension, it is also important to investigate whether these same students were also achieving higher rates of comprehension than the rest of the sample. **Figure 20**, below, shows the range of reading comprehension scores of the children who read more than 40 correct words per minute, compared to children who read fewer than 40 correct words per minute.

Figure 20: Reading Comprehension Scores by Subgroups of Achievement



* Standard 2 and Standard 4 combined; unequated scores.

The students in the group that read with a fluency of 40 cwpm or more were also more consistently demonstrating high levels of comprehension. The largest proportion of students in the overall sample read with no comprehension (no correct answer given), but the largest proportion of students in the 40+ group achieved 80% comprehension. The procedures for test administration require the assessor to stop the exercise at the end of 60 seconds, whether the child completed the passage or not, and the questions asked correspond only to the number of words of the text the child read (so they are never asked a question that they could not answer because they didn't read the text that corresponds to the answer). Therefore, children in the ORF < 40 group couldn't get 100% comprehension because they would not have been asked the 5th question. However, we can see from the graph that as children achieve a higher level of fluency, they also are more likely to understand all or most of what they read than those at a lower rate of fluency who are asked the same number of questions.

Logistic regression uses a dichotomous variable instead of a continuous variable as the outcome variable in the regression equation. Typically this dichotomous variable is thought of as success vs. failure. In **Table 23** below, logistic regression is used to show the odds of a student being in the 40+ words per minute group based on his or her fluency in other subtests. For example, a student with an odds ratio of 2 would be twice as likely to be in the 40+ group for each increment in the corresponding explanatory variable. This allows us to see how changes in performance in one area can be expected to change scores in the ORF for this subpopulation.

Table 23: Subtest Results for All Three Years, and Effect Sizes

ORF 40+ group	Odds ratio	Standard error	t	P > t	95% Conf. interval	
					Lower	Upper
Initial sound ident.	1.4	0.1	8.37	0	1.277	1.486
Syllable segmentation	1.3	0.0	6.95	0	1.192	1.371
Nonword reading	1.2	0.0	4.08	0	1.081	1.250
Familiar word reading	1.2	0.0	5.24	0	1.107	1.252
Syllable reading	1.1	0.0	9.98	0	1.060	1.091
Listening comp	1.9	0.2	6.28	0	1.543	2.296
Reading Comp	10.7	3.8	6.72	0	5.355	21.572

From this table we can see that for every additional question of the 10 initial sound identification questions that a child answered correctly, that child was 1.4 times more likely to be in the 40+ group. These odds ratio coefficients are much lower for the fluency tasks, but that is because each individual item is a multiplier. Thus, a child who read, for example, 31 familiar words correctly would be 1.2 times more likely to achieve 40+ ORF than a child who only read 30 familiar words correctly.

A logistic regression of the subpopulation of 40+ readers against the variables in the exit interview uncovers only three statistically significant variables (see **Annex 2**):

- The standard the child was in last year: Odds ratio of 3.22, meaning that a child who was repeating a class was 3 times more likely to read fewer than 40 words per minute than a child who was not repeating.
- The standard the child is in this year: Odds ratio of 6.63, meaning that from Standard 2 to Standard 3 or 3 to 4, a child was more than six times more likely to read 40 words per minute or more.
- If the child has the school textbook or not: Odds ratio of 2.31, meaning that a child with a school textbook was twice as likely to be in the 40+ group of readers.

These correlations support findings in other international literature and other EGRA studies. Students who repeat a grade are often found to have lower reading scores than others, and also students who are over the normal age for their grade. This can be explained because students who have been asked to repeat a grade are by definition those who are experiencing difficulty in school, possibly because of other characteristics such as poverty, high absences, etc. Therefore they were poor readers before being asked to repeat a grade. What is important to note, however, is that repeating a grade does not usually help to raise the level of achievement of the child. See, for example, research from Senegal by Glick and Sahn (2010)¹², which recommends “alternative measures to improve the skills of lagging students” besides repeating; or the literature review by CfBT on grade repetition in Sub-Saharan Africa, which concludes that “The majority of the studies undertaken to explore the effects of repetition on learning suggest that the practice does more harm than good.”¹³

Teacher and School-Level Factors

The analysis (linear regression) of the 2012 dataset reveals far fewer predictive factors than 2011. In fact, the only variables that stood out as statistically significant based on teacher or school characteristics are:

- If the teacher sounds out new words two or more days per week: 5.0 words per minute advantage;
- If students read aloud to the teacher two or more days per week: -13.2 words per minute *disadvantage*;
- If the PTA meets every 2-3 months or more often: 4.3 words per minute advantage; and
- Class size: Larger class size has a strong negative relationship with a child’s reading ability (see further statistical explanation below).

Because the teacher-level variables are asking questions of teachers in Standard 2 or 4, but the assessment actually measures learning that took place in the previous school year (Standard 1 or 3), the teacher level factors cannot be assumed to accurately reflect the influence that the teacher has on the scores. The responses are also self-reported and not otherwise verified by the research team. It is possible that during the first month of school (October 2012) prior to the assessment, the Standard 2 or Standard 4 teacher helped

¹² Glick, P. and Sahn, D. (2010), “Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis.” *The World Bank Economic Review*, 24 (1): 93-120. Washington, D.C.: The World Bank.

¹³ Ndaruhutse, S. (2008). “Grade repetition in primary schools in Sub-Saharan Africa: an evidence base for change”. Berkshire: CfBT Education Trust.

students review the previous year’s material and regain any losses that occurred over the break, and therefore sounding out unfamiliar words was part of this review process. The reason why there would be such a strong negative correlation with the “reading aloud” variable is a mystery. It could be that teachers are only telling the researchers what they think they want to hear, or it could be that “reading aloud” is really being interpreted as “recitation,” which is a common instructional strategy but not one that builds the sound- and syllable-level skills that are linked to fluent reading.

Class size was estimated using the teacher-reported enrollment of boys and girls in the class where students were sampled for the study, or enrolment figures were obtained from the head teacher using the school enrolment table or class attendance register. The data from this national study (**Table 24**) show that each 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).

Table 24: Linear Regression : Class Size vs. Oral Reading Fluency

Class size	Coefficient	Standard error	t	p > t	95% Confidence interval	
Class size	-0.02	0.008	-2.87	0.005	-0.040	-0.007
Constant *	10.34813	1.052	9.83	0	8.271	12.42

*“Constant” refers to the y-intercept, or the ORF of a child who is in a class of “0” students.

Conclusions and Recommendations

The past three years have been an opportunity for USAID and the Malawi MoEST to study reading achievement and school characteristics nationwide. At each stage, recommendations have been made that have informed the MoEST’s curriculum review and teacher-training programs. Some important changes that have been addressed during the curriculum review process spearheaded by MIE (2011-2012) include:

- Teaching all letters of the alphabet rather than just vowels;
- Focusing on explicit literacy development strategies in Standard 1 (distinct from language learning and including essential components);
- Reducing the Standard 1 “Introduction to school life and learning program” from 14 weeks to 7 weeks, to allow more time for basic skills building;
- Developing and delivering a national CPD module titled *Developing Literacy: Effective Teaching Techniques* (MoEST, 2012b);
- Completing a comprehensive review of textbooks and teachers’ guides in Standards 1–4, and substantial revisions to the Chichewa textbooks that render them much more useful for the teaching of essential decoding skills to learners; and
- Employing strategies for reducing class size. These strategies include ensuring that learners enroll at the expected age of 6 years; enforcing measures to minimize repetition; redeploying teachers within the school to ensure a maximum class size of 60 in the early grades; promoting use of double shifting; deploying only one trained

teacher per class; and, where not all teachers are trained, prioritizing the deployment of trained teachers to the early grades.

Additionally, efforts supported by development partners include expansion of the *Maziko a Kuwerenga* reading program and provision of the CPD literacy modules nationwide. In October 2012, a national conference on early grade literacy under the theme *Together we can build a literate nation: It starts with you* was organized. Presided over by the Minister of Education, it included participants from MoEST, international and local organizations, private and public schools, and other academic institutions to share experiences and explore best practices of early grade reading in Malawi. Thus momentum is clearly building for improving reading skills, and the results from the 2012 study suggest that these efforts are paying off. Nationwide, students' scores are increasing, particularly in Standard 4 (a reflection of learning from Standard 1 to 3), yet they remain well below established benchmarks and targets.

As mentioned in the limitations section above, the process of data collection changed from midterm to endline to use of a digital data collection tool instead of paper and pencil. Additionally, MoEST employees were used to collect data. This was done in an effort to build local capacity to conduct this type of assessment. It is possible that this could have increased the risk of assessor bias towards inflated results; however, by this logic Standard 2 and 4 scores would have increased at the same rate. Since it was only Standard 4 results that increased, it is unlikely that assessors conspired to increase scores at endline just for this group, and by similar margins across regions.

A separate randomized control study carried out under MTPDS aimed to show how much of an effect could be made on reading scores when a systematic focus on early literacy skills is introduced into Standard 1. This study included a small sample of schools in Salima and Ntchisi that received the *Maziko a Kuwerenga* reading program and compared results after one year to control schools that did not have the intervention. The results showed that while overall performance remains low, large absolute and relative gains in reading performance were achieved in the intervention schools that did not occur in control schools. Students in control schools demonstrated little, if any, measurable pre-reading skill, including in naming the letters of the alphabet. Although scores for students in the intervention schools were still very low, they were close to the level of achievement demonstrated by Standard 4 learners described in the present national study report.

From both the national study and the intervention study, there were few characteristics of students, teachers, and schools that stood out as influencing scores significantly. Class size, grade, repeating a grade, having the textbook, and receiving supportive visits from an instructional coach have all emerged as significant factors, but not in every specific sub-sample studied. Therefore it is difficult to draw broad conclusions about what can be done on a national scale to improve outcomes across the board. On the one hand, this is frustrating, but on the other hand it is also encouraging, because we can also be confident about things that are NOT a factor; for example, gender is no longer a determining factor, nor is socioeconomic status. Region, language, school size, preschool attendance, or a supportive home environment are not factors. Being in a school receiving the *Maziko a Kuwerenga* reading program and receiving supporting instructional coaching visits seems to be the only certain factor from our data, and it is likely that the program will be effective in a variety of environments and contexts. Other things that would compound the gains would be addressing class size, text book availability, and teacher attitudes and abilities through more capacity building.

From this we can conclude that by addressing fundamental reading skills such as phonological awareness and alphabetic principles in the first year of primary school, in a systematic and sustained way, reading can be measurably improved. The quantitative analysis from this national study and the smaller intervention study confirms that students who succeed on basic skills such as letter naming and familiar word reading have higher ORF and comprehension scores. Therefore it is important to focus on these basic skills early in Standard 1. Although carefully scripted lessons, teacher training, coaching, and providing text books are having a positive effect, and the gains are large relative to control schools that showed no reading skills at baseline or endline, they are not large enough to classify students as readers or allow them to read to learn. Additional efforts need to be made to improve classroom and home environments to promote reading and multiply the effects of the school reading programs in order to meet established benchmarks. We also need more information on the extent to which the teachers in the schools were actually using the inputs provided to know whether the gains measured are a minimum or maximum to be expected. This study shows improvement from 2010 to 2012, so it is likely that as the reading intervention becomes more established and teachers begin integrating these strategies into their teaching more systematically, these gains will increase.

Apart from continuing to act on recommendations made in 2010 and 2011 (see list below), and promoting proven best practices in reading instruction, some elements that stand out from this study include:

- Looking into the issue of absences, repeaters, and ages of students in each grade in order to minimize class sizes and maximize the use of class time to focus on skills that are appropriate for a majority of students in the class.
- Investigating the lower scores for the initial sound identification subtest.
- Conducting a case study in select regions (i.e., Shire Highlands) to determine if there are certain instructional practices that are being implemented in these regions to account for larger gains in these areas.
- Continued efforts to reduce class size, ensure that students are in an appropriate grade for their age, and ensuring increased instructional time devoted to basic reading practices in the first year of schooling.

Recommendations for Action

1. Start teaching the alphabet and letter sounds at the beginning of Standard 1
2. Teach decoding skills to early grade learners
3. Maximize time on task for reading practice (including lengthening the school day, and reinforcing reading concepts throughout the day, across subject areas)
4. Minimize the turnover of specialist teaching staff in the early grades
5. Prioritize investment in provision of appropriate early grade reading materials
6. Strengthen efforts to reduce class size
7. Review and develop strategies to address the needs of non-Chichewa-speaking learners
8. Focus on the needs of learners currently assessed with zero scores
9. Intensify coaching and supervision of teachers to support improved early grade reading
10. Review pre-service teacher training curriculum to strengthen early literacy acquisition
11. Expand comprehensive reading interventions

12. Publicize and monitor literacy benchmarks
13. Develop, document, publicize, and implement a National Early Literacy Strategy

Annex 1. Statistically Significant Regressions for the Entire National Sample

2012 Class size

Number of strata	=	10	Number of obs	=	4510
Number of PSUs	=	193	Population size	=	1371403.9
			Design df	=	183
			F(1, 183)	=	8.23
			Prob > F	=	0.0046
			R-squared	=	0.0112

eq_orf	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
class_size	-.024405	.008509	-2.87	0.005	-.0411934	-.0076167
_cons	10.4984	1.067809	9.83	0.000	8.391602	12.6052

Annex 2. Statistically Significant Regressions on the Subpopulation of 40+ cwpm

Exit interview 12: What grade were you in last year?

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5177
Number of PSUs	=	202	Population size	=	1634275.4
			Design df	=	192
			F(1, 192)	=	180.27
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
exit_interview12	3.224145	.2811187	13.43	0.000	2.714727	3.829156
_cons	.0024002	.0006046	-23.95	0.000	.0014605	.0039447

Note: variances scaled within each stage to handle strata with a single sampling unit.

Exit interview 14: Do you have the school textbook?

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5186
Number of PSUs	=	202	Population size	=	1644395.5
			Design df	=	192
			F(1, 192)	=	13.63
			Prob > F	=	0.0003

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
exit_interview14	2.31226	.5250117	3.69	0.000	1.477545	3.618535
_cons	.0364289	.0058762	-20.53	0.000	.0265016	.050075

Note: variances scaled within each stage to handle strata with a single sampling unit.

Grade

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	81.96
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
grade	6.631466	1.38578	9.05	0.000	4.391424	10.01414
_cons	.0000558	.0000452	-12.09	0.000	.0000113	.0002756

Note: variances scaled within each stage to handle strata with a single sampling unit.

Fluency subtests

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5198
Number of PSUs	=	202	Population size	=	1650576.5
			Design df	=	192
			F(1, 192)	=	122.11
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
clpm	1.066945	.0062565	11.05	0.000	1.054676	1.079357
_cons	.0066071	.0017612	-18.83	0.000	.0039055	.0111776

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	99.51
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
csspm	1.075806	.0078802	9.98	0.000	1.060375	1.091461
_cons	.0037825	.0012397	-17.02	0.000	.0019817	.0072199

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	27.48
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
cwpm	1.177378	.0366721	5.24	0.000	1.107223	1.251978
_cons	.0006165	.0006685	-6.82	0.000	.0000726	.0052329

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	16.66
			Prob > F	=	0.0001

		Linearized				
--	--	------------	--	--	--	--

orf40plus	Odds Ratio	Std. Err.	t	P> t	[95% Conf. Interval]	
cnonwpm	1.162059	.0427611	4.08	0.000	1.080705	1.249537
_cons	.005013	.003474	-7.64	0.000	.0012778	.0196661

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	45.20
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
read_comp_score	10.748	3.796447	6.72	0.000	5.354953	21.57244
_cons	.0003477	.0004079	-6.79	0.000	.0000344	.003516

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5194
Number of PSUs	=	202	Population size	=	1647514.2
			Design df	=	192
			F(1, 192)	=	39.45
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
list_comp_score	1.88242	.1895776	6.28	0.000	1.543294	2.296066
_cons	.0096865	.0036065	-12.45	0.000	.0046476	.0201884

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5199
Number of PSUs	=	202	Population size	=	1650842.1
			Design df	=	192
			F(1, 192)	=	48.26
			Prob > F	=	0.0000

orf40plus	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
pa_num_sound_score	1.278462	.0452086	6.95	0.000	1.192331	1.370814
_cons	.0102265	.003243	-14.45	0.000	.0054712	.0191148

Note: variances scaled within each stage to handle strata with a single sampling unit.

Survey: Logistic regression

Number of strata	=	10	Number of obs	=	5009
Number of PSUs	=	202	Population size	=	1592611.2

Design df = 192
 F(1, 192) = 70.09
 Prob > F = 0.0000

```

-----
      orf40plus |      Linearized
                | Odds Ratio  Std. Err.      t    P>|t|    [95% Conf. Interval]
-----+-----
pa_init_sound_score |  1.377634  .0527168    8.37  0.000    1.277482    1.485637
   _cons |  .0355215  .0058491   -20.27  0.000    .0256708    .0491523
-----

```

Note: variances scaled within each stage to handle strata with a single sampling unit.

Annex 3. EGRA 2012 Instrument

Subtest	2010	2011	2012	Notes
Letter Naming				Ability to say the names of the letters of the alphabet accurately, without hesitation and naturally. This is a timed test that assesses automaticity and fluency of letter name knowledge—measured in correct letters per minute.
Syllable Segmentation				The first of two measures of phonemic awareness. The subtest measures learners’ ability to hear a word and break it up into syllables. This is one of the first skills needed to understand how to read new words by decoding.
Initial Sound Identification				The second measure of phonemic awareness (the understanding that words are made of sounds). This subtest measures learners’ ability to hear and isolate the first sound in a word. This is generally considered a pre-reading skill and can be assessed in a variety of ways.
Syllable reading				This subtest is used because Chichewa is considered by Malawians to be syllabic in nature. This subtest asks students to read the most commonly occurring syllables in a particular language—measured by correct syllables read per minute.
Familiar Word Reading				Ability to read high-frequency words. This assesses whether students can process words quickly—measured by words read correctly per minute.
Nonsense Word Reading	(✓) ¹⁴			Ability to decipher “words” that follow the linguistic rules but do not actually exist in Chichewa. The nonsense words used for EGRA are truly made-up words. This subtest assesses a child’s ability to “decode” words fluently—measured by nonsense words read correctly per minute.
Oral Reading Fluency	(✓)			Ability to read a passage, approximately 60 words long—measured by words read correctly per minute.
Reading Comprehension	(✓)			Ability to answer comprehension questions based on the passage read—measured by percent correct out of five comprehension questions.
Listening Comprehension	(✓)			Ability to follow and understand a simple oral story. This assesses a child’s ability to concentrate and focus to understand a very simple story and answer both literal and inferential questions without the burden of reading the story. It is a more complete measure of learners’ ability to comprehend stories—measured by percent correct out of five comprehension questions.

¹⁴ In 2010, the EGRA instrument contained a decision rule intended to simplify the assessment administration for learners who were doing very poorly on the cognitively easier subtests. Learners scoring zero on the test of familiar word reading were not required to attempt the nonsense word reading, oral reading fluency, or reading comprehension sections. Instead, it was assumed that learners would likely score zero on these tests as well. However, due to some inconsistencies among enumerators in applying the decision rule in 2010, it was not used in 2011. Instead, all learners were tested on all subtests.

Annex 4. Benchmarks and Targets for EGRA, 2012–2017

Subtest	Measure	Suggested benchmark Std. 1	Suggested benchmark Std. 3	Possible	Std. 1 2017	Std. 3 2017	Std.1 2013	Std. 3 2013	Std. 1 2012	Std. 3 2012
Letter naming	correct letters per minute (clpm)	24+ [2.3%]	50	(unlimited)	60%	60%	25%	25%	10%	10%
Syllable segmentation	% correct	70[0%]	80[3.6%]	10/20/30...100	60%	70%	40%	50%	20%	30%
Initial sound identification	% correct	80 [0%]	90 [0%]	10/20/30...100	30%	30%	10%	10%	5%	5%
Syllable reading	correct syllables per minute (cspm)	30 [1.8%]	60 [9.7%]	(unlimited)	50%	60%	20%	30%	10%	20%
Familiar word reading	correct words per minute (cwpm)	20 [1.7%]	45[3.7%]	(unlimited)	50%	50%	25%	25%	10%	10%
Nonsense word reading	correct words per minute (cwpm)	15 [1.3%]	40[0.6%]	(unlimited)	30%	30%	10%	10%	5%	5%
Oral reading fluency	correct words per minute (cwpm)	20 [1.6%]	50[2.7%]	(unlimited)	50%	50%	25%	25%	10%	10%
Reading comprehension	% correct	40 [1.6%]	80 [0.8%]	20/40/60/80/100	30%	30%	10%	10%	5%	5%
Listening comprehension	% correct	60 [6%]	80 [24.1%]	20/40/60/80/100	60%	60%	40%	40%	30%	30%

Note: Benchmarks and targets agreed at the MoEST-convened EGRA Coordinating Committee Meeting, Crossroads Hotel, Lilongwe, October 27, 2011.

Annex 5. Technical Notes on National Study Sampling

Sampling for the Malawi National Study assessment in 2010 was set in two stages. The first stage was a simple random sample of schools stratified by division. At the second stage, students were selected in each school, stratified by grade or standard (Standard 2 & Standard 4).

Stage	Method	Sampling Unit	Stratification
1	SRS	School (99)	Division (6)
2	SRS	Student (~20)	Standard (2 nd , 4 th)

A similar weighting framework was used at midterm and endline, with the additional stratification variable of gender added in the second stage.

Stage	Method	Sampling Unit	Stratification
1	SRS	School (Mid=150, End=202)	Division (6)
2	SRS	Student (~20)	Standard (2 nd , 4 th), Gender

There are six Divisions in Malawi, and each Division has several Districts within, ranging from 4 Districts to 7 Districts per Division. Four Districts were left out of these studies in the original school selection process, as well as during the baseline and midterm report cycles because these Districts had a separate Reading Intervention study being administered. Data from these four Districts, Ntchisi (Central Eastern Division), Salima (Central Eastern Division), Mwanza (Southern Western Division), and Dedza (Central Western Division) were later added into the National Study dataset in order to allow the endline report to be entirely nationally representative.

As a result, schools from the four Reading Intervention Districts were sampled at a much higher rate than the rate for the remaining 30 Districts in Malawi. In order to adjust for this unequal probability of selection, schools in these four Districts were separated out into their own strata, and the weights were adjusted appropriately. The remaining schools in each Division are listed in the table below. The table is based on the baseline data (2010), using the most current EMIS available at the time (2009) for illustrative purposes. Midterm and endline data are similar, but use the next year's EMIS database to increase accuracy.

Strata	Schools in Division	Schools in Reading Intervention Districts of this Division	Modified Finite Population Correction (FPC)	Sample Size within Strata	Stage 1 Weight (FPC/Sampled)
Central Eastern Division	1014	276	738	7	105.4286
Central Western Division	1231	220	1011	11	91.90909
Northern Division	1316	0	1316	7	188
Shire Highlands Division	551	0	551	7	78.71429
Southern Eastern Division	822	0	822	8	102.75
Southern Western Division	749	49	700	7	100
Ntchisi District			139	16	8.5625
Salima District			137	27	8.176471
Mwanza District			220	10	22
Dedza District			49	8	6.125

Because such a small proportion of the schools in each Division was sampled, these weights are considerably larger than the Reading Intervention districts' weights, where a larger proportion of the population was sampled. Thus, when calculating the average reading scores across the entire country, the values of the low-weighted Reading Intervention Districts play a much smaller role in the calculation than the high-weighted Divisions. This ensures that the process of adding in Reading Intervention District schools into the National Study database will not incorrectly skew the results, but will offer a more correct snapshot of the reading levels in Malawi on a truly national scale.