USAID
FROM THE AMERICAN PEOPLE

Early Grade Reading Activity (EGRA) Impact Evaluation

Baseline Report

November 2013

This publication was made possible by the support of the American people through the United States Agency for International Development (USAID). It was reviewed by the USAID/Malawi Education team and prepared by Social Impact, Inc. (SI) through Dr. Joanne Capper, Dr. Pedro Carneiro, Dr. Benjamin Hermoso, Brenda Sinclair, Melissa Chiappetta, Molly Brune, and Aaron Foss.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACRONYMS

CAADP	Comprehensive Africa Agriculture Development Program
CADECOM	Catholic Development Commission
CARE	Cooperative for Assistance and Relief Everywhere
CDCS	Country Development and Cooperation Strategy
CPD	Continuous Professional Development
DAI	Development Alternatives, Inc.
DIAS	Department of Inspection and Advisory Services
EGRA	Early Grade Reading Activity
EMIS	Education Management Information Systems
EP	Evaluation Policy
EPA	Extension Planning Area
FtF	Feed the Future
FUM	Farmers Union of Malawi
GHI	Global Health Initiative
GoM	Government of Malawi
HT	Head Teacher
ICC	Inter-Cluster Correlation
IE	Impact Evaluation
IP	Implementing Partner
IKI	Invest in Knowledge Initiative
INVC	Integrating Nutrition in Value Chains
JHU-CCP	Johns Hopkins Bloomberg School of Public Health – Center of Communication Programs
MoEST	Ministry of Education, Science, and Technology
MOH	Ministry of Health
MOU	Memorandum of Understanding
MSCE	Malawi Schools Certificate of Excellence
MTPDS	Malawi Teacher Professional Development Support
NASFAM	National Smallholder Farmers Association of Malawi
NESP	National Education Sector Plan
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PEA	Primary Education Advisor
PEPFAR	President's Emergency Plan for AIDS Relief
PIRLS	Progress in International Reading and Literacy Study
PMI	President's Malaria Initiative
PSM	Propensity Score Matching
RA	Reading Assessment

RTI	Research Triangle Institute
SAT	Stanford Achievement Test
SI	Social Impact, Inc.
SSDI	Support for Service Delivery Integration
SOW	Scope of Work
US	United States
USAID	United States Agency for International Development
WASH	Water, Sanitation, and Hygiene

I. PROJECT BACKGROUND

INTRODUCTION

Although the Government of Malawi (GoM) has been working to improve the quality of education in its schools, the country still places well below average on worldwide education indices. According to the United Nations' 2013 Human Development Report Index, Malawi ranks 153rd out of 187 countries for the quality of its education, behind Zimbabwe, Congo, and Iraq.[1] The country has been quite successful in getting children into school but less so in ensuring quality education once learners are enrolled. After the abolition of school fees in 1994, primary school enrollment in Malawi went from 1.9 million to nearly 2.9 million. [2] While this result is laudable, multiple measures of learning outcomes indicate widespread underperformance of schools. For instance, the 2010 Early Grade Reading Assessment (EGRA) baseline study conducted by the United States Agency for International Development (USAID) through the Malawi Teacher Professional Development Support (MTPDS) Project shows that learners consistently and dramatically underperformed in measures of basic literacy. Specifically, test results show that 73 percent of Standard 2 youth could not read a single word of a prompted story, and 97 percent could not answer a single comprehension question correctly.[3]

In an effort to improve the quality of education in Malawi, USAID/Malawi funded the three-year MTPDS Project, which sought to improve educational outcomes by building Ministry of Education, Science, and Technology (MoEST) capacity, improving both teacher capacity to teach reading and school management and leadership. The project provided continuous professional development (CPD) to 34,000 lower primary school teachers across the country and higher-intensity literacy programming in seven districts. The result was that familiar-word

---

[1] Malik, Khalid. "2013 Human Development Report: The Rise of the South, Human Progress in a Diverse World," United Nations Development Program, 2013. <http://issuu.com/undp/docs/hdr_2013_en?mode=window>.

[2] Ampiah, Joseph; Byomugisha, Albert; Chimombo, Joseph; Kunje, Demis; Mikiko, Nishimura; Ogawa, Keiichi; Sawamura, Nobuhide; Sifuna, Daniel; and Yamada, Shoko, "A Comparative Analysis of Universal Primary Education Policy

in Ghana, Kenya, Malawi, and Uganda," CICE Hiroshima University, *Journal of International Cooperation in Education*, Vol.12 No.1, 2009. <CICE Hiroshima University, *Journal of International Cooperation in Education*, Vol.12 No.1>.

[3] USAID/Malawi. "Early Grade Reading Assessment: National Baseline Report," 2010. <www.eddataglobal.org/reading/index.cfm/Malawi%20National%20Baseline%20EGRA%202010.pdf?fuseaction=throwpub&ID=354>. The numbers from Standard 4 are 42 percent and 69 percent, respectively.

reading scores among children from beneficiary schools increased by 41.8 percent in just two years, and reading comprehension scores increased by 42.7 percent.[4]

However, despite these gains, the MTPDS end-of-program evaluation, conducted by USAID through Creative Associates International, Inc., showed that at the end of the project more than 70 percent of Standard 1 and 3 learners who benefited from the MTPDS intervention still could not answer a single comprehension question after reading a simple short story, and nearly 50 percent remained unable to read even a single word. As such, to support the MoEST in addressing the challenges associated with high levels of illiteracy, USAID/Malawi has designed a follow-on project that will build on the lessons learned and successes of MTPDS. The Early Grade Reading Activity (EGRA) Project, awarded to RTI (Research Triangle Institute) International in July 2013, is a multifaceted educational development approach intended to enable sustained literacy among children, promote a literate community, and help the country increase economic growth and reduce poverty. RTI began implementing the EGRA Project in August of 2013.

EARLY GRADE READING ACTIVITY BACKGROUND

The EGRA Project derives from a strong base of educational research that has revealed the importance of developing learners' fluent reading abilities by the end of Standard 3.[5] Learners who are not fluent readers by that time are unlikely to ever catch up, not only in reading, but also in all the other learning areas that require reading facility. Several factors are critical to ensuring that all learners develop reading fluency in Malawi, and EGRA, like MTPDS, is designed to address each of these factors, including teacher training, reading instruction, parents and community involvement, and political perceptions related to education policies. Specifically, the EGRA Project objectives are to:

- Improve the capacity of Standard 1 to 3 teachers to provide quality reading instruction to learners
- Improve the learning outcomes of Standard 1 to 3 learners
- Increase parental and community engagement to support learner reading
- Reduce repetition and dropout rates in the early grades by providing a quality learning environment

---

[4] USAID/Malawi. "Malawi Reading Intervention—Early Grade Reading Assessment: Final Assessment," 2013.

[5] USAID Malawi. (2010). "Early Grade Reading Assessment: National Baseline Report." <www.eddataglobal.org/reading/index.cfm/Malawi%20National%20Baseline%20EGRA%202010.pdf?fuseaction=throwpub&ID=354>. And: USAID Malawi. "Malawi National Early Grade Reading Midterm Assessment, 2011," June 2012.

To accomplish these objectives, the EGRA Project implementing partner (IP), RTI, will:

- Conduct teacher training, including practicums
- Provide scripted lesson plans to teachers
- Provide in-service teacher support and mentoring (or coaching)
- Provide rewards for high-performing teachers and schools
- Develop and distribute books, story cards, letter cards, and bookshelves
- Organize reading fairs and other events to engage parents, caregivers, and the community in learner learning
- Invite parents to participate in their learners' classrooms and/or become engaged in extracurricular activities
- Ensure a supportive policy environment by attending the Basic Education Technical Working Group and writing Memorandums of Understanding (MOUs) to gain necessary MoEST support and buy-in
- Facilitate the extension of school instructional time

Since EGRA has only recently been launched, it's possible that some components of the project may change over the course of the project's three-year timeframe.

COMPLEMENTARY DEVELOPMENT INTERVENTIONS

In keeping with USAID/Forward[6] and the country ownership and responsibility clauses of USAID's Global Education Strategy and the Paris Declaration, EGRA is aligned with USAID/Malawi's Country Development and Cooperation Strategy (CDCS), which aims to increase Malawi's self-sufficiency, economic development, and provision of quality essential services. In turn, it is expected that accomplishing these outcomes will, over time, decrease Malawi's dependence on foreign aid.

USAID/Malawi's CDCS focuses on the interconnectedness of development interventions and Malawi's ownership of the development process. Thus, USAID is implementing the EGRA Project alongside a Feed the Future (FtF) Project called Integrating Nutrition in Value Chains (INVC) and a Global Health Initiative (GHI) Project called Support for Service Delivery Integration (SSDI), both of which are also expected to affect the success of learners in school. Through this interconnected strategy, USAID/Malawi hopes to have a greater impact on

[6] USAID Forward, http://www.usaid.gov/usaidforward, accessed on August 23, 2013.

outcomes of interest (such as learning reading scores) than would otherwise be possible through one project alone. Additionally, USAID/Malawi officials hope to learn more about what types of interventions most affect learner reading outcomes. To accomplish these goals, USAID is implementing these three projects in three districts simultaneously; in other districts, the organization is implementing only one or two of the projects to allow for comparison of results between districts. Community engagement will feature in all of these efforts, especially in the implementation of the EGRA. Ultimately, USAID anticipates that its investments in early grade reading and the improved learning outcomes that are expected to result from increased literacy will contribute to an increase in communities' capacities for economic growth and development. The FtF and GHI components of this interconnected strategy are briefly described in the paragraphs below.

Rationale for FtF and GHI Projects

While overall poverty rates and kilocalorie intake in Malawi have improved in recent years, the country's food security and economic development are constrained by high rates of undernutrition, HIV/AIDS, malaria, underdeveloped markets, and low agricultural productivity. Factors contributing to these challenges include high population density and growth, a single rainy season per year, and significant barriers to regional and international trade. These include high freight costs, regional conflicts, limited infrastructure, and unpredictable economic policies. In addition, Malawi sees cholera outbreaks each year during the rainy season. Lack of access to improved water sources and proper sanitation systems contributes to these outbreaks.[7] The agricultural sector in Malawi employs 80 percent of the population, with the vast majority of smallholder farmers producing only enough food for household self-sufficiency, with little to no grain left to take to market.[8] The FtF INVC and GHI SSDI Projects were designed to address many of these challenges to Malawi's economic and social development.

Feed the Future: Background

Feed the Future is the US Government's contribution to the New Alliance for Food Security and Nutrition, a global collaboration of G8 and African countries; it aims to raise 50 million people out of poverty by 2022. With a focus on smallholder farmers, particularly women, FtF supports partner countries in developing their agricultural sectors to spur the economic growth that increases incomes and reduces hunger, poverty, and undernutrition. To address these concerns, the GoM has elevated agriculture and nutrition as key national policy priorities. In consultation with key stakeholders, government officials have developed an

---

[7] Feed the Future, Countries/Malawi, http://feedthefuture.gov/country/malawi, accessed on August 23, 2013.

[8] *Ibid.*

Agricultural Sector-Wide Approach and made targeted commitments under the Comprehensive Africa Agriculture Development Program (CAADP).

Specific FtF strategies aimed at achieving these goals include:

1. Supporting country-owned plans for results-based agriculture and nutrition programs and partnerships while promoting transparency, democracy, and good governance.
2. Promoting economic growth through improved linkages along the entire value chain—from farm to market—by improving connections to local, regional, and global markets; promoting sustainable intensification; and supporting an enabling environment for agricultural trade to minimize the impact of food price hikes.
3. Strengthening strategic coordination to mobilize and align the resources of diverse partners and stakeholders—including the private sector, civil society, and multilateral institutions.
4. Ensuring an integrated approach that accelerates inclusive agriculture-led growth, reduces undernutrition, and increases the impact of humanitarian food assistance.

In Malawi, the FtF INVC Project seeks to help vulnerable women, children, and family members—mostly smallholder farmers—escape hunger and poverty, prevent stunting, and reduce child mortality. The project aims to address barriers to food security by improving nutritional outcomes through changing communication behaviors, increasing access to food, and investing in high potential value chains (dairy and legumes—soybeans and groundnuts). The ultimate goal is to develop domestic and export markets and improve nutritional options for families.

USAID awarded the five-year contract for the INVC Project to Development Alternatives, Inc. (DAI) in April 2012, and DAI initiated its activities in May 2012. Since then, the INVC Project has already reached farmers in seven districts (Mchinji, Lilongwe, Dedza, Ntcheu, Balaka, Mangochi, and Machinga), and DAI plans to expand its efforts in these districts and others with the goal of reaching an estimated 275,000 households by 2017. The project is focused on smallholder farmers who own between one and three and a half acres of land in the central and southern regions of Malawi. INVC functions as a grant-making mechanism, ultimately supporting the efforts of several sub-grantees, including the Catholic Development Commission (CADECOM), the National Smallholder Farmers Association of Malawi (NASFAM), the Farmers Union of Malawi (FUM), and the Malawi Milk Producers Association.

Global Health Initiative: Background

While recent years have shown considerable improvement in various measures of health status in Malawi,[9] much remains to be done. Malawi's population is projected to more than triple in 30 years, with an annual growth rate of 3.2 percent. Although use of contraceptives is well established, total fertility remains high at six children per woman, with access to family planning limited for youth and those living in rural communities. Moreover, Malawi has one of the highest maternal mortality ratios in the world at 807/100,000.[10] Infant and under-five mortality rates remain high at 72/1,000 and 122/1,000 live births, respectively. Malaria, which is the leading cause of mortality and morbidity in Malawi, is also a serious issue, with approximately six million cases occurring each year. Finally, nutrition remains a major health and development problem, with nearly half of Malawi's children stunted, and approximately 12 percent of Malawi's15- to 49-year-olds HIV-positive.[11]

In May 2009, the US Government announced the Global Health Initiative (GHI), a six-year, comprehensive effort to reduce the burden of disease and promote healthy communities and families around the world. Historically, the US has funded health programs in Malawi that support specific disease or program efforts of the GoM, including HIV/AIDS under the President's Emergency Plan for AIDS Relief (PEPFAR); malaria under the President's Malaria Initiative (PMI); and other programs targeting health issues such as tuberculosis, family planning and reproductive health, maternal, neonatal and child health, and nutrition programs. These programs are now functioning under a single, coordinated health strategy—GHI—with the primary goal being to increase access to quality health care to foster a healthier Malawian populace able to participate in the country's economic development. USAID/Malawi, in collaboration with the GoM, civil society, and other partners, has developed a strategy to focus GHI in Malawi on the following priorities: 1) the provision of quality care to reduce maternal, neonatal, and child mortality and morbidity, 2) the reduction of unintended pregnancies, and 3) the reduction of new HIV infections.[12]

The specific GHI Project being implemented in coordination with EGRA and INVC Projects is the SSDI Project, which aims to integrate and strengthen healthcare services for families throughout Malawi. Initially, USAID designed the SSDI Project to include three cooperative

---

[9] Demographic and Health Survey, 2004.

[10] Demographic and Health Survey 2004; Currently three different maternal rates for Malawi are in circulation, including 510/100,000 and 1100/100,000. The statistic 807/100,000 is the approved figure used by the MOH.

[11] http://www.ghi.gov/whereWeWork/docs/MalawiStrategy.pdf

[12] http://www.ghi.gov/country/malawi/documents/160474.htm

agreements and provide approximately $100 million in total USAID funding over a five-year period (2011–2016).[13]

Different aspects of the project are being implemented by different IPs. Specifically, the Johns Hopkins Bloomberg School of Public Health – Center of Communication Programs (JHU-CCP) (in partnership with Save the Children-Malawi) is implementing the $16 million project that focuses on multimedia health messaging in the districts of Mangochi, Machinga, Phalombe, Nsanja, Chikhawa, Kasungu, Dowa Salima, Lilongwe, Chitipa, and Karonga. Abt Associates is carrying out a five-year, $10 million project to support Malawi's Ministry of Health (MOH) in improving the nation's health system by building capacity and financial efficiency within the MOH. Jhpiego (an affiliate of Johns Hopkins University) is leading a consortium partnership of Save the Children, CARE, and Plan-International to implement an $8.5 million project that aims to reduce fertility and population growth, lower the risk of AIDS, and reduce maternal and infant mortality rates from October 2011 to September 2016.[14] This includes mobilizing populations to seek services from Malawi's health system and to scale up access to these resources.

[13] https://www.devex.com/en/projects/malawi-support-for-integrated-service-delivery

[14] http://www.planusa.org/content2669431

II. EVALUATION PURPOSE AND QUESTIONS

In April 2013, USAID/Malawi contracted with Social Impact, Inc. (SI), an Arlington, Virginia–based development consulting firm to conduct four tasks:

1. A baseline assessment for an impact evaluation (IE) of the USAID/Malawi EGRA Project on addressing learner reading outcomes in Standards 2 and 4;
2. A survey of a sub-sample of the households of the Standard 2 and 4 learners selected for the impact evaluation (IE) sample;
3. A national reading assessment (RA) of Standard 1 and 3 learners; and
4. A final impact evaluation of the EGRA Project and the CDCS hypothesis that learner reading outcomes will improve even more from the implementation of EGRA, INVC, and SSDI combined than they will from EGRA alone.

Together, Tasks 1, 2, and 4 will provide a rich, multifaceted impact evaluation of USAID/Malawi's EGRA. Task 3, a national-level reading assessment, to be carried out in 2014 and 2016, will provide valuable data on learner reading performance for all stakeholders and will contribute to the trend within Malawi's education system toward greater accountability and evidence-based decision-making.

EVALUATION AND ASSESSMENT PURPOSES

The purpose of the EGRA impact evaluation is:

1. To measure the effect of EGRA on learner reading outcomes (versus a comparison group);
2. To test the hypothesis that integrating USAID interventions in education, agriculture, health, and community strengthening in the same communities results in increased learning among primary school youth, as described in the USAID/Malawi CDCS. This includes measuring:
    a. How integration of USAID programming across sectors (education, health, and agriculture) in the same geographic areas impacts learner reading outcomes; and

b. How community strengthening through capacity-building of local institutions and promotion of citizen participation impacts the sustainability of reading interventions.

The purpose of the national reading assessment during years 2 and 4 is to allow the MoEST, USAID, and other stakeholders to monitor education in Malawi over time using nationally representative data. The results from both studies will be used by all of these parties to improve the effectiveness and efficiency of future education interventions.

EVALUATION AND ASSESSMENT QUESTIONS

This evaluation and assessment has been divided into four discrete tasks, each task with its own set of questions, outlined below. The answers to these questions represent the key indicators for this project. This baseline report addresses primarily the questions related to Tasks 1 and 2, although the information collected to date and presented in this report lays the groundwork for answering the questions outlined in Task 4. Questions under Task 3, the national reading assessment, are not addressed here, but will be addressed in 2014 and 2016.

The tasks described below will be carried out over a five-year period, as specified in the scope of work (SOW) for this impact evaluation, and are presented in Table 1.

Table 1. Timeline of Assessment and Evaluation Tasks

	Baseline		Midpoint		Endline
Tasks	**May 2013**	**May 2014**	**May 2015**	**May 2016**	**May 2017**
1. Evaluation of the USAID/ Malawi EGRA on Standard 2 and 4 learners reading outcomes	X		X		X
2. Household survey of sub-sampled Standard 2 and 4 learners	X		X		X
3. National reading assessment for Standard 1 and 3 learners		X		X	
4. Final impact evaluation of EGRA and CDCS hypotheses					X

Task 1: Evaluation of the Effects of the EGRA on Standard 2 and 4 Learner Reading Outcomes

Four instruments were used to help answer the evaluation questions for Task 1, described below. The instruments were for the reading assessment, which was given in two languages (Chichewa and English), though English was given only in three districts.

1. What proportion of primary school learners is able to read with comprehension, according to Malawi's curricular goals, by the end of lower primary school (Standard 4)?
    a. What proportion of learners, who by the end of Standard 4 are able to read grade-level text, as measured by the number of correct words per minute?
    b. What proportion of learners, who by the end of Standard 4 are able to answer comprehension questions after reading grade-level text, as measured by the number of correct comprehension questions answered correctly?
2. What proportion of learners, who by the end of Standard 2, demonstrate that they can read and understand the meaning of grade-level text?

a. What proportion of learners, who by the end of Standard 2 are able to read grade-level text, as measured by the number of correct words per minute?
b. What proportion of learners, who by the end of Standard 2 are able to answer comprehension questions after reading grade-level text, as measured by the number of correct comprehension questions answered correctly?

Task 2: Household Survey of Standard 2 and 4 Learners

In the first, third, and fifth years of this evaluation, SI is tasked with conducting a survey of the households of the learners selected to take part in the reading assessment described above. The data from this household survey (HHS) will be used to help understand the influence that factors outside of the school, including household- and community-level influences, may have on reading outcomes. The HHS also includes questions intended to capture the influence at community and household levels of other USAID interventions operating under GHI and FtF. The questions of interest are:

1. What household and community factors correlate with learner reading outcomes?
2. What level of household and community resources is dedicated to schooling overall and reading specifically?
3. How have health and agricultural interventions at the household and community levels influenced schooling and reading outcomes?
4. What factors at the household and community levels correlate with changes in the rates of learner grade repetition and early departure from school, and are girls and boys treated differently in the household in ways that influence academic achievement?

Task 3: National Reading Assessment for Standard 1 and 3 Learners

The National Reading Assessment will use the same EGRA reading assessment (RA) instrument as was used in Task 1 to examine how primary learners in Standards 1 and 3 are progressing toward reaching MoEST reading EGRA Coordinating Committee benchmarks; it will collect data to answer the following questions:

1. What proportion of primary school learners is at the Standard 1 benchmark for reading skills?
    a. What is the breakdown of learners grouped by subdivisions and progress toward attaining EGRA Coordinating Committee benchmarks in Standard 1?

b. What is the relationship of Standard 1 reading skill acquisition to additional factors that relate to or predict achievement, including classroom size, for Standard 1?

2. What proportion of primary school learners is at the Standard 3 benchmark for reading skills?
   a. What is the breakdown of learners grouped by subdivisions and progress toward attaining EGRA Coordinating Committee benchmarks in Standard 3?
   b. What is the relationship of Standard 3 reading skill acquisition to additional factors that relate to or predict achievement, including classroom size, for Standard 3?

Again, these questions will be addressed in 2014 and 2016.

Task 4: Final Impact Evaluation of EGRA and CDCS Hypotheses

SI will draw upon data collected under all of the preceding tasks, including data gathered during the national reading assessments (Task 3), to determine the overall impact of the EGRA on schooling and reading outcomes, as well as the influence of the combined GHI and FtF interventions. The key questions to be answered under this task include:

1. What is the USAID/Malawi EGRA's impact on children's reading abilities (disaggregated by sex) in terms of the following:
   a. Impact of level of effort of reading instructor on learners' reading abilities
   b. Effect of extracurricular reading activities
   c. Effect of time-on-task in improving reading outcomes
2. Which components have the largest effect and what is the relative cost-effectiveness of these various components?
3. How do teachers' classroom behavior and practices impact the ability of children to read?
   a. How did the level of coaching impact teacher behavior and learner reading outcomes?
4. How does the level of integration with other USAID/Malawi FtF and GHI programs, and other related development program interventions in the target districts, impact the reading outcomes of learners?
   a. What interactions can be identified with other major USAID/Malawi Mission interventions in agriculture and health?

b. What other multiplier effects have been identified over the life of EGRA?

c. What are the key external factors that were found to have a multiplier effect, i.e. early childhood development (ECD) attendance, participation in school feeding, change in water, sanitation, and hygiene (WASH) behaviors, access to a secondary school, etc.?

d. How does the provision of non-cash incentives to performing teachers and schools translate into changes in children's reading abilities?

5. What secondary effects can be attributed to EGRA?
   a. Impact on repetition rate
   b. Impact on dropout rate
   c. Impact on school completion, particularly for girls and learners with disabilities
6. What is the effect on the effectiveness and sustainability of education and learning outcomes of USAID/Malawi's investments in institutional capacity-building and community engagement?

These questions will be addressed in the 2017 final evaluation report for this contract.

III. METHODS AND APPROACH

IMPACT EVALUATION OVERVIEW

Answering causal questions such as *What is USAID/Malawi EGRA's impact on children's reading abilities?* and attributing change to a specific program require ruling out alternative possible causes for changes in outcomes. Recognizing the multitude of possible alternative causes for changes in the EGRA Project's key outcomes (learner reading ability, dropouts, and repetition rates) between government programs, other donor initiatives, and political and/or economic development, the USAID 2011 Evaluation Policy (EP) requires that impact evaluations use a carefully selected comparison group to rule out possible alternative causes for key outcomes through estimating the counterfactual, or the level of change in project participant outcomes expected in the absence of the project. By comparing project participants with a comparison group, it is possible to "subtract away" the contextual changes (or those caused by other interventions or natural changes such as time) that affect both project participants and non–project participants (the comparison group). If project participation is the only substantive difference between participants and the comparison

group, then any differences in outcomes between the two groups can be attributed to the project.

In order to test possible complementary or multiplier effects of the EGRA, FtF, and GHI Projects, the EGRA Project and associated impact evaluation is being implemented in and evaluated for four distinct treatment levels:

- **Treatment Level 1**—Three districts where current or future FtF and GHI interventions are planned, along with the EGRA intervention, providing the opportunity to evaluate the impact of a thoroughly integrated development approach on early-grade reading outcomes. This baseline assessment and the supplemental midterm and endline assessments already include or will include collecting data for beneficiaries from mission-integrated districts at this level. These districts include Lilongwe Rural, Balaka, and Machinga.
- **Treatment Level 2**—The district where EGRA will overlap with only the GHI intervention. This will serve as a test ground for the hypothesis that synergies between education and health initiatives catalyze changes that are greater than the sum of their parts. Data for this treatment level was collected from Salima, and follow-up assessments will also include data collection in this district.
- **Treatment Level 3**—The district where EGRA will overlap with only the FtF intervention. This will serve as a test ground for the development hypothesis that synergies between education and agricultural livelihood and nutrition initiatives catalyze changes that are greater than the sum of their parts. Data for this treatment level was collected from Ntcheu, and follow-up assessments will also include data collection in this district.
- **Treatment Level 4**—Districts that only receive the EGRA initiative. These districts will be used to test the EGRA theory of change that integrated education support leads to improved literacy and general education outcomes. Data for this treatment level was collected from Thyolo, Mzimba North, Blantyre Rural, Zomba Rural, and Ntchisi, and follow-up assessments will also include data collection in these districts.

SI will compare learner reading scores from each of these four groups with a comparison group located within the same district(s) as the beneficiary group to determine how effective each type of treatment is. This comparison group will not receive an intervention of any type for at least one year. However, after that point, RTI will phase in support to these areas to ensure equal access to support through the project. The study will also include comparisons of the types of treatment with one another, although it will be more difficult to isolate the effects of the different types of treatment from district-level effects in such a comparison since each of the treatment types is being rolled out in separate districts (this is explained in

more detail in the treatment assignment section of this report). The evaluation team plans not to follow the same set of learners longitudinally but, instead, to follow schools longitudinally given the need to assess different standards in different years.

To collect the necessary data, SI used the data collection tools listed in Table 2 and described more fully in a later section.

Table 2. Data Collection Instruments Used on Each Task

	Task			
Instrument	**1**	**2**	**3**	**4**
Reading Assessment (Chichewa and English)	X		X	X
Learner Interview Protocol	X		X	X
Teacher Interview Protocol	X		X	X
Head Teacher Interview Protocol	X		X	X
Classroom Observation Protocol	X		X	X
School Environment Observation Protocol	X		X	X
Household Survey		X		X

SAMPLE SIZE

Original Sample Size Determination

The sample size for Tasks 1, 2, and 4 was determined by identifying the necessary number of schools that SI would need to visit to allow evaluators to compare reading outcomes between treatment levels and treatment types and make statistically valid conclusions about the effects of different treatment. Prior to collecting baseline survey data, SI conducted power calculations using MTPDS endline data and found that in order to detect effects of around 6.4 points (0.37 to 0.38 standard deviations) between learners from different treatment types for the number of correct letter sounds per minute (which has a mean of 9.42 and standard deviation of 17.3), the sample needed to include at least 40 schools within each treatment

level and at least 10 learners per school.[15] Likewise, evaluators determined that they would need to include 40 control schools per treatment level. However, rather than selecting only 10 learners per school, SI believed it would be possible to assess 15 learners per school in the same amount of time (one school per day per team of three enumerators),[16] and, therefore, increased the number of learners assessed per standard per school to 15. This was done to help mitigate risks associated with attrition in the team's longitudinal design due to dropouts or lack of full treatment due to excessive absenteeism. The GoM reports both absenteeism and dropouts rates averaging more than 25 percent in the lower standards, and cohort analyses indicate that only 53.1 percent of boys and 45 percent of girls who initially enrolled in primary school made it to Standard 8 in 2010. For this reason, SI thought it important to oversample learners, with the hope that doing so would mean that at least 10 learners per standard per school would remain in the sample throughout the life of the project. Power analysis conducted prior to the baseline survey suggested that as long as this number of learners remained at their current school, it would still be possible to detect statistically significant differences from impact evaluation data.

To further address potential attrition and contamination challenges, when selecting the sample of learners from each sample school (the process for selecting sample schools is described below), SI enumerators asked teachers a few brief questions about the randomly selected learners chosen for the sample to determine whether they constituted appropriate sample units. These questions included the following: 1) *Has the learner been enrolled at this school since the beginning of the school year?* and 2) *About what percent of school days would you say the learner attends classes?* If the teacher did not know, learners were also asked. If the answers to these questions were 1) "Yes" and 2) "At least about 75 percent of the time," the learner was included in the sample. If not, to avoid contaminating the sample and likely biasing the observed treatment effect, enumerators did not administer the assessment to that learner and instead chose a randomly selected alternate learner.

Based on the above information, SI's sample selection framework included the following:

- 40 schools with each different treatment type (160 total treatment schools)
- 160 control schools
- 15 learners per standard per school

---

[15] The minimum detectable effect sizes for other tests are: 5.5 points in correct syllables per minute (mean 5.9 and standard deviation 14.9), 3.37 points in correct words per minute (mean 3.58 and standard deviation 9.11), 2.25 points in correct invented words per minute (mean 2.22 and standard deviation 6.09), 3.02 points in oral fluency (mean 2.79 and standard deviation 8.17).

[16] This belief was based on recommendations made in the EGRA toolkit and also the recollection of MTPDS enumerators.

- 320 total schools
- 9,600 total learners

Sample Size Revisions

Using the now available baseline data, the evaluation team reworked the power analyses to assess the accuracy of the initial power calculations. As a result, the team found that the original power calculations were quite conservative and, therefore, evaluators may be able to detect smaller effect sizes than originally expected for several types of comparisons.

The new power calculations are based on clustering at the zonal level, since zone is the unit of treatment in many cases. Conducting power analysis at the zonal level provides for the most conservative power calculations, as there is likely to be less inter-cluster correlation (ICC) within zones than within schools, meaning learners from the same school are more likely to score more similarly than learners across schools in the same zone.[17] As such, when comparing schools or learners, evaluators should be able to detect even smaller differences. SI has included both zone- and school-level calculations for learner oral reading fluency test scores.

Due to sampling and data collection challenges, the sample is somewhat smaller than was originally called for in the design. There are several reasons for this. The first is that during sampling, the evaluation team realized that there were only 28 possible control schools within Level 3. Thus, the remaining 13 control schools for this level were moved to Level 4 to allow for more *pure* control schools, as described later. Second, the team found that the GHI SSDI Project had already spread across the districts it was working in before baseline data could be collected, contaminating possible control zones and schools. Because of these issues, rather than include 40 schools per treatment and 40 schools per control level, the evaluation team decided to move the control schools from Level 2 treatment to Level 4 since Levels 3 and 4 were the only levels where pure control schools were feasible (given contamination by the GHI Project). This was done to avoid comparing EGRA + GHI to GHI only, which would essentially be answering the same question as that of Level 4. Finally, several challenges arose during baseline data collection (i.e., the reading assessment taking longer than expected, school letting out early or starting late, and schools lost due to data collection supervisors leaving their teams, etc.—all of which are discussed more fully in the Limitations section). Therefore, the final sample size is:

- 34–40 schools per treatment type (150 total treatment schools)

---

[17] ICC is important because the higher it is (meaning that units or learners within a cluster—schools or zones—are more similar than units between clusters), the more clusters are needed; the lower it is, the more units are needed to allow for detection of statistically significant differences in outcomes of interest.

- 160 control schools
- 14.37 learners per standard per school, on average
- 310 total schools
- 8,910 total learners

Table 3 shows the number of schools and the average number of learners per school per treatment type.

Table 3. Number of Schools and Learners in Each Treatment Group

Level	Treatment	Schools	Average Per School	Total Learners
L1	Treatment	40	27.5	1,101
L1	Control	40	30	1,198
L2	Treatment	32	28.4	908
L3	Treatment	39	30	1,171
L3	Control	27	27.7	748
L4	Treatment	40	26.8	1,071
L4	Control	92	29.5	2,713

For Level 1 treatment, evaluators were able to collect data for 18 treatment zones (40 schools, 1,101 learners; on average, 61 learners per zone, 28 learners per school) and 13 control zones (40 schools and 1198 learners; on average, 30 learners per zone and 30 learners per school). With this sample, if the number of respondents is maintained, the evaluation team should be able to detect statistically significant changes (improvements or reductions) in oral reading fluency test scores over time of approximately 5 completed words per minute (cwpm) at the zone level and 4.1 cwpm at the school level.[18]

[18] The intra-cluster correlation in oral reading fluency test scores for Level 1 is 0.0673 for treatment areas and 0.0368 for control areas (0.0934 and 0.0642, respectively, if the cluster is the school). The standard deviations for these scores in treatment and control areas are 19.07 and 17.94, respectively. Therefore, if all sample respondents are maintained, evaluators should be able to detect an effect size of 5.0467 (about 25–30 percent

Level 2 treatment was not considered in the "by treatment type" power analysis because the evaluation team did not designate any Level 2 control schools and instead will only compare Level 2 treatment schools with pure control schools from Levels 3 and 4 (as described above). Power analysis for this comparison is included in Annex 8 at the end of this report.

For Level 3, evaluators collected data for 11 treatment zones (39 schools and 1,171 learners; on average, 106 learners per zone and 30 learners per school) and four control zones (27 schools and 748 learners; on average, 187 learners per zone and 28 learners per school). With this sample, if the number of respondents is maintained, the evaluation team should be able to detect statistically significant changes (improvements or reductions) in oral reading fluency test scores over time of approximately 5.5 completed words per minute (cwpm) at the zone level and 4.6 cwpm at the school level for Level 3. [19]

For Level 4, evaluators collected data for 27 treatment zones (40 schools and 1,071 learners; on average, 40 learners per zone and 26.8 learners per school) and 41 control zones (92 schools and 2,713 learners; on average, 66 learners per zone and 30 learners per school). With this sample, if the number of respondents is maintained, the evaluation team should be able to detect statistically significant changes (improvements or reductions) in oral reading fluency test scores over time of approximately 4 completed words per minute (cwpm) at the zone level and 4.1 cwpm at the school level for Level 3.[20]

These minimally detectable effect sizes tend to be even smaller when comparing each of the treatment levels with the pure controls (those where no other intervention exists, which are control zones and schools in Level 4). More information about this is included in Annex 8.

---

of a standard deviation; the effect size is 4.0940 if the cluster is the school) when comparing Level 1 treatment to Level 1 controls, with 80 percent power and a 5 percent level of significance.

[19] The intra-cluster correlation in oral reading fluency test scores for Level 3 is 0.0890 for treatment areas and 0.0080 for control areas (0.0978 and 0.0880, respectively, if the cluster is the school). The standard deviations for these scores in treatment and control areas are 17.45 and 16.62, respectively. Therefore, with the sample size used in this study, at endline, evaluators should be able to detect an effect size of 5.5158 (about 30–35 percent of a standard deviation; the effect size is 4.6036 if the cluster is the school) when comparing Level 3 treatment to Level 3 controls, with 80 percent power and a 5 percent level of significance.

[20] The intra-cluster correlation in oral reading fluency test scores is 0.0707 for treatment areas and 0.0451 for control areas (0.0706 and 0.0601, respectively, if the cluster is the school). The standard deviations for these scores in treatment and control areas are 17.63 and 17.66, respectively. Therefore, with the sample size used in this study, evaluators should be able to detect an effect size of 4.0278 (about 20–25 percent of a standard deviation; the effect size is 4.1426 if the cluster is the school) when comparing Level 4 treatment to Level 4 controls at endline, with 80 percent power and a 5 percent level of significance.

TREATMENT ASSIGNMENT

As described in the sample size section above, in consultation with USAID, the evaluation team decided to include treatment and comparison schools within the same districts rather than having some districts be treatment districts and some control, as was specified in the original SOW for this evaluation. This was done to increase the likelihood that the evaluation will be able to attribute identified changes to EGRA, FtF, and/or GHI rather than to non-project-related differences between districts. One thing the evaluation team was not able to do, however, was to include beneficiaries from each of the treatment types (Level 1, 2, 3, and 4) within each district. As such, it will be difficult for evaluators to compare changes in reading outcomes across different types of treatment. While evaluators will be able to determine whether EGRA is better than no EGRA and whether EGRA plus FtF and GHI is better than no treatment, it will not be possible to say whether EGRA plus FtF and GHI is better than EGRA alone. Designing an Impact Evaluation (IE) that allowed for this would require randomizing treatment within districts to ensure there were some schools with each different type of treatment in each district (since schools within a district are likely to have a lot in common). This is not possible since the FtF and GHI interventions have already begun, and especially because of the near-universal coverage of the GHI interventions in the districts where it is operating. The inability to randomly assign treatment within a district means that the likelihood of selection bias in these samples is high. To help address this concern, SI's evaluation design includes the use of Propensity Score Matching (PSM, see Annex 1) to match learners in treatment schools with learners in control schools using learner-, household-, and school-level characteristics. This method will help to mitigate the risk of selection bias and produce the best matches possible between the treatment and control groups.

Level 1

Level 1 districts, namely Lilongwe Rural, Balaka and Machinga, contain zones and schools that will be exposed to all three interventions: EGRA, FtF, and GHI. GHI is expected to have universal or near-universal coverage in all of these districts by the end of the five-year evaluation period, making it impossible to identify schools within Level 1 districts to serve as "pure" control schools (ones that would receive none of the interventions). FtF is implemented in certain Extension Planning Areas (EPAs), which represent geographic areas within each district. At least some aspects of the EGRA intervention (teacher training, for example) will be implemented by educational zone, which is a grouping of adjacent schools determined by the MoEST. Since the EPAs and educational zones are determined by different agencies within the GoM, they do not map neatly onto one other, meaning that some zones may fall within two or more EPAs, and EPAs overlap multiple educational zones. Therefore, in order to designate treatment and control zones as well as schools for the Level 1 districts, SI worked with the implementers of FtF's INVC Project to determine in which EPAs FtF is being

implemented and in which EPAs it will be implemented within the next five years. Using the geospatial coordinates for the schools in each district, based on a 2008 GoM National Statistical Office dataset, SI was then able to determine which schools fell inside of the EPAs where FtF was or would be implemented. With this information and with assistance from the Department of Inspection and Advisory Services (DIAS), the Education Management Information Systems (EMIS) office, and the District Education Offices in Lilongwe, Balaka, and Machinga, the SI evaluation team was able to identify some educational zones where no schools were expected to be exposed to FtF and others where all schools were expected to be exposed to FtF, or, in other words, zones that fell completely inside or completely outside of the EPAs where FtF was or would be operating. This process resulted in 18 educational zones outside of the FtF treatment areas, which could be called FtF-control zones, and 21 educational zones inside the FtF treatment areas, or FtF-treatment zones. Combined, these 39 zones contained 342 schools, which became the sample frame for Level 1. SI randomly selected 40 schools within the 18 designated FtF-treatment areas to serve as treatment schools, meaning that they would receive all three interventions (EGRA, FtF, and GHI), and 40 schools within the 21 designated FtF-control zones to serve as control schools, meaning that they would receive only GHI intervention, not EGRA or FtF.

Level 2

The Level 2 district, Salima, has 10 zones, all of which will be exposed to the GHI intervention by the end of the evaluation period. As such, there was no possibility to select control zones within Level 2. For this reason, four schools from each of the 10 zones were randomly selected as treatment schools for inclusion in the sample, for a total of 40 treatment schools in Level 2. The 40 control schools originally intended for selection within Level 2 were reallocated to the Level 4 districts, so that they could serve as "pure" control schools (those not exposed to any of the three interventions). Doing so will increase the size of the pool of control schools available to match against, increasing the likelihood of securing high-quality matches during the analysis phase.

Level 3

The Level 3 district, Ntcheu, has 18 zones. Using a process similar to the one described above for Level 1, SI was able to determine which educational zones fell inside of and outside of the EPAs in which FtF was being or would be implemented, resulting in three FtF-control zones and nine FtF-treatment zones. (Six zones contained both FtF and non-FtF schools and so were omitted from the sample.) These 12 zones have 96 schools, and these schools formed the Level 3 sample frame. From the nine FtF-treatment zones, 40 treatment schools were randomly selected for the Level 3 treatment sample. The three FtF-control zones, however, contained only 27 schools, all of which were selected for inclusion in the control group. The remaining 13 schools, which would be needed in order to reach the total of 40 control

schools, were reallocated to the Level 4 districts in order to maintain the intended sample size. As noted above with regard to Level 2, doing so increases the total size of the pool of "pure" control schools, increasing the likelihood of identifying strong matches using principal component matching.

Level 4

Level 4 includes the districts of Mzimba North, Blantyre Rural, Zomba, Thyolo, and Ntchisi, which together include 79 zones. From these 79 zones, 41 control zones and 38 treatment zones were selected based on a pair-wise random sampling technique using learner-teacher ratio for each school.[21] All zones were ranked, from highest average learner-teacher ratio to lowest. Zones that were adjacent to each other on the ranked list were then formed into pairs, so that #1 and #2 formed one pair, #3 and #4 formed another pair, and so on. The zones were then assigned to the treatment and control groups randomly: for example, the #1 school was assigned to the treatment group and the #2 school to control, with the decision whether #1 or #2 was assigned to the treatment group being random. This method maintains the benefits of random selection while increasing the likelihood that the control group and treatment group will be comparable, at least according to learner-teacher ratio.

From the 38 treatment zones, SI selected 40 EGRA-only treatment schools. From among the 41 control zones, SI selected 92 total control schools, which would not be exposed to any of the interventions. Forty of these 92 control schools were originally intended as control schools in Level 4, 40 were reallocated from Level 2 (as described above), and 12 were reallocated from Level 3 to make up for the smaller number of control schools available in Level 3.

Phased-Treatment Design

SI recognizes USAID and the MoEST's desire to eventually roll the program out to all schools in all treatment districts. Therefore, SI plans to work closely with USAID and RTI to implement a phased-treatment plan for the schools currently designated as control schools. This plan will allow for some of the control schools at each treatment level to be added to the list of EGRA treatment schools in each year of the three-year project until all schools in the designated districts are receiving the EGRA intervention. Phasing in the treatment of control schools over time—rather than providing them all with treatment in the second year of the project, for instance—will allow the evaluation team to compare the effects of two years of the EGRA intervention with the effects of one year of the intervention. However, the phased

[21] Learner-to-teacher ratio was one of the only factors for which the evaluation team had evidence by zone prior to baseline data collection. This is the reason that the team chose to use this particular variable for pair-wise sampling.

design does limit the evaluation team's ability to measure the effects of more than two years of EGRA treatment.

TESTING IN ENGLISH

Though English and Chichewa are the official languages of Malawi, learners in the early grades often are taught in Chichewa until they reach Standard 5, at which point all learners transition to English-only instruction. To date, instructional materials for Standards 1 through 4 have been produced only in Chichewa and in English as a second language. Learners begin learning English as a second language in Standard 1. To better understand how this additional language and the process of transitioning affect reading performance in English, SI also conducted the RA with a subset of learners in English.

Overall, 8,910 learners were assessed in Chichewa (as described above) and 997 of those were also assessed in English. Though the original sample for learners assessed in English was intended to be larger, some records were lost among this group for the same reasons described in the sample size section above. However, since the sample of learners assessed in English was never intended to be statistically significant for this study, the evaluation team was not too concerned with this reduction in records. Despite the lack of statistical significance among these groups, these data—especially when combined with the accompanying qualitative data collected from the learners, teachers, head teachers, and households—still provide valuable information to stakeholders about the influence that learning in Chichewa and English may have on reading achievement .

INSTRUMENTS

Ten instruments were developed, adapted, and/or adopted for use in evaluating reading achievement in the sampled schools, as well as to obtain information on a number of potentially explanatory variables. Two instruments are reading assessments, one for each of the following languages: Chichewa and English. Five instruments were intended to capture explanatory data on the school site and from learners, teachers, head teachers, and households.

EGRA Chichewa and English Reading Assessment Instruments

The primary source of outcome data (reading performance) is a standardized reading assessment (RA) instrument of Standard 2 and Standard 4 learners' reading abilities in Project Years 1, 3, and 5 (2013, 2015, and 2017, respectively). This instrument was developed and

used in Chichewa and English in the MTPDS project and was developed based on guidelines[22] provided by the originators of the EGRA model, RTI International, and, as such, has been widely used in Malawi. The instrument also has been validated and used in a variety of contexts and languages throughout the world to measure the reading ability of young learners. It includes the following subtasks:

1. Letter name knowledge
2. Syllable segmentation
3. Initial sound identification
4. Syllable reading
5. Familiar-word reading
6. Unfamiliar (non-word) reading
7. Oral reading fluency
8. Reading comprehension
9. Listening comprehension

Other In-School Data Collection Instruments

In addition to the EGRA reading assessments developed for Chichewa, and English, SI developed and used five qualitative instruments, described below, to capture additional relevant information during in-school data collection. All of these instruments were reviewed and revised based on input from USAID/Malawi and the DIAS, and were further revised following pilot testing conducted during the enumerator and supervisor training sessions, described below.

- A **classroom observation protocol** was used in one Standard 2 and one Standard 4 classroom at each sampled school. It also will be used during the mid- and end-point evaluations conducted in 2015 and 2017. This instrument was used to gather data regarding teacher and learner behavior in the classroom that would indicate the extent to which the EGRA intervention is being implemented in the classroom; it also includes generally recognized good teaching practices across all subjects, as well as behavior representing the equal treatment of all learners, the number of classroom resources, and the extent to which the classroom is a supportive learning

[22] RTI International and International Rescue Committee (2011). *Guidance Notes for Planning and Implementing EGRA*, Research Triangle, NC.

environment. The classroom observations were conducted by enumerators, such as teacher trainers, who had sound knowledge of good teaching practice.

- A **head teacher interview protocol** was administered to head teachers by an enumerator and contained a range of relevant questions, including involvement of the community, amount of EGRA training received, and EGRA coaching or supervision of teachers.
- A **teacher interview protocol** was administered to the teachers from whose classes learners were selected for the reading assessment; it included topics such as years of teaching experience, qualifications, the amount and quality of the EGRA training and coaching received, learner attendance, learner repetitions, and perceived barriers to learning.
- A **learner interview protocol was** administered to each learner participating in the EGRA assessment, with questions that address factors such as learners' attitudes toward school, reasons for not attending school regularly, how often and what they eat, and whether they are read to at home.
- A **school climate checklist** was used to rate factors such as cleanliness of the school and classrooms, whether there are latrines and how clean they are, and other items related to the physical and environmental condition of the school, such as instances of bullying.

The final versions of these instruments are included as Annexes 2 through 6 of this report.

Household Survey

The final instrument used in this study was a household survey, which was designed by the evaluation team using validated modules from USAID's FtF population survey, World Bank surveys, and several learner assessment surveys given in Malawi. The purpose of the survey was to collect information on household and community factors that may affect learner reading outcomes. This survey is where the evaluation team asked questions about FtF and GHI Projects as well as household resources and support of learner education. The survey was administered to all of the households for learners assessed during the reading assessment, though some households were lost because they could not be located or because enumerators visited the house three times and could not find a valid respondent. The final household survey instrument is included in Annex 7 of this report.

BASELINE ACTIVITIES

Supervisor and Enumerator Training

In addition to the RA adaptation workshop described above, the evaluation team was involved in several activities to ensure smooth administration of the baseline assessment. The team trained supervisors and enumerators for the reading assessment and accompanying in-school data collection from May 22 to 27, 2013, in Lilongwe. The MoEST selected more than 70 educators and other Ministry staff, many of whom had prior experience with the EGRA, to serve as enumerators and supervisors for the school visits, supplemented by 10 additional supervisors from the Invest in Knowledge Initiative (IKI), SI's local data collection partner. The training was conducted by team members from SI and IKI.This training consisted of a review of the instruments and data collection protocols, practice using the instruments and tablet computers, pilot testing of the instruments in Lilongwe and Kasungu North, and revision and finalization of the instruments based on the results of the pilot testing.

Data Collection

The enumerators and supervisors were formed into 18 teams of between three and seven enumerators each, and began school-based data collection on May 28. Each team was headed by a supervisor who helped to coordinate the work of the team and helped to ensure that all procedures were consistently followed. Based on the EGRA Guidance Notes mentioned earlier, SI planned that one enumerator would be able to complete each reading assessment and the learner interview in approximately 15 minutes, and so would be able to complete all the required assessments for one standard at one school in one day. Thus, most teams had three enumerators—one each to conduct the reading assessments for 15 learners from each of the two standards, and the third enumerator to conduct classroom observations.

Classroom observations were conducted in the classrooms of the teachers whose learners had been selected to participate in the reading assessment, and usually lasted around 90 minutes per standard. The tasks were delegated so that the team members with the greatest knowledge of instructional techniques, often a teacher trainer, primary education advisor (PEA), or other MoEST representative, conducted the classroom observations. The teams that administered the RA in both Chichewa and English included additional enumerators to facilitate administering a larger number of assessments in the same amount of time.

The intended schedule was for the RAs, learner interviews, and classroom observations to all occur during normal school hours, and that after the school day had ended and the learners had left, the enumerators would interview the teachers and head teacher and conduct the school climate assessment. At the beginning of data collection, as enumerators and

supervisors were becoming more familiar with the data collection process, some teams took longer than expected to complete the learner RAs and interviews, meaning that some of the reading assessments were conducted in the afternoon after regular school hours. However, within a week, barring other challenges, all of the teams were able to complete the learner portions of the data collection within the normal school day. The team supervisors helped to coordinate all of the in-school data collection activities and helped to ensure that all procedures were consistently followed.

School-based data collection continued on every school day from May 28 until June 25. Enumerators collected data using paper instruments for the first two or three days of in-school data collection due to delays in receiving and programming sufficient numbers of tablet computers. Later, the data from these paper forms were entered into the tablet computers for uploading to the online server. Throughout the course of in-school data collection, two teams of monitors, one from the MoEST and one from IKI, traveled throughout the 10 districts to ensure consistency of data-collection procedures and quality of the data collected. The SI project manager also visited Malawi from June 2 to June 13 as an additional measure of quality control.

Data Cleaning and Analysis

Following the conclusion of Task 1 data collection, IKI cleaned and compiled the data and submitted it to SI. In July, SI performed an initial review of the data, and in August, upon receipt of the results of the household survey, completed initial analysis of the data. However, upon further analysis, SI found the data to contain many errors, inconsistencies, and mismatched data between datasets. As such, IKI and SI invested in more extensive cleaning of the data throughout September and much of October.

Quality Control

Enumerators were equipped with electronic tablet computers, which allowed them to record the learner responses during the test. This method helped to reduce errors at the time of data collection by automatically prompting required skips, ensuring that all questions were answered in an acceptable format through logic checks, and avoiding data entry (errors can occur when entering data from paper forms into an electronic database). Enumerators made the data available to IKI as soon as it was uploaded to the cloud server from the tablet computers. Having this real-time feedback allowed IKI to catch errors soon after they happened, to alert the enumerators' attention to the issues immediately, and instruct them on ways to remedy these errors. Throughout the process, the data was stored on the tablets and in a remote "cloud" server hosted by the platform Tangerine. Once all of the data was collected and cleaned, it was backed up on high-quality DVDs encrypted with passwords that are changed on a regular basis. The data was also backed up on a protected high-volume

external hard drive to provide an additional layer of redundancy and protect against inadvertent data loss.

ANALYSIS

After all data-collection activities concluded in late summer 2013 and all data was cleaned in early fall, SI analyzed the data to answer the evaluation questions for Task 1 (related to learner reading abilities for Standard 2 and 4 learners); provide some summary statistics for data gathered during the baseline study; relate those summary statistics for learners, as well as teachers, head teachers, classrooms, schools, households, and communities to learner reading outcomes through regression analysis (described below) in order to address Task 2 evaluation questions; and, finally, examined differences between the treatment and control groups both before using PSM (described briefly below and in more detail in Annex 1) and after.

Principal Component Analysis of Factors Used in Regressions

As described above, the evaluation team used regression analysis to look at the effects of various learner, household, teacher, head teacher, classroom, school, and community factors on learner reading outcomes—most specifically oral reading fluency scores, dropouts, and repetition rates.

The baseline data collection process produced a large dataset, including hundreds of variables. Having a large number of variables is necessary in order to capture complex concepts such as school quality and quality of home environments. However, it is not practical to use all these variables in an unrestricted way during data analysis, for two main reasons. When a regression model incorporates several correlated variables, the problem of multicolinearity may emerge. Multicolinearity occurs when several variables within the model are related to each other (correlated) and tend to be observed in similar ways across the same learners. Multicolinearity can cause large standard errors for the coefficients on the correlated variables, sometimes even resulting in a situation where two variables that are correlated and that should have the same signs actually end up with opposite signs. It can also cause two different but related independent variables that have been shown to have an effect on a dependent variable appear to have no significant effect whatsoever. This is because each one diminishes the effects of the other. These kinds of unanticipated results also contribute to a second problem, which is that regression models with large numbers of variables are difficult to interpret. The sheer number of variables leads to complex and unwieldy findings statements, and if the standard errors are large, the regression results often become more confusing to explain.

It is usually much more informative to aggregate these variables into indices, which then convey the main information contained in a group of variables. One way to construct these indices is to use a method called "principal component analysis" (PCA). This method decomposes a set of correlated variables into another set of linearly unrelated components. The single component that is found through statistical analysis to have the most explanatory power, the one that explains the highest amount of variance of the index as a whole, is chosen as the principle component. In a sense, it is then taken to represent all of the other components of the index, but using it in place of the others avoids the problems outlined above related to large numbers of correlated variables.[23] One advantage of using this method over other ways of constructing an index (such as adding or averaging all variables in a group) is that it allows the data itself to guide the construction of the index rather than some external determinant.

In selecting the principal component, PCA also produces a number by which households, learners, or schools can be ranked, allowing for classification of units according to an independent variable of interest. The SI evaluation team used PCA analysis to group some variables together but also included some of the more significant and important variables in regression models on their own. The grouping of variables conducted using PCA included the following:

Wealth. A number of studies have been conducted on the topic of measuring wealth in developing countries, in large part because wealth is difficult to measure in poor populations, due to the population's tendency to rely on unconventional and inconsistent methods of generating income, acquiring and trading goods, and supporting the needs of their families. The majority of experts agree that assets are usually the best indicator of household wealth among impoverished families in developing countries. In fact, one study specifically assessed the validity of this indicator in Malawi and found that asset indices accurately measured livelihoods.[24]

Income is difficult to measure because families tend to make their living farming—often only at the subsistence level, doing day labor, selling small goods, or selling crops grown in excess of subsistence, all of which usually produce inconsistent and unpredictable income. Cash is often scarce among the poor, so households often rely on trade to obtain goods they do not produce themselves. Because of this, household expenditures also are not a particularly

---

[23] There are several textbook treatments of these methods. One example is Stanley Mulaik, *Foundations of Factor Analysis, Second Edition*, Chapman and Hall/CRC, London: 2009.

[24] Carletto, Calogero; Christiaensen, Luc J.M.; Hoddinott, John; and Morris, Saul. "Validity of Rapid Estimates of Household Wealth and Income for Health Surveys in Rural Africa," International Food Policy Institute: October, 1999.

reliable indicator of wealth for this population. Consumption is also an unreliable indicator because families tend to vary their consumption patterns throughout the calendar year because of natural variations in their annual wealth arising from harvest versus "lean seasons" and tourist seasons (for those who sell small goods or rely on tourists for day labor). Thus, assets tend to be the most reliable indicator of long-term wealth, with the caveat that assets do not indicate recent changes in wealth: It takes time for families to acquire assets after they have gained more wealth (either in the form of money or goods to sell/trade). Thus, SI also included more short-term indicators of wealth such as consumption and expenditure.

For this specific study, SI used multiple measures of wealth to allow for triangulation. The evaluation team found that consumption, expenditure, and assets were all highly correlated, so assets were used as the proxy indicator for household wealth. This asset (wealth) index was compiled using the following household variables:

- Main source of drinking water
- Type of toilet facility
- Whether that facility is shared with other families
- Household assets, including whether the family has each of the following: a paraffin lamp, cell phone, table, bed with mattress, sofa, radio, television, ox plow, gold/silver jewelry, a motorcycle or motorized scooter, refrigerator, car or truck, and/or tractor.
- The types of animals that the family owns, including whether they own each of the following: cows, pigs, sheep, goats, chickens, horses, donkeys, oxen/bull.
- The type of fuel that the household uses for cooking
- The main source of lighting for the house
- The materials that the house is made of
- The materials that the roof is made of
- The materials that the floor is made of

Food Security. To set a baseline for the FtF intervention and its possible impacts on household food security and reading outcomes, questions from the FtF Population Survey food security module were included in the HHS. A PCA was performed to determine which of the variables from this module best predicts food insecurity. These include whether, in the four weeks preceding the baseline survey, the household respondent (usually the head of the household) or any other household member:

- Expressed concern that the household would not have enough food
- Was not able to eat the kinds of foods they preferred

- Had to eat a limited variety of foods due to inability to obtain other foods
- Had to eat fewer meals in a day because there was not enough food
- Went to sleep hungry at night because there was not enough food to eat
- Went a whole day or night without eating anything because there was not enough food
- Did not have any kind of food to eat because of a lack of resources to get food

The same thing was done for health variables for GHI. However, when lumped together, those variables did not make much sense, so the evaluation team chose to disaggregate the health variables.

School resources. The evaluation team also created a PCA score for school resources using the school climate protocol and head teacher questionnaire questions. The following factors were included in that PCA score:

- Whether the school had a library
- Whether the library was well stocked
- Whether the school had electricity
- Whether the school had a school feeding program
- Whether the windows on the school buildings were in good shape
- Whether there were locks on classroom doors
- Whether there was sufficient classroom space
- Whether there was a teacher lounge
- Whether the school had access to clean water
- Whether the school had latrines
- Whether the school had teacher latrines
- Whether there were sufficient desks for learners

Outside of this PCA score, evaluators also considered whether the school had a girls' latrine as well as the learner-to-teacher ratio for the standard level being assessed. These were taken out of the PCA score due to their perceived importance and explanatory capability.

Teacher Use of Best Practices in Teaching Reading. The final PCA used by the evaluation team was teacher use of best practices in teaching reading. This PCA was developed through analysis of all of the variables included in the table for the classroom observation protocol, which can be found in Annex 6.

Balancing Treatment and Control Groups

Ensuring that treatment and control groups are as similar as possible at baseline is important for ruling out selection bias (the possibility that the treatment group and the control group are somehow different, and that those differences might then be the cause of changes in key outcomes rather than the project(s) being assessed). As such, to verify that the four treatment types and control groups have similar observable characteristics, SI developed balancing tables that compare important variables between the groups. This was done before matching pairs of learners between treatment and control groups (through PSM, which seeks to match individuals based on multiple variables, as described in Annex 1) and after. Despite this matching, it is quite possible that some selection bias exists in the evaluation sample, since the evaluation team was unable to use a completely randomized evaluation design. To control for this in later studies in 2015 and 2017, SI will use the difference-in-differences methodology, which seeks to compare changes among groups rather than comparing the groups' initial starting points and ending points. Thus, even if the treatment group is somewhat different than the control group at baseline, as long as the two groups can be expected to experience increases in outcomes at the same rate, difference-in-differences will allow the evaluation team to rule out selection bias and attribute differences in outcomes between treatment and control groups to the interventions being studied.

THREATS AND LIMITATIONS

The experimental portion of the evaluation design allows for robust conclusions about the impact of the EGRA intervention on learner performance in various settings (across treatment types). Nevertheless, there are still several important potential threats to the validity of the study. The following section discusses these threats and limitations.

Attrition, Absenteeism, and Grade Repetition

One important threat is the depletion of the sample size. A reduction in sample size reduces evaluators' ability to be confident in identified changes in key outcomes. As described earlier, the GoM reports absenteeism rates averaging more than 25 percent of the learner population in the lower standard levels.[25] This can affect whether learners are receiving the full advantages of treatment. To mitigate this threat, as described above, enumerators asked teachers about absentee rates of the children assessed before they selected them to take part

[25] USAID Malawi. "2010 Early Grade Mathematics Assessment (EGMA): National Baseline Report," 2010. http://mtpds.org/images/stories/pdfs/baseline_reports/malawi_national_egma_2010_baseline_report.pdf, accessed August 26, 2013.

in the study and attempted to only include learners in the study with an attendance rate of at least 75 percent.

In addition to absenteeism, reputation is also a major issue in Malawi. In 2010, Malawian primary school repetition rates ranged from 12.8 percent in Standard 7 to 24.6 percent in Standard 1. Repetition often arises because learners are not doing well in school and can be related to excessive absences. Issues of absenteeism and repetition often culminate in learners dropping out of school before completing the full eight levels of compulsory education. The result is extremely high dropout rates. In 2010, the Standard 1 dropout rate in Malawian schools was 12.7 percent.[26] Furthermore, cohort analyses indicate that only 53.1 percent of boys and 45 percent of girls who initially enrolled in primary school enrolled in Standard 8 in 2010.[27] To help mitigate the effects of attrition, SI oversampled learners. However, to avoid overestimating the impacts of the EGRA Project (due to loss of the lower-performing learners to attrition), the evaluation team also plans to track those learners who drop out and follow up with them to find out why and to what extent the project affected their decision.

A final issue related to attrition is contamination of control schools when learners or teachers move from a treatment school to a control school, or of diminishing the potential impact of the treatment when a learner, teacher, or head teacher moves mid-term from a control school to a treatment school. To help reduce the likelihood for such contamination, enumerators selected only learners who had been in the classroom since the beginning of the academic year.

Spillover

The threat of spillover is somewhat related to that of contamination. Spillovers occur when the intervention taking place in the treatment schools also affects the control schools. This can occur, for example, if treatment and control schools are located in nearby areas and teachers in treatment schools interact with teachers in nearby control schools and share knowledge of the EGRA intervention, thereby potentially influencing the teaching practices in control schools. Spillover might be somewhat more likely to occur with the FtF INVC Project than with the EGRA Project, but both types of spillover could affect evaluators' ability to measure the true effectiveness of the projects.

SI is working closely with the EGRA IP to ensure that their intervention minimizes spillover, and will, to the degree possible, continue to collect data throughout the course of the impact evaluation to identify when and how spillover may affect the results. Further, when selecting

[26] This was the highest dropout rate of any of the standard levels.

[27] *Ibid.*

control zones, SI left out the zones that bordered the FtF treatment zones too closely to help avoid problems with spillover of agricultural knowledge and information.

Comparability of Contexts: Internal Validity

Another limitation inherent to the design of this evaluation (and that of any evaluation that does not rely on completely random selection of treatment and control groups) is the possibility that the contexts where the four treatment levels are being implemented are different in ways that influence treatment and control groups differently. Central to this evaluation is the ability to make valid comparisons between treatment and control groups. To do this, the evaluation design assumes that these groups of learners, and all of the factors acting on them that may influence reading performance, are similar enough so that observed differences in performance can be attributed to the intervention itself and not to other, confounding factors. However, even after collecting data on a wide range of teacher-, school-, household-, and community-level variables, the possibility remains that the learners being compared in treatment and control groups are somehow dissimilar in an *unobservable* way that might have an effect on their reading performance. If this were the case, then the internal validity of the evaluation would be weakened, and the results of the IE may not be a completely accurate reflection of the true impact of the EGRA intervention.

The fact that the four treatment levels were assigned to geographic districts in a non-random way means that there are almost certainly important differences among the learners in these treatment levels that are likely to affect their reading performance. The PSM process, described in Annex 1, helps to minimize these differences by identifying treatment and control schools that are similar to each other based on a set of variables. However, it is still possible that some important factor has been left out and that the groups will remain meaningfully dissimilar. Selecting a robust set of variables to include in the PSM process, based on established literature in the field and expert knowledge of the education system and community contexts Malawi, helps to reduce this risk, as does collecting data on many factors that may be influencing learner reading scores, which SI has done.

Comparability of Contexts: External Validity

Even if the evaluation were completely valid internally, there is still the question of whether the conclusions are valid for contexts outside of the intervention area. Just as there is no guarantee that learners in the Level 1 districts are sufficiently similar to those in Level 4 districts to allow for accurate conclusions between the two levels, evaluators cannot be certain that the learners in these 10 intervention districts are similar enough to learners in other districts in Malawi to claim that the results measured in this evaluation would be replicable in other parts of the country. If the learners in these 10 districts are not representative of the whole population of learners in Malawi, then it is possible that

expanding the EGRA Project or similar future projects to other districts may not lead to the same results. This is a potential weakness in the evaluation's external validity that exists for all evaluations, especially impact evaluations, which are heavily dependent on context.

Having extensive data about the learners, their households, and their communities will help users of this evaluation to assess how similar the context of the learners sampled for the evaluation is to other contexts to which they may be interested in expanding. This will allow USAID and other stakeholders to make an informed determination about how appropriate it would be to extrapolate the results of this evaluation to those other contexts.

Sample Size

Another concern about the IE analysis is the possibility that the sample sizes will be too small to make meaningful conclusions for some levels of disaggregation—especially should issues of attrition and contamination arise. The evaluation team is confident in its power analysis and the belief that the data will be adequate to estimate the impact of EGRA interventions in different types of districts and to differentiate effects between each level of treatment and their respective comparison group (Level 1 vs. Level 4; Level 2 vs. Level 4; Level 3 vs. Level 4). However, it may be more difficult to draw statistically significant conclusions comparing across intervention levels (Level 1 vs. Level 2, for example) or across different districts within a level. The reason is that differences in effects across interventions are likely to be much smaller than the differences within a treatment level between the treatment and control groups, and larger sample sizes are required to measure smaller effects. This is not to say that future reports will not be able to provide information on which levels or districts appear to have benefited most from the EGRA intervention—data that could be informative for the Mission and other stakeholders. However, without significantly increasing the sample size and, thus, data collection costs, it may not be possible to be fully confident in these results to a level of statistical significance.

Past reading assessment studies have encountered similar problems when trying to assess differences in test scores by sex.[28] In fact, the 2010 and 2011 reading assessments were unable to detect statistically significant differences between female and male learners.[29]

---

[28] RTI International. "Early Grade Reading Assessment Toolkit," March 2009. <http://eddataii.rti.org/documents/index.cfm/EGRA_Toolkit_Mar09.pdf?fuseaction=throwpub&ID=149>.>.<http://eddataii.rti.org/documents/index.cfm/EGRA_Toolkit_Mar09.pdf?fuseaction=throwpub&ID=149>.

[29] USAID Malawi. (2010). Early Grade Reading Assessment: National Baseline Report. <www.eddataglobal.org/reading/index.cfm/Malawi%20National%20Baseline%20EGRA%202010.pdf?fuseaction=throwpub&ID=354>. And: USAID Malawi. "Malawi National Early Grade Reading Midterm Assessment, 2011," June 2012. <http://mtpds.org/images/stories/pdfs/result_area_3_resources/malawi_%20egra_2011_midterm_report.pdf>.

However, those studies only used samples of 1,000 and 3,000 learners, respectively, whereas this evaluation uses much larger sample sizes, which are well above 7,000, the figure recommended by the EGRA Toolkit to ensure statistically representative effects by gender.

To further mitigate data threats, quantitative analyses provided through the reading assessment are supplemented by the qualitative school, learner, classroom, and household data to triangulate results and minimize the threat of false positives. Finally, the team has and will continue to ensure that all reports of findings include clear cautions regarding responsible interpretation and use of the results.

Use of Government Employees as Supervisors and Enumerators

Another potential threat to the accuracy and reliability of the data is the use of MoEST staff and other GoM employees as enumerators and supervisors. SI recognizes the value of involving the MoEST in this process: It capitalizes on existing experience and expertise, especially of those individuals who have been involved in EGRA-related activities in the past; it increases ownership of the MoEST of study results; and it builds the capacity of the MoEST. However, there is always a risk when the same actors who are responsible for overseeing or implementing a project are asked to evaluate the project. It may be in the interest of some individuals or groups within the MoEST to show improved reading outcomes as a result of the EGRA, or to show no change or a negative trend. In any case, when individuals who may have conflicting interests are involved directly in evaluation activities, there is always a risk that they may somehow inappropriately influence the results of the evaluation. On the other hand, GoM personnel, and in particular, MoEST staff, have been involved in data collection activities for other projects in the past and have conducted themselves in a professional and objective manner. For this reason, there is reason to believe that this risk is relatively low.[30]

Also, these data serve an important purpose for the GoM and especially the MoEST. As such, to help inform their decisions related to reading teaching and learning, these bodies have a vested stake in obtaining accurate information from these assessments. It was then expected that, in addition to SI and IKI's procedures to ensure that all data collection protocols were faithfully followed during data collection, the leaders of these government bodies would also communicate to their staffs the importance of abiding by these protocols and capturing an accurate picture of learner performance. SI very much found this to be the case during data collection, and the evaluation team worked closely with IKI and its GoM partners to minimize these risks and obtain the highest quality data possible.

---

[30] GoM staff have been involved in data collection activities for the Southern and Eastern Africa Consortium for Monitoring Education Quality, Paralegal Advisory Service, and Monitoring Learning in Africa, as well as the previous reading assessments conducted under the Malawi Teacher Professional Development Support (MTPDS) Project.

Matching Data between Datasets

During data cleaning for this baseline assessment, IKI and SI had difficultly matching the various databases of information (matching learner assessment information to teacher interview and school climate protocol information, for instance) to one another. The reason for this is that enumerators were not consistent in their capturing of key information, such as EMIS ID number, for each of the different data collection instruments administered at the schools. Likewise, there were problems related to consistency of learner ID numbers between the assessments and household surveys. As such, matching up data proved difficult during data cleaning and resulted in the loss of some records for purposes of overall analysis. Despite this fact, SI feels confident in the matches that were made, and, as described in the sample size section of this report, the sample has proven more able to allow for detection of even smaller results than originally planned. Also, this problem should be solved through more training, clear protocols, and the use of tablets for all data collection instruments in the future.

IV. LITERATURE REVIEW FINDINGS

FACTORS THAT INFLUENCE LEARNER ACHIEVEMENT IN READING

Prior to conducting its assessment, SI conducted a literature review to determine factors that may influence learner reading ability. These factors were incorporated into survey instruments and are analyzed in relation to reading assessment results later in this report.

Many factors have been found to have a strong influence on learning to read, including factors in the home, at school, and in the classroom. Based on a review of the research and evaluation literature regarding those factors, data was gathered through the various instruments developed and/or adapted for this study. Unless otherwise noted, much of the findings regarding factors found to influence learning to read and reported below were drawn from the 2011 Progress in International Reading and Literacy Study (PIRLS),[31] an international assessment of learner reading that tested more than 300,000 fourth-grade learners across 45 countries (including the two sub-Saharan African countries of Botswana and South Africa).

HOME FACTORS THAT INFLUENCE READING

Parents' Education and Home Environment

A recent review of the literature on the correlation in education between parents and children in most countries is quite high, with correlations varying between 0.30 and 0.50. The intergenerational persistence is also high for other outcomes and traits such as income and cognitive ability.[32] The authors concluded that in all countries for which they have data, more than 50 percent of the variation in years of schooling can be attributed to factors shared by siblings, such as parents, family wealth, and the home environment.

Another study covering a 40-year period found that children of parents who are educated are far more likely to do well in school and later in life. The study followed almost 900 children from Standard 3 until they were 48 years old and found that parents' educational level when

---

[31] Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

[32] Bjorklund, A., and Salvanes, K.G., (2010). "Education and Family Background: Mechanisms and Policies." IZA DP No. 5002, in Handbook of the Economics of Education, Volume 3, Eds: Hanushek, E.A., Machin, S.J., and Woessmann, L. (2010). Amsterdam: North Holland.

the child was eight significantly predicted that child's educational and occupational success up to 40 years later.[33]

Parental Expectations

The most consistent predictors of children's academic achievement and social adjustment are parental expectations of the child's academic attainment and their satisfaction with their child's education at school. Parents of high-achieving learners set higher standards for their children's educational activities than do parents of low-achieving learners.[34]

Parental Involvement

A 2002 synthesis of research on parental involvement found that learners with involved parents are more likely to get better grades, score higher on tests, enroll in higher-level programs, be promoted, pass their classes, attend school regularly, have better social skills, graduate from secondary school, and attend postsecondary education.[35]

Another synthesis of research also found that the earlier parents become involved in their child's education, the more powerful the effects.[36] The most effective forms of parental involvement are those that engage parents in working directly with their children on learning activities at home.

The PIRLS study found that homes where family members engage their children in literacy activities on a regular basis are also far more likely to produce fluent readers. These activities include activities such as reading to children frequently.

The PIRLS study also found that children who had attended three or more years of pre-primary education scored 44 points higher on their reading assessment than did those who had never attended pre-primary. Even children who had attended one year or less scored 18 points higher than did those who had never attended pre-primary.

---

[33] Palmer, M. Q (Wayne State University Press). 2009 July; 55(3): 224–249. doi: 10.1353/mpq.0.0030

[34] Michigan Department of Education (2002). "What Research Says About Parent Involvement In Children's Education." M

[35] Southwest Educational Development Laboratory (2002). "A New Wave of Evidence."

[36] Cotton, K., Wikelund, K., Northwest Regional Educational Laboratory, School Improvement Research Series. "In Parent Involvement in Education."

SCHOOL FACTORS THAT INFLUENCE READING

Resources

The resources available to support learner learning were also found to be a factor in reading achievement. Typical resources found in the schools with higher reading scores on the PIRLS were instructional materials, books, computers, library books, audiovisual equipment and supplies, as well as sufficient numbers of and well-maintained buildings, adequate classroom space, and well-trained and sufficient numbers of staff. These higher-performing schools also had specialized staff, such as teachers who were more highly trained in supporting learners' development of reading skills. Schools that were well resourced in the PIRLS study scored an average of 45 points higher than schools that were poorly resourced.

Strong Academic Focus

Factors such as a strong school leader (head teacher), effective teachers, learners who want to do well, and parental support were highly correlated with reading achievement. The effect on learner achievement is greatest when there is a collective influence, when the school leaders, teachers, and parents support and trust in learners' achievement.

Early Start to Reading

Learners who attended schools that focused on developing essential reading skills at or before Standard 2 scored higher on the reading assessment than did those schools that delayed the focus on reading skills until Standards 3 or 4.

Safe and Orderly Schools

Learners in schools that principals and teachers reported are orderly and provide a safe environment for learning performed 43 points higher on the PIRLS than did learners in schools with moderate security problems, such as bullying, lack of discipline, and classes that do not start and end on time. Learners who reported almost never being bullied scored an average of 34 points higher than did learners who reported being bullied about weekly.

Experienced Teachers

Learners performed somewhat better on the PIRLS when their teachers had more than 10 years of teaching experience, and in that study, most of the teachers had specialized training in the teaching of reading. Even though the PIRLS study did not specifically measure the correlation between the amount of training teachers had in teaching reading and learners' performance in reading, there has been growing evidence that teacher preparation is a

powerful predictor of learners' achievement, perhaps even overcoming socioeconomic background factors.[37]

Confidence in Reading Ability

Learners who reported liking to read scored higher than did learners who reported not liking to read; the authors of this study stated that being a good reader is more likely to result in enjoying reading, while the reverse is also likely to be true—not being able to read fluently is likely to undermine one's enjoyment of reading. In fact, the learners who reported being confident in their reading ability scored a substantial 91 points higher on the PIRLS reading assessment than did those who were not confident, and 45 points higher than those who were only somewhat confident.

Learner Engagement and Interest in School

Learners who reported being more engaged in their lessons scored higher than those who reported not being engaged. Teacher practices that encouraged engagement at the fourth-grade level included summarizing the lesson's learning goals, questioning to elicit reasons and explanations, and bringing interesting things to class.

Class Size

Class size is a complex issue and the research on class size is mixed. Several studies in the US have found considerable academic benefits to lower class sizes, but other studies have not. For example, a highly credible study conducted in US State of Tennessee found that reducing class size from an average of 22 learners to an average of 15 (a 32 percent reduction in class size) increased learner achievement over a four-year period comparable to adding three months of instruction to the school year.[38] Similar studies in Texas[39] and Israel also found benefits of smaller classes, but the gains were smaller than in the Tennessee study.[40] That study also estimated the cost-benefit of the smaller class size and found a positive rate-of-return of about 6 percent.

---

[37] Darling-Hammon, Linda, "Teacher Quality and Student Achievement," *Education Policy Analysis Archives* v 1 p. 8, (2000).

[38] Alan B. Krueger, "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 115(2): 497–532 (1999).

[39] Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 73(2): 417–458 (2005).

[40] Chingos, M. M. and Whitehurst, G. J. (2011), "Class Size: What Research Says and What it Means for State Policy," Brookings Institution, Washington DC.

In the international setting, Woessman and West looked at the impact of class size on performance on the TIMSS international examination in 11 countries and found mixed results.[41] However, the countries with the most significant impact of smaller class sizes were those with less qualified teachers and low salaries, suggesting that teachers with less experience, lower levels of qualifications, and lower salaries are more likely to have an easier time teaching in classrooms with fewer learners, while more qualified teachers, who are better paid, are more likely to teach as well in larger classes as they are in smaller ones.

Chingos and Whitehurst caution that conclusions drawn from class-size studies should be considered tentative, arguing that the substantial variability in the settings, methods, grades (standards), and magnitude of class sizes studied makes it difficult to generalize.[42]

However, studies seem to show that very large class-size reductions, on the order of 7 to 10 fewer learners per class, can have significant long-term effects on learner achievement and other meaningful outcomes. The effects seem to be largest when introduced in the earliest grades and for learners from less advantaged family backgrounds.

Additionally, almost none of the studies considered class sizes as large as many classrooms in Malawi, but instead were looking at reductions in the range of 40 to15. The issue and the research are further complicated by the fact that many very small classes in rural areas are multi-grade classrooms in poorly-resourced schools with learners from extremely poor households, each of which is likely to undercut the benefits of smaller class sizes.[43]

Girls' Education

Extensive research has found substantial social, health, and economic benefits to educating girls. When girls are educated, their families are healthier, they marry later, have fewer children, have more children who survive childhood, provide better care for their children, and have more opportunities to generate income.[44] Returns on girls' education in developing countries are substantial and in most cases exceed those observed in developed countries and those based on the education of boys. A cross-country study of data from 50 countries estimated that each additional year of education for girls increases long-term growth by 0.58

---

[41] Ludger Woessmann and Martin West, "Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS," *European Economic Review*, 50(3): 695–736 (2006)..

[42] Chingos, M. M. and Whitehurst, G. J. (2011), "Class Size: What Research Says and What it Means for State Policy," Brookings Institution, Washington DC.

[43] Cotton, K., Wikelund, K., (ND) School Improvement Research Series. "In Parent Involvement in Education," Northwest Regional Educational Laboratory, Oregon.

[44] Chaaban, J., and Cunningham, W., (2011). "Measuring the Economic Gain of Investing in Girls: The Girl Effect Dividend." World Bank Policy Research Working Paper N0. 5753, Washington, D.C., World Bank.

percentage points per year.[45] A 1999 World Bank study that used simulated data from 100 countries found that increasing the secondary education of girls by as little as 1 percent results in an increase of 0.3 percent per capita, a substantial increase for many developing countries.[46]

The authors of one study argued that the 40 percent decline in malnutrition in the developing world since 1970 is due to the increased numbers of educated girls and women. Their rationale was that since the majority of the farmers in the developing world are women, the increases in girls' education has resulted in more productive farming, which in turn has led to the dramatic decline in malnutrition.[47]

Malawi has had great success in increasing the number of girls in school, now reaching parity with boys' enrollment. However, between 2002 and 2007, although both boys and girls showed reasonable improvement in learning achievement, boys consistently outperformed girls in reading by more than 10 points, with regional variations in the learning growth and the gender-learning gap. This gap was most notable in the South West Division, where the reading scores for both girls and boys declined during that five-year period. However, the girls' decline was much greater and was especially significant since there had been no gender difference in 2002.[48]

Insufficient Numbers of Female Teachers and Head Teachers

This same study found that female representation in the teaching force had stagnated over time and even declined slightly between 2002 and 2007, with females representing only 25 percent of Standard 6 teachers and head teachers only 12 to 13 percent during that period. The MoEST responded by committing to increase female representation in the teaching force to at least 35 percent in the National Education Sector Plan (NESP), [49] acknowledging that the

---

[45] Hanushek, E. & Woessmann, L. (2007). "The role of education quality in economic growth." Policy Research Working Paper 4122. Washington, DC: World Bank.

[46] Dollar D., Gatti R. (1999). "Gender Inequality, Income, and Growth: Are Good Times Good for Women?" World Bank Policy Research Report on Gender and Development, Working Paper Series 1. Washington, D.C: World Bank.

[47] Center for Global Development, (ND). "Education and the Developing World," Washington DC. http://www.cgdev.org/files/2844_file_EDUCATON1.pdf

[48] SACMEQ Policy Brief No. 6 (2011), "Progress in Gender Equality in Education in Malawi." http://www.sacmeq.org/sites/default/files/sacmeq/publications/06_mal_policy_brief_-_gender_newformat_ver12_final_111004.pdf

[49] MoEST. (2008a). "The national education strategy plan 2008–2017, a statement." Lilongwe: Government of Malawi.

lack of female teachers had a harmful impact on female learners, in part due to the dominance and conduct of male teachers, especially in the rural areas.[50]

Health and Nutrition

The PIRLS study found that the 27 percent of learners whose teachers reported that malnutrition among learners was a problem scored 24 points lower than when the teachers reported that learners were sufficiently nourished. Sleep deprivation was also correlated with lower scores on the PIRLS reading assessment.

Research in the developing world has consistently found that education and health are very closely linked and mutually dependent. Studies show that children who are healthier do better in school, partly because they do not miss as many days of schools as those who are unhealthy. The direction of the cause-effect relationship shifts after people are educated, according to the Center for Global Development, which has found that those who are educated are more likely to be healthy and to use health services effectively. For example, those who complete primary education are half as likely to contract HIV as those with little or no schooling.[51]

As described above, the evaluation team used these literature review findings to help develop data collection instruments to capture information on factors shown to affect reading scores in other studies. In the findings section on Task 2 Evaluation Questions, below, the evaluation team presents analysis of these factors and their correlation with reading scores in Malawi. Many of these findings concur with the findings reported here.

---

[50] MoEST. (2008b). "Education management information system 2008 report." Lilongwe: Government of Malawi.

[51] Center for Global Development, (ND). "Education and the Developing World," Washington DC. http://www.cgdev.org/files/2844_file_EDUCATON1.pdf

V. BASELINE STUDY FINDINGS

OVERVIEW

The findings section is presented with the goals of showing overall learner reading outcomes and examining the factors that affect those outcomes. However, since this is just the baseline assessment (and, thus, a snapshot), all results on factors that correlate with learner reading outcomes are preliminary. Therefore, evaluators first present some summary statistics on those factors that have been shown by other studies to affect learner reading outcomes or success in school (with numbers and/or percent of learners meeting those criteria) before presenting analysis on what seems to matter based on the preliminary results from this study.

The analysis of learner reading outcomes includes a summary table that presents the answers to all of the Task 1 evaluation questions described above. This table is followed by more detailed analysis of reading assessment results by subtask and according to the stages of reading. The analysis of what matters and what does not when it comes to learner reading scores is presented by providing preliminary answers to the four Task 2 evaluation questions noted above. Finally, the end of the findings section looks at the differences between this study's treatment and comparison groups in an effort to inform the balance of this study moving forward. An outline of subsections under findings follows:

1. Overall RA results of all learners assessed on the Chichewa RA across the 10 sample districts.[52]
2. Summary statistics on factors that have been shown (in other studies, such as those described in the literature review) to affect learner reading outcomes. This section does not attempt to tie these factors to learner reading outcomes from this study but instead presents a snapshot of learners and schools that this baseline study assessed to give reviewers an idea of what the study population looks like as a whole.
3. Analysis of which of the above factors actually makes a difference (based on snapshot results from this baseline survey) in learner reading outcomes, community engagement, and success in school. This analysis is presented in the form of preliminary answers to this study's Task 2 evaluation questions, which seek to understand these relationships.
4. A comparison of the treatment and comparison groups for this study.

[52] Note that, as described above, findings are not representative at the district level. Instead, they are representative of each of the four different EGRA treatment arms.

READING ASSESSMENTS

A Brief Explanation of the Stages of Learning to Read

Regardless of the language, all children who learn to read advance from being non-readers (unable to read words), to partial readers (can read some items but not others), to fluent readers (can read all or a majority of words). At each stage, they develop a different set of competencies, from listening comprehension to initial reading skills to comprehension and fluency (can read with ease and speed). The EGRA RA measures children's abilities according to the first three of the four stages of reading development:[53]

Stage 1: Pre-Reading Skills. In Stage 1, which typically lasts from birth to kindergarten, children should learn oral language and listening comprehension skills. These pre-reading skills are measured by the *Listening Comprehension* subtask.

Stage 2: Initial Reading/Decoding Skills. In Stage 2, between Standards 1 and 2, children should learn to associate letters with their corresponding sounds. Using phonics, children sound out letters, decoding the word that is formed when they combine the sounds. Readers at this stage often focus on individual words and phrases and miss the larger meaning of the story. These initial reading/decoding skills are measured by the following initial reading subtasks: *Letter Name Recognition*, *Initial Sound Identification*, *Familiar-word reading*, and *Non-Word Reading*.

Stage 3: Confirmation and Fluency. In Standards 2 to 3, children should become more fluent in recognizing or decoding words. Rereading familiar books and reading stories with familiar or stereotyped structures, children gain speed, fluency, and confidence in their reading ability. They should also be able to read unfamiliar words through the process of decoding and gaining meaning from text. Comprehension and fluency are measured through the *Oral Reading Fluency* (reading passage) and *Comprehension* subtasks. Based on scores from the tests in this section, learners fall into one of three categories: non-readers, partial readers, or fluent readers. Zero scores determine non-readers; partial readers are identified by comparing the mean score to the benchmark; and, fluent readers are those meeting the established EGRA Coordinating Committee benchmarks for oral reading fluency, which is 20 correct words per minute (cwpm) in Standard 1 and 50 cwpm in Standard 3. For Reading Comprehension, the benchmark is 80 percent of the questions answered correctly.

Stage 4: Reading to Learn. Once children have acquired fluency, they reach stage 4, which allows them to move to less familiar material. By Standard 4, children should have transitioned from "learning to read" to "reading to learn," focusing on fact-based information

---

[53] Chall, J. (1983). *Stages of Reading Development.* N.Y.: McGraw-Hill Book Company.

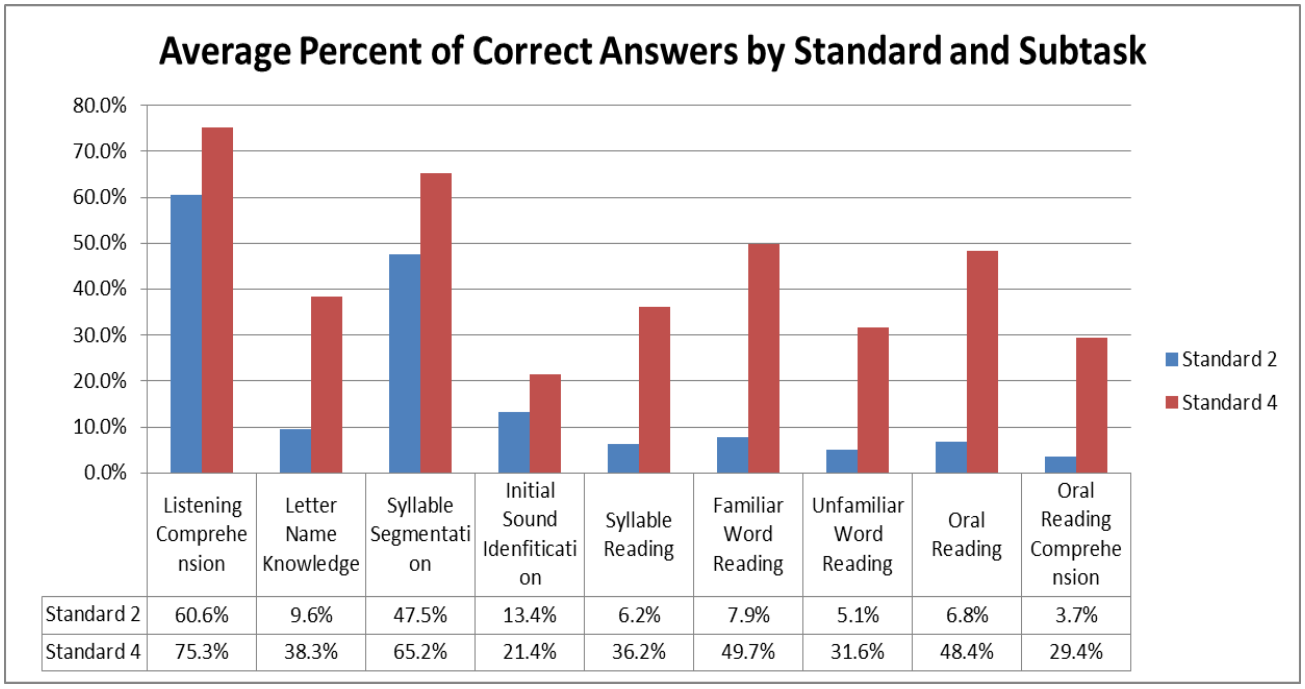
and academic subjects like science, history, and geography. While the Reading Comprehension subtask may begin to address whether learners can learn to read, the EGRA tool does not measure whether learners have reached stage 4 because its focus is on early-grade reading.

Because learners in Malawi are taught primarily in Chichewa through Standard 4, all 8,910 of the sampled learners were tested in Chichewa, and the results from that assessment are provided below. The reading assessment results reported here are compared against the recommended benchmarks established for reading by the MoEST National EGRA Coordinating Committee in 2011. While the EGRA Coordinating Committee recommended benchmarks were established for Standards 1 and 3, SI's SOW for this evaluation specifies that this baseline assessment and the mid- and final-evaluation assessments be conducted with learners in Standards 2 and 4. Thus, learners in this study are compared against benchmarks for the standard below the one in which they were enrolled at the time of this study. The performance of learners in Standards 1 and 3 will be assessed in 2014 and 2016 through the National Reading Assessment.

Summary of Reading Assessment Results

Generally, many learners received zero scores on multiple subtasks and very few learners met the EGRA Coordinating Committee recommended benchmarks for the standard below their current standard. However, the majority of learners in both standards achieved or surpassed the EGRA Coordinating Committee recommended benchmarks for listening comprehension, thus, suggesting they can competently understand spoken Chichewa. This suggests that the majority of learners from this study have mastered pre-reading skills. As far as the ability to read fluently (confirmation and fluency skills) goes, despite the very low scores on the reading fluency and reading comprehension subtasks, the gains in scores from Standard 2 to Standard 4 indicate that some learning is evident: Figure 1 shows that learners in Standard 4 score higher on all subtasks, though the gains in learning from Standard 2 to Standard 4 vary substantially by subtask. These differing results are not surprising, however, as subtasks differ in the level of difficulty and skills needed. Figures 2 and 3 show the mean scores for the sampled Standard 2 and 4 learners as compared with the EGRA Coordinating Committee recommended benchmarks for Standards 1 and 3. The large majority of Standard 2 and Standard 4 learners did not the Standard 1 and 3 benchmarks, respectively. Table 4 provides a summary of learner reading results by reporting on all of the indicators and questions from Task 1 of this evaluation contract (See page 10 of this report for the Task 1 questions).

Figure 1. Average Percent of Correct Answers by Standard and Subtask



	Listening Comprehension	Letter Name Knowledge	Syllable Segmentation	Initial Sound Idenfiticaion	Syllable Reading	Familiar Word Reading	Unfamiliar Word Reading	Oral Reading	Oral Reading Comprehension
Standard 2	60.6%	9.6%	47.5%	13.4%	6.2%	7.9%	5.1%	6.8%	3.7%
Standard 4	75.3%	38.3%	65.2%	21.4%	36.2%	49.7%	31.6%	48.4%	29.4%

Figure 2. Average Percent Correct by Subtask for Standard 2 Learners

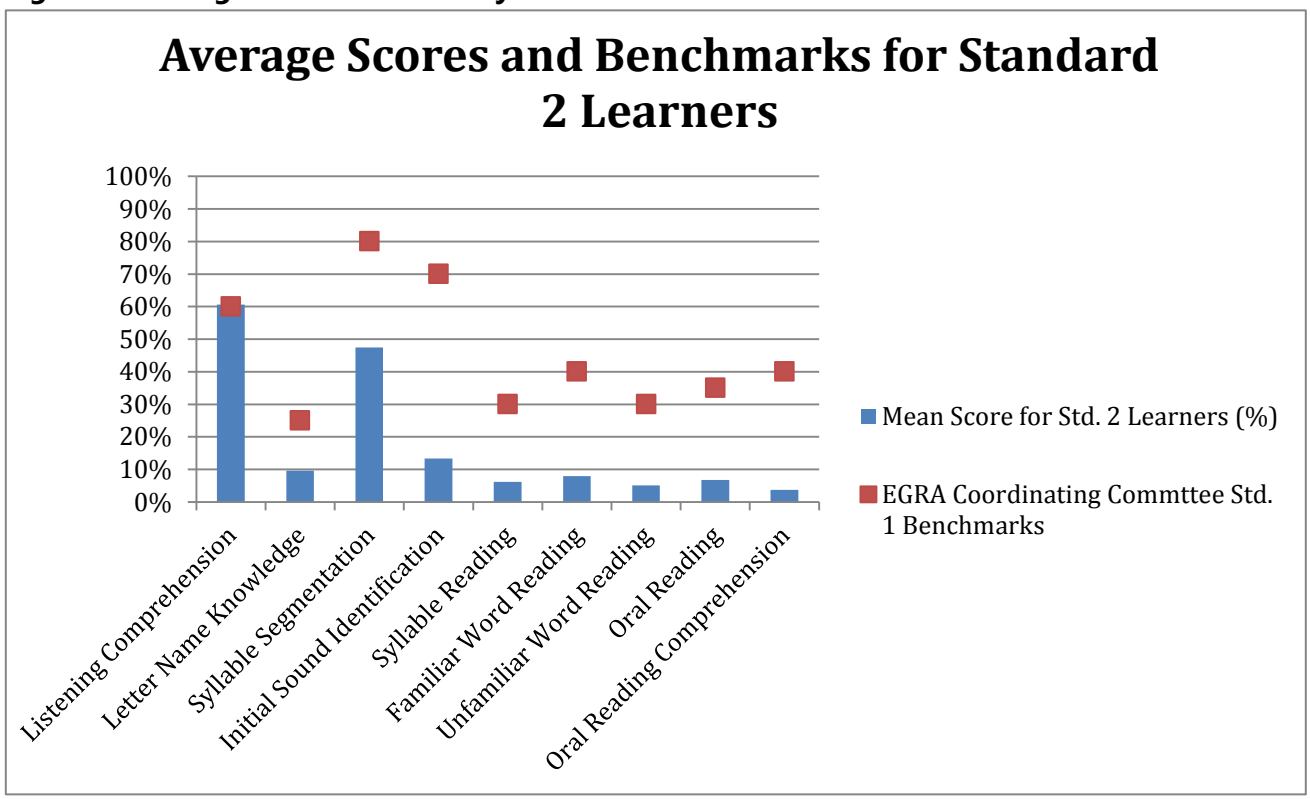


Figure 3. Average Percent Correct by Subtask for Standard 4 Learners

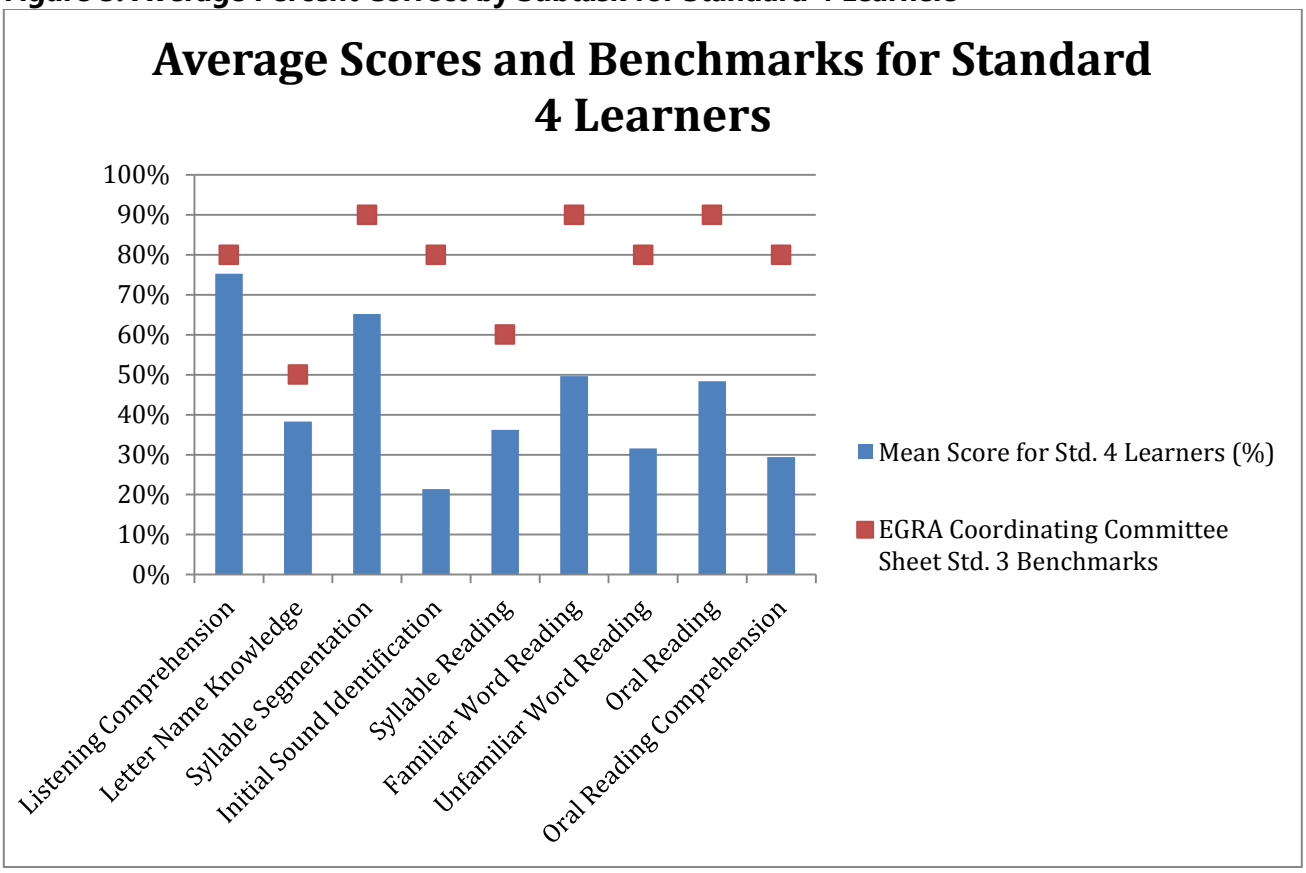


Table 4. Task 1 Evaluation Questions and Indicators

Proportion of primary school learners who are able to read with comprehension, according to Malawi's curricular goals, by the end of lower primary school (Standard 2).	
Proportion of learners who are able to read grade-level text by the end of the second school year	
MoEST benchmark: 20 cwpm	Percentage meeting benchmark: 8.4%
Proportion of learners who are able to answer comprehension questions after reading grade-level text by the end of the second school year	
MoEST benchmark: 40% correct	Percentage meeting benchmark: 3.4%
Proportion of primary school learners who are able to read with comprehension, according to Malawi's curricular goals, by the end of lower primary school (Standard 4).	
Proportion of learners who are able to read grade-level text by the end of the fourth school year	
MoEST benchmark: 50 wpm	Percentage meeting benchmark: 10.8%
Proportion of learners who are able to answer comprehension questions after reading grade-level text by the end of the fourth school year	
MoEST benchmark: 80% correct	Percentage meeting benchmark: 10.8%

READING ASSESSMENT FINDINGS BY SUBTASK

This section reports results of the EGRA by subtask and is organized based on the three stages of reading development described above. Each section includes a brief description of the subtask(s), the mean score for the sample in each standard, and the percentage of learners achieving the EGRA Coordinating Committee recommended benchmarks for Standards 1 and 3. Zero scores are also provided by standard and district. However, it should be noted that while the sample from each district may be sufficiently large to allow the evaluation team to be confident in reported results, the sampling methodology resulted in a sample that may not be representative of the district as a whole.

Stage 1: Pre-Reading Skills

Listening Comprehension. Testing of listening comprehension is important because it is a complete measure of a learner's ability to comprehend stories due to the different ways in

which learners approach, process, and respond to text. It is also a skill that learners should develop before learning to read. The purpose of this assessment is to see whether the learner can listen to a passage being read aloud and then answer several questions correctly with a word or simple statement. The listening passage read to learners during the RA consisted of a short fable, about 30 words long, which was followed by four comprehension questions. Scores for this subtask are based on the percentage of those four questions answered correctly.

Listening comprehension results (see Table 5) indicate that about 60 percent of learners in Standards 2 and 4 are meeting the EGRA Coordinating Committee benchmarks for Standards 1 and 3, respectively. The mean percentage of questions answered correctly in Standard 2 was 60.6 percent (or somewhere between two and three correct answers out of four), which is just more than the benchmark for Standard 1. Learners in Standard 4 had a mean score of 75.3 percent, which shows improvement from Standard 2 but is below the Standard 3 benchmark (Standard 3 learners should correctly answer 80 percent of the questions). As stated above, despite a lower mean, the majority of learners in Standard 4 (63.6 percent) did meet the EGRA Coordinating Committee recommended Standard 3 benchmark for listening comprehension. Less than 5 percent of learners in both standards scored zero on the listening subtask.

Table 5. Percentage of Learners Reaching Listening Comprehension EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score	MoEST Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	5	60.6%	60% (Std. 1)	62.2%
4	5	75.3%	80% (Std. 3)	63.6%

Stage 2: Initial Reading and Decoding Skills

Letter Name Knowledge. This subtask is the first and most basic of learner reading assessments (not pre-reading skills). It measures learners' ability to say the names of the letters of the alphabet accurately. Automaticity and fluency of letter name knowledge is a predictive skill for later reading success. During the RA, learners were given a page of 100 randomly distributed upper- and lowercase letters and asked to state the names of as many letters as possible within one minute. The subtask is scored by the number of letters that

learners correctly name in one minute (correct letters per minute—clpm) out of a total of 100 possible letters.

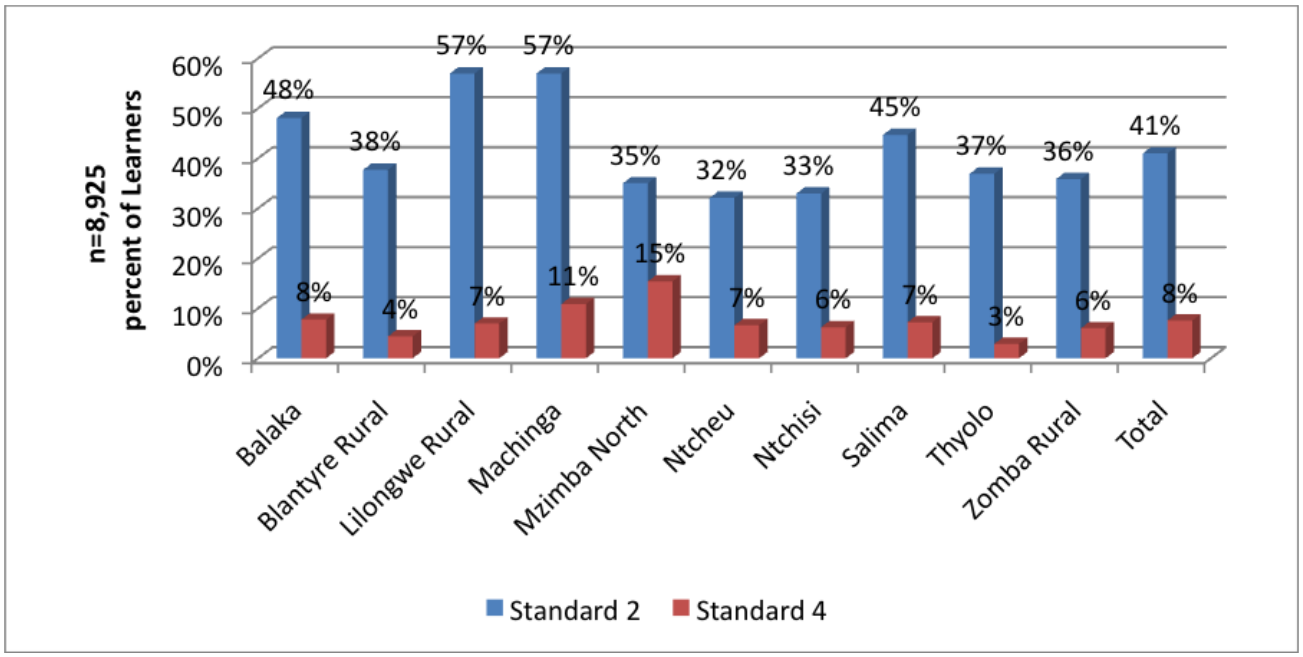
Mean scores in Table 6 show that learners in Standard 2 read an average of 9.58 letters correctly per minute, which is far below the Standard 1 benchmark of 24+ correct letters per minute. According to the benchmark achievements in Table 6, only 13.5 percent of learners in Standard 2 are meeting the benchmark of reading 24 or more correct letters per minute. Standard 4 learners show a mean score of 38.3 clpm, which is lower than the EGRA Coordinating Committee benchmark of 50 clpm. About 33 percent of Standard 4 learners met the Standard 3 EGRA Coordinating Committee benchmark, showing an increase in the percentage of learners meeting the benchmark from Standard 2 to Standard 4. Still, two-thirds of Standard 4 learners are unable to recognize letters of the alphabet at a pace necessary for oral reading fluency.

Table 6. Percentage of Learners Reaching Letter Name Knowledge EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score	EGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	100	9.58 clpm	24+ clpm (Std. 1)	13.5%
4	100	38.31 clpm	50 clpm (Std. 3)	33%

Figure 4 shows that 41 percent of learners in Standard 2 cannot name one single letter of the Chichewa alphabet. The districts with the highest proportion of zero scores in Standard 2 were Balaka (48.1 percent), Lilongwe Rural (57 percent), and Machinga (57 percent). In Standard 4, Machinga and Mzimba North are the only two districts where more than 10 percent of learners received zero scores. This could be because these are the only two districts where Chichewa is not the first language of most learners. A more in-depth examination of what might be causing these differences falls outside the scope of this report.

Figure 4. Percentage of Learners in Standards 2 and 4 Scoring Zero on Letter Name Knowledge



Syllable Segmentation. Research has found that phonemic awareness is the number one predictor of success in reading.[54] Phonemic awareness, a subset of phonological awareness, is the ability to identify sounds in words, to separate words into sounds, and to manipulate those sounds. In the EGRA instrument, phonemic awareness is tested in two ways: phoneme (syllable) segmentation and identification of onset and rhyme sounds (first and last sounds, respectively). In the syllable segmentation subtask, which is the first phonemic awareness measure, the RA assessor read aloud a word and asked learners to divide the word into syllables. There were a total of 10 words segmented into syllables in this subtask. As this subtask was untimed, the score reflects the number of items the learner answered correctly regardless of how long it took him or her to do so.

[54] RTI. (2009). *Early Standard Reading Toolkit*. Prepared for the World Bank Office of Human Development and USAID.

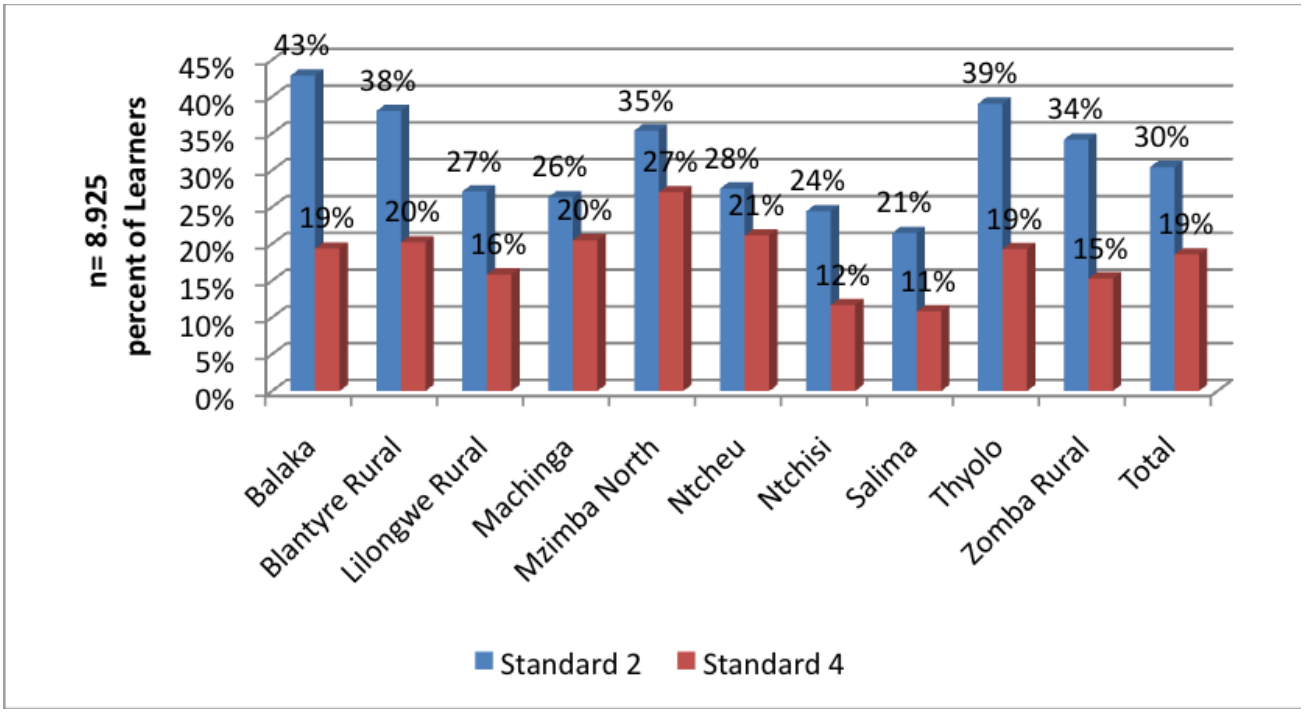
Table 7. Percentage of Learners Reaching Syllable Segmentation EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score (%)	EGRA Coordinating Committee Benchmark (%)	% Reaching Std. 1 and Std. 3 Benchmark
2	10	47.5%	70% (Std. 1)	45%
4	10	65.2%	80% (Std. 3)	57%

Table 7 shows the mean percent correct for Standard 2 learners is 47.5 percent, far below the Standard 1 benchmark of 70 percent. Standard 4 learners decoded an average of 65.2 percent of the syllables, which is also below the Standard 1 benchmark (as well as the Standard 3 benchmark of 80 percent). Less than half of the Standard 2 learners reached the Standard 1 benchmark (45 percent), and just more than half (57 percent) of Standard 4 learners met the Standard 3 benchmark.

Figure 5 shows that 30 percent of learners in Standard 2 and 19 percent of learners in Standard 4 could not read even one syllable. There were five districts where more than one third of Standard 2 learners received zero scores. In four out of five of these districts, between 19 and 21 percent of Standard 4 learners received zero scores.

Figure 5. Percentage of Learners Scoring Zero on Syllable Segmentation



Initial Sound Identification. Initial sound identification is the second measure of phonemic awareness (the understanding that words are made of sounds and the sound of a letter is not the same as the name of the letter). This subtask measures learners' ability to hear and isolate the first sound in a word. The test is administered in a similar manner to the syllable segmentation subtask. The learner is read a word aloud and asked to identify the first sound, or phoneme, in the word. This test is also untimed and has 10 word segments for identification. It is measured according to the number of items that the learner answered correctly.

The phonemic awareness subtask was difficult for many learners, as shown in Table 8. Far from the 80 percent benchmark (8 correct sounds out of 10), results for this subtask showed that, on average, Standard 2 learners could identify only 13 percent of initial sounds correctly (just more than 1 correct out of 10 possible). The mean score for Standard 4 learners was 21.4 percent (just more than 2 correct out of 10). The combined results show that learners have either not been taught phonemic awareness or do not yet understand it, and there is virtually no progress between Standard 2 and 4.

Table 8 shows the percentage of learners achieving EGRA Coordinating Committee benchmarks for Initial Sound Identification. As can be expected due to the very low mean scores, less than 10 percent of learners in both standards are meeting the EGRA Coordinating Committee benchmarks. The difference in results between syllable segmentation and initial sound identification is striking, given that 45 and 57 percent of learners in Standards 2 and 4, respectively, could segment words into syllables but less than 8 percent could identify the first sound of a word.

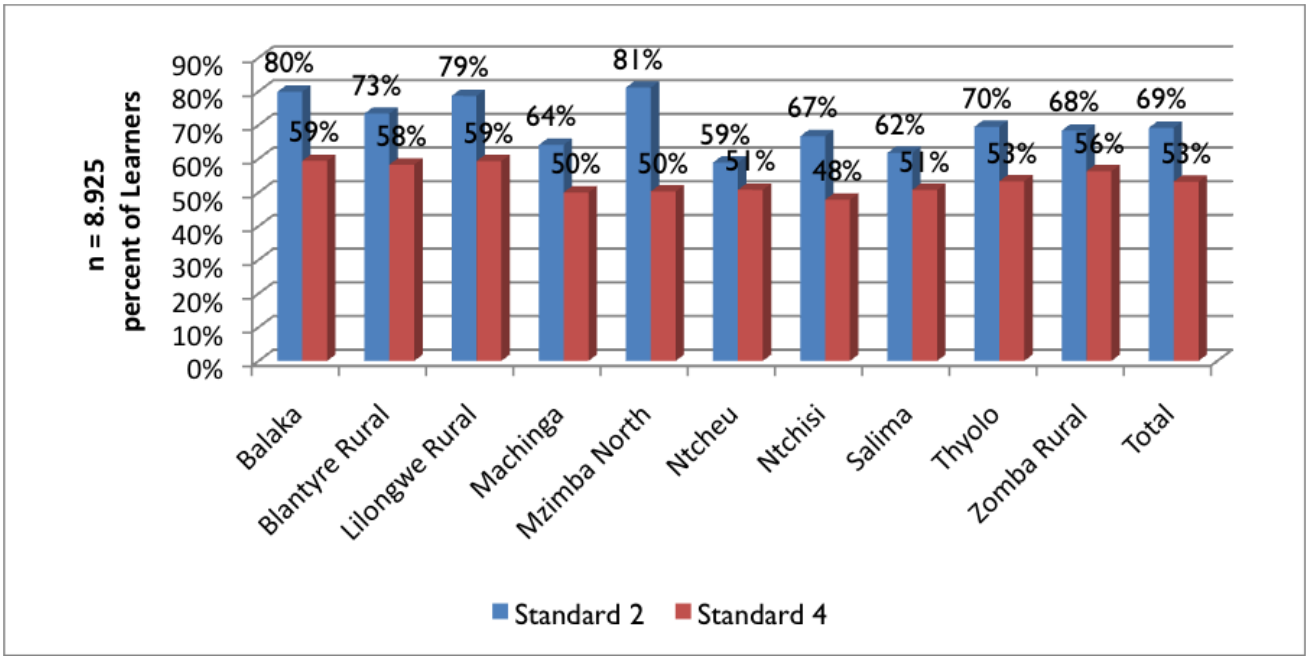
Table 8. Percentage of Learners Reaching Initial Sound Identification EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score (%)	EGRA Coordinating Committee Benchmark (%)	% Reaching Std. 1 and Std. 3 Benchmark
2	10	13.4%	80% (Std. 1)	5.3%
4	10	21.4%	90% (Std. 3)	7.1%

Figure 6 shows that 69 percent of Standard 2 learners scored zero across all districts, and about 80 percent of Standard 2 learners in Mzimba North (81 percent), Balaka (80 percent), and Lilongwe Rural (79 percent) could not identify even a single initial sound.

Overall, 53 percent of Standard 4 learners received zero scores on initial sound identification. In Standard 4, roughly 60 percent of learners scored zero in three districts: Balaka (59 percent), Blantyre Rural (58 percent), and Lilongwe Rural (59 percent). These percentages are higher than expected because, according to the EGRA Coordinating Committee benchmarks, learners should correctly identify 80 percent of initial sounds by Standard 1 and 90 percent of initial sounds by Standard 3.

Figure 6. Percentage of Learners Scoring Zero on Initial Sound Identification



There are several possible explanations for the poor scores achieved by learners for this subtask. Since teachers use the syllabic method of teaching, learners are taught to segment words into syllables and, thus, it may be easier for them to do that than to isolate the initial sound of a word. Also, initial sound identification is not a skill that is currently taught in Malawi classrooms, and the RA may have been the first time that learners were asked to do this. Therefore, learners (and assessors who have not learned these skills themselves and struggled to master the sounds during enumerator training) may have found it difficult to differentiate the first sound of the word.[55] It is also important to note that both phonological awareness tests are administered orally and sequentially. Because this task follows the syllable

[55] SI staff noticed during the supervisor and enumerator training sessions that many of the enumerators for the reading assessment did not know what a syllable was and had trouble identifying the correct syllable even when it was explained. This, of course, makes sense if they themselves were not taught phonics, but signals that more time should be devoted to training in phonics for the next round of data collection.

segmentation task, it may be that learners were responding with syllable sounds instead of phonemes (i.e., "ka" in the case of *kala*, when the correct answer should have been /k/).

Syllable Reading. In the Syllable Reading subtask, children were given a table of 100 randomly ordered common syllables and asked to read as many syllables as possible within one minute, such as "mi", "po," or "mle." The test is scored suing the number of correct syllables read per minute (cspm).

The syllable-reading subtask was added to the original EGRA instrument because the Malawian language experts advised that Chichewa is considered to be syllabic in nature and syllable reading is believed to be essential for language acquisition, although this theory has not been substantiated by research. The language experts reported that Chichewa-language instruction usually involves teaching learners to read groups of syllables (ma-me-mi-mo-mu) instead of letter sounds and then joining them together to form words.[56] This is called the "syllabic method" of language teaching.

If this theory is valid, it can be assumed that learners who are taught this technique would read syllables with automaticity and faster than whole words, thus scoring higher on this task versus the familiar or non-word reading subtasks.

However, the data suggests that if this method is used, it is likely used later than Standard 2, as learners at Standard 2 only got an average of 6.2 syllables correct out of 100, as shown in Table 9**.** There is quite an improvement at Standard 4, with learners reading 36.2 syllables. Despite these improvements from Standard 2 to Standard 4, learners from both standards are performing significantly below the EGRA Coordinating Committee benchmarks. Only 7.4 percent of Standard 2 learners met Standard 1 EGRA Coordinating Committee benchmarks and only 19.7 percent of Standard 4 learners met the Standard 3 EGRA Coordinating Committee benchmarks.

Table 9. Percentage of Learners Reaching Syllable Reading EGRA Coordinating Committee benchmarks

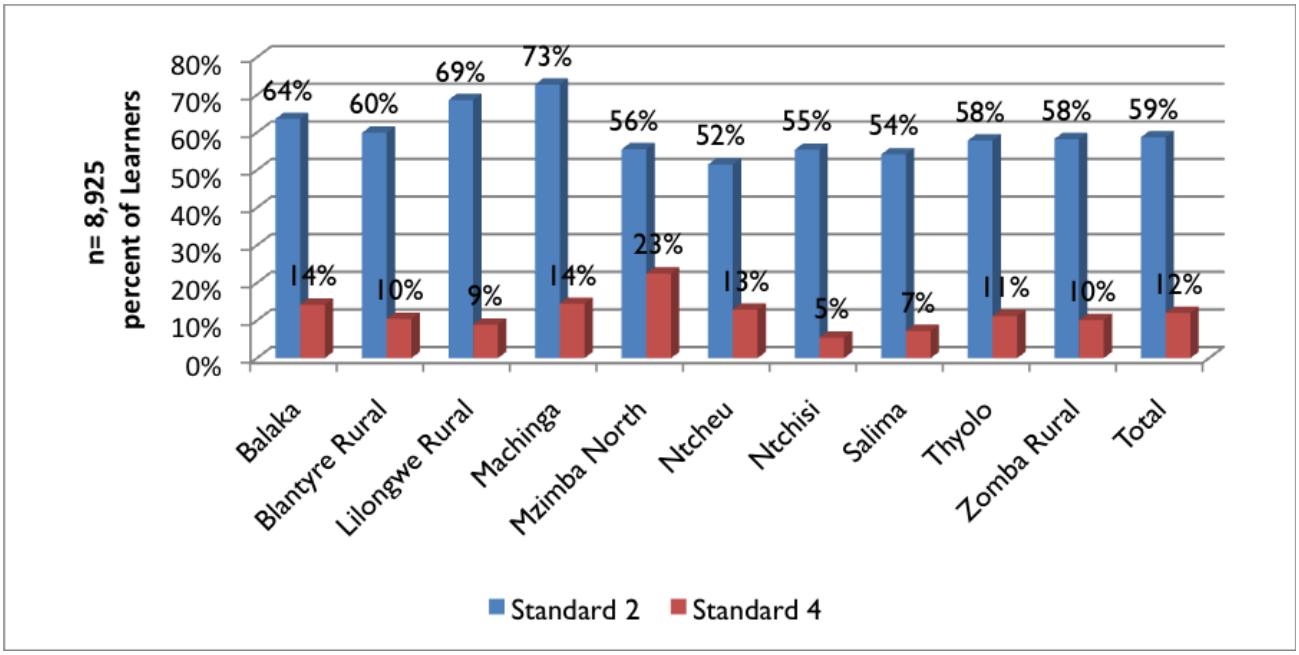
Standard	No. of Items	Mean Score	MEGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	100	6.2 cspm	30 cspm (Std. 1)	7.4%

[56] *Ibid.*

4	100	36.2 cspm	60 cspm (Std. 3)	19.7%

The low achievement in the syllable reading subtask is further evidenced by the high percentage of zero scores in Standard 2, shown in Figure 7 (59 percent). The highest proportion of non-syllable readers is in four districts: Machinga (73 percent), Lilongwe Rural (69 percent), Balaka (64 percent) and Blantyre Rural (60 percent). Only 12 percent of Standard 4 learners were not able to read a single syllable during the baseline RA.

Figure 7. Percentage of Learners Scoring Zero on Syllable Reading



Familiar Word Reading. Learners' decoding skills are often assessed using reading lists of unrelated words. This allows for a purer measure of word recognition and decoding skills than does reading comprehension paragraphs, as children are unable to guess the next word from the context. In the familiar-word reading subtask, learners were given a list of 50 common, simple words with instructions to read as many as they could in one minute (e.g., *atate*, *chiwala*, *zovala*). This assesses whether children can process words quickly and with accuracy, both of which are necessary to become fluent readers and to read to learn. Familiar-word reading is a timed test scored by number of correct words read per minute (cwpm).

For learners to achieve fluency EGRA Coordinating Committee benchmarks of 20 cwpm in Standard 1 and 50 cwpm in Standard 3, learners should be reading one correct word every three seconds in Standard 1 and one correct word every 1.2 seconds in Standard 3. Mean scores in Table I0 show that learners in Standard 2 are reading an average of four correct

words per minute, or 15 seconds per correct word. The pace improves to 2.5 seconds per correct word in Standard 4, but this is still below the target of 1.3 seconds per correct word. This indicates that the average learner in Standard 4 is reading too slowly to achieve fluency EGRA Coordinating Committee benchmarks required for memory recall and comprehension.

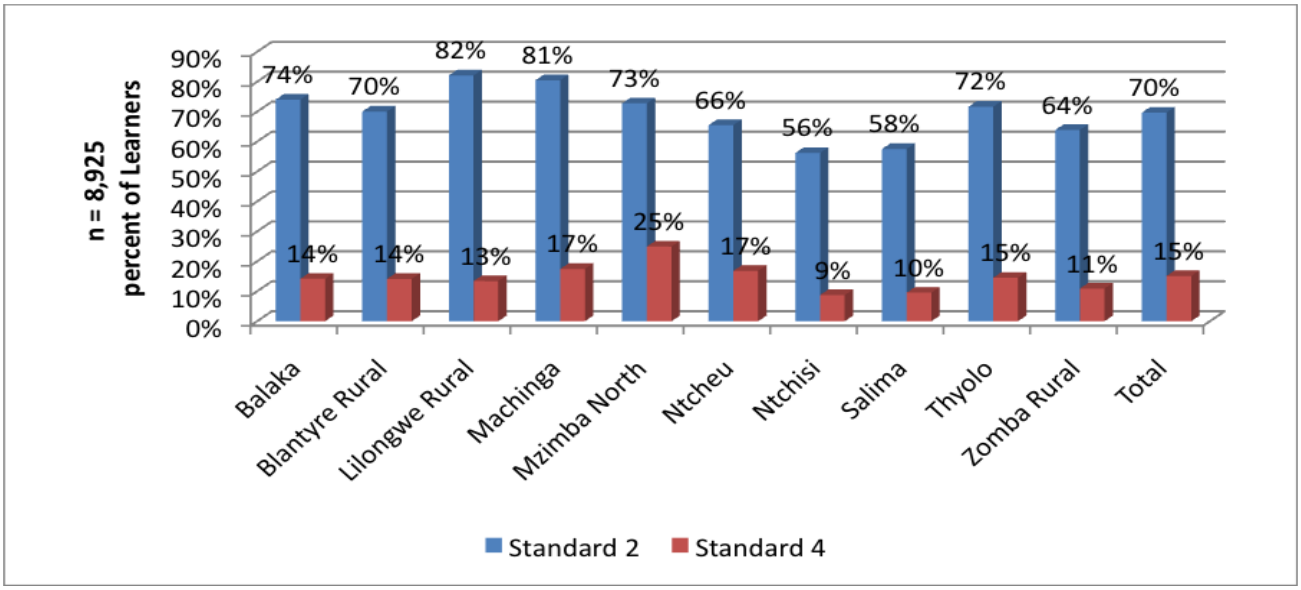
As seen in Table 10, only 7.7 percent of Standard 2 learners and 12.5 percent of Standard 4 learners are reaching the reading fluency EGRA Coordinating Committee benchmarks for familiar-word reading.

Table 10. Percentage of Learners Reaching Familiar-Word Reading EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score	EGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	100	4.0 cwpm	20 cwpm (Std. 1)	7.7%
4	100	24.87 cwpm	45 cwpm (Std. 3)	12.5%

The poor reading performance on this subtask is further illustrated by a large number of zero scores shown in Figure 8. In Standard 2, where learners should be reading approximately 32 words per minute (calculated as the midpoint between Standard 1 and Standard 3 EGRA Coordinating Committee benchmarks), 70 percent of the children tested across 10 districts are not able to read a single word. Again, the districts with the highest percentages of non-readers are Machinga (81 percent), Lilongwe (82 percent), Balaka (74 percent), and Blantyre Rural (74 percent). These are also the same districts with high percentages of zero scores on the other initial reading subtasks (letter name recognition, syllable segmentation, initial sound identification, and syllable reading). The district with markedly lower results in Standard 4 is Mzimba North, where 25 percent of learners could not read even one word. Considering learners should be reading between 45 to 60 words per minute in Standard 4, these scores are very low.

Figure 8. Percentage of Learners Scoring Zero on Familiar-Word Reading



Non-Word Reading. Non-word reading (also sometimes called invented word reading or nonsense word reading) is a measure of decoding ability and is designed to avoid the problem of sight recognition of words. Many children in the early standards learn to memorize or recognize by sight a broad range of words. To be successful readers, children must combine both decoding and sight-recognition skills. Tests that do not include a decoding exercise can overestimate children's ability to read unfamiliar words, as the words tested may be part of learners' sight-recognition vocabulary. Therefore, this subtask measures learners' ability to decipher "words" that follow the linguistic rules but do not actually exist in Chichewa. Learners were provided with a table of 50 made-up words and asked to read as many words as they could in one minute. The non-word reading test is a timed test measured by the number of correct non-words read per minute (cwpm).

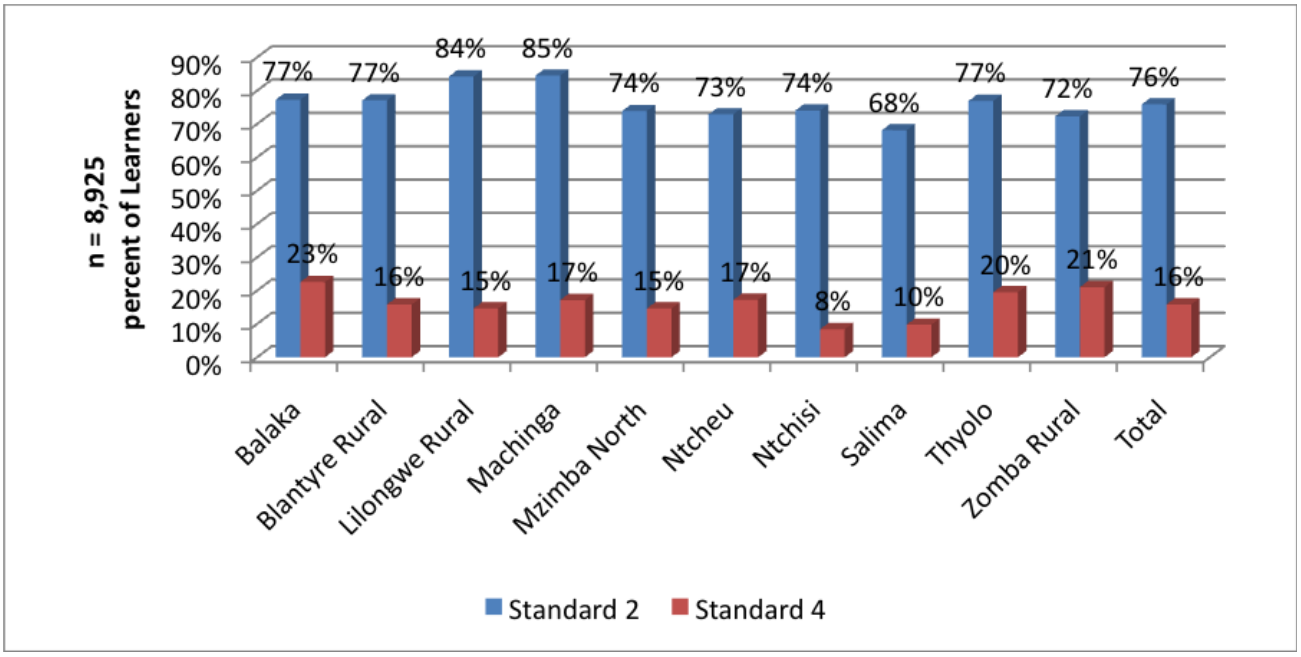
Table 11 shows that out of 50 possible non-words, Standard 2 learners could read an average of 2.57 correct non-words per minute and Standard 4 learners 15.78 correct non-words per minute. This table also shows that 6 percent of Standard 2 learners reached the Standard 1 benchmark of 15 or more correct non-words per minute, while in Standard 4 only 2.5 percent reached the higher Standard 3 benchmark of 40 cwpm.

Figure 9 shows that an average of 76 percent of Standard 2 learners tested were not able to read a single non-word. More than 84 percent of learners in the districts of Lilongwe Rural and Machinga, and 77 percent in Balaka and Blantyre Rural, scored zero. In Standard 4, more than 20 percent of learners in Machinga, Mzimba North, Blantyre Rural, and Ntcheu earned zero scores. This overall poor performance indicates that learners lack key decoding skills and, thus, have very limited vocabularies.

Table 11. Percentage of Learners Reaching Non-Word Reading EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score	EGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	100	2.57 cwpm	15 cwpm (Std. 1)	6.0%
4	100	15.78 cwpm	40 cwpm (Std. 3)	2.5%

Figure 9. Percentage of Learners Scoring Zero on Non-Word Reading



Stage 3: Confirmation and Fluency

Oral Reading Fluency. Oral reading fluency is a measure of overall reading competence. It is defined by the ability to translate letters into sounds, unify sounds into words, process connections, relate text to meaning, and make inferences to fill in missing information.[57] As skilled readers translate text into spoken language, they combine these tasks in a seemingly

[57] Hasbrouck, J., & Tindal, G. A. (2006). "Oral reading fluency norms: A valuable assessment tool for reading teachers." *The Reading Teacher, 59*(7), 636–644.

effortless manner, otherwise known as automaticity. Because oral reading fluency captures this complex process, it can be used to characterize overall reading competency. Tests of oral reading fluency have been shown to have a strong correlation with comprehension on more complex assessments. For example, there is a correlation of 0.91 with the Reading Comprehension subtask of the Stanford Achievement Test (SAT 10).[58,59] Poor performance on a reading comprehension tool suggests that the learner had trouble with decoding, with reading fluently enough to comprehend, or with vocabulary. In this subtask, learners were asked to read a very short story. The assessor stopped them after one minute and recorded the number of words read correctly (cwpm). Subsequently, learners were asked to answer comprehension questions about the story.

Table 12 shows that learners are failing to read with accuracy and fluency. An overall mean score of 15.25 cwpm reveals that learners sampled are reading below EGRA Coordinating Committee benchmarks set for Standard 1. In Standard 2, learners read an average of 3.77 words per minute. This is a slower pace than the familiar word mean score of four words per minute. However, a slightly higher percentage of learners met the benchmark on this task (8.4 percent in Standard 2 versus 7.7 percent on the familiar-word task). This result could be due to the connected text of related words in the passage, which allows learners to guess the next word from the context. Nonetheless, more than 90 percent of learners in Standard 2 did not meet the Standard 1 benchmark. In Standard 4, only 10.8 percent of learners met the benchmark in oral reading fluency while 89.2 percent are reading below the Standard 3 benchmark. Overall results for the fluency tests shown in Table 12 demonstrate that nearly 90 percent of learners in both standards are performing below standard level EGRA Coordinating Committee benchmarks. Figure 10 shows that 75 percent of pupils in Standard 2 and 20 percent in Standard 4 in nearly all districts tested are unable to read even a single word in a passage of Chichewa text. Figures 11 and 12 show the distribution of reading assessment scores. The majority of Standard 2 learners received zero scores, while the scores of Standard 4 learners are relatively equally distributed.

---

[58] Fuchs, L., Fuchs, D., Hosp, M. K., & Jenkins, J. (2001). "Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis." *Scientific Studies of Reading, 5*(3), 239–256.

[59] The Stanford Achievement Test Series, usually referred to simply as the "SAT 10," are standardized achievement tests utilized by school districts in the United States and in American schools abroad, such as the American International School of Johannesburg (AISJ), for assessing children from kindergarten through high school, according to the Wikipedia page located at: http://en.wikipedia.org/wiki/Stanford_Achievement_Test_Series.

Table 12. Percentage of Learners Reaching Oral Reading Fluency EGRA Coordinating Committee benchmarks

Standard	No. of Words in Story	Mean Score	EGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	55	3.77 cwpm	20 cwpm (Std. 1)	8.4%
4	55	26.62 cwpm	50 cwpm (Std. 3)	10.8%

Figure 10. Percentage of Learners Scoring Zero on Oral Reading Fluency

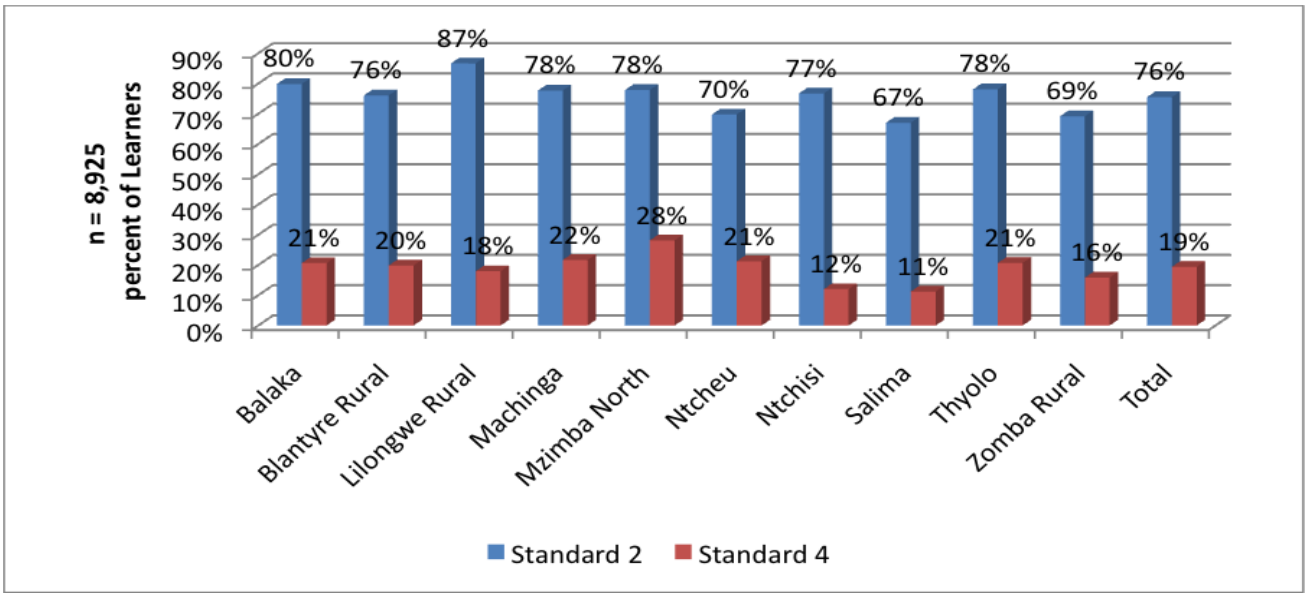


Figure 11. Percentage of Standard 2 Learners Receiving the Following Range of Oral Reading Fluency Scores

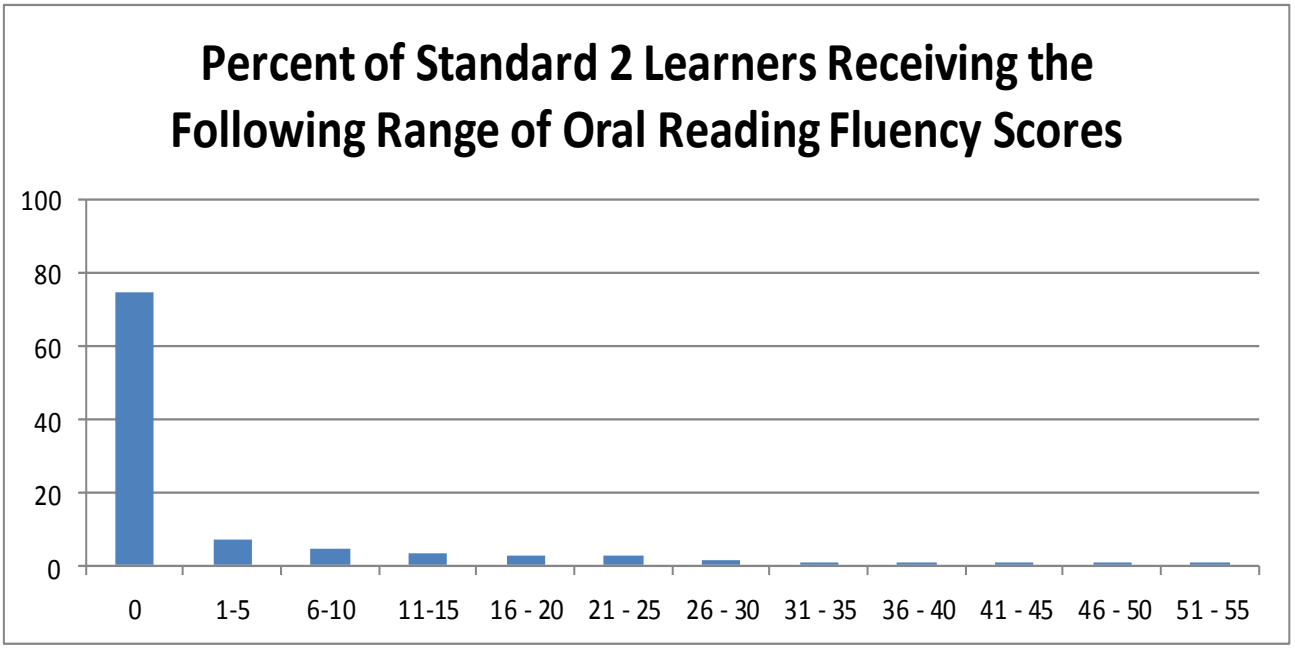
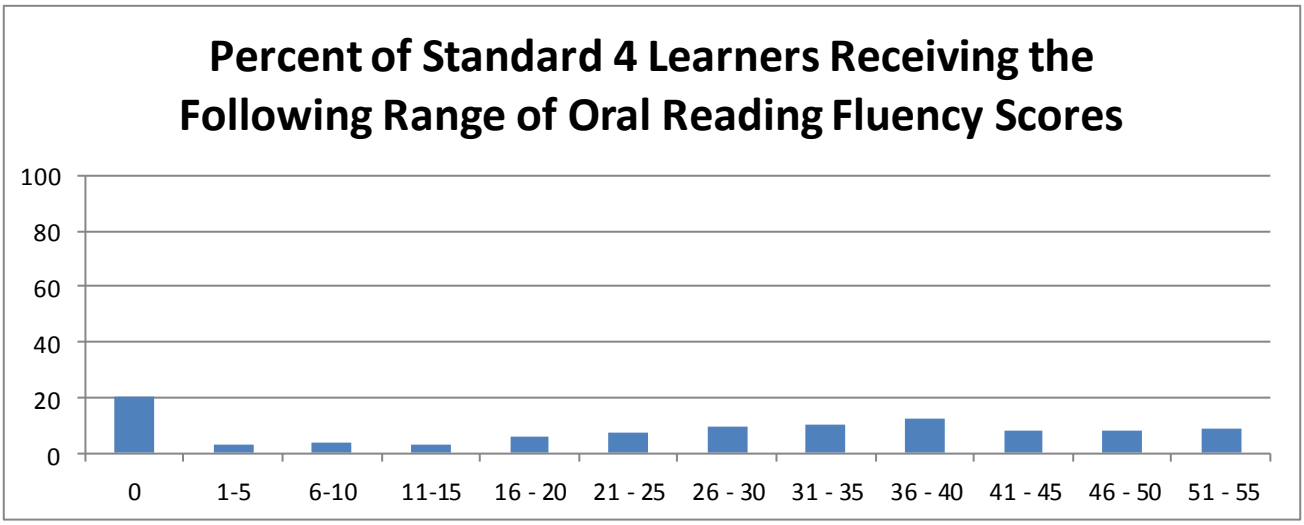


Figure 12. Percentage of Standard 4 Learners Receiving the Following Range of Oral Reading Fluency Scores



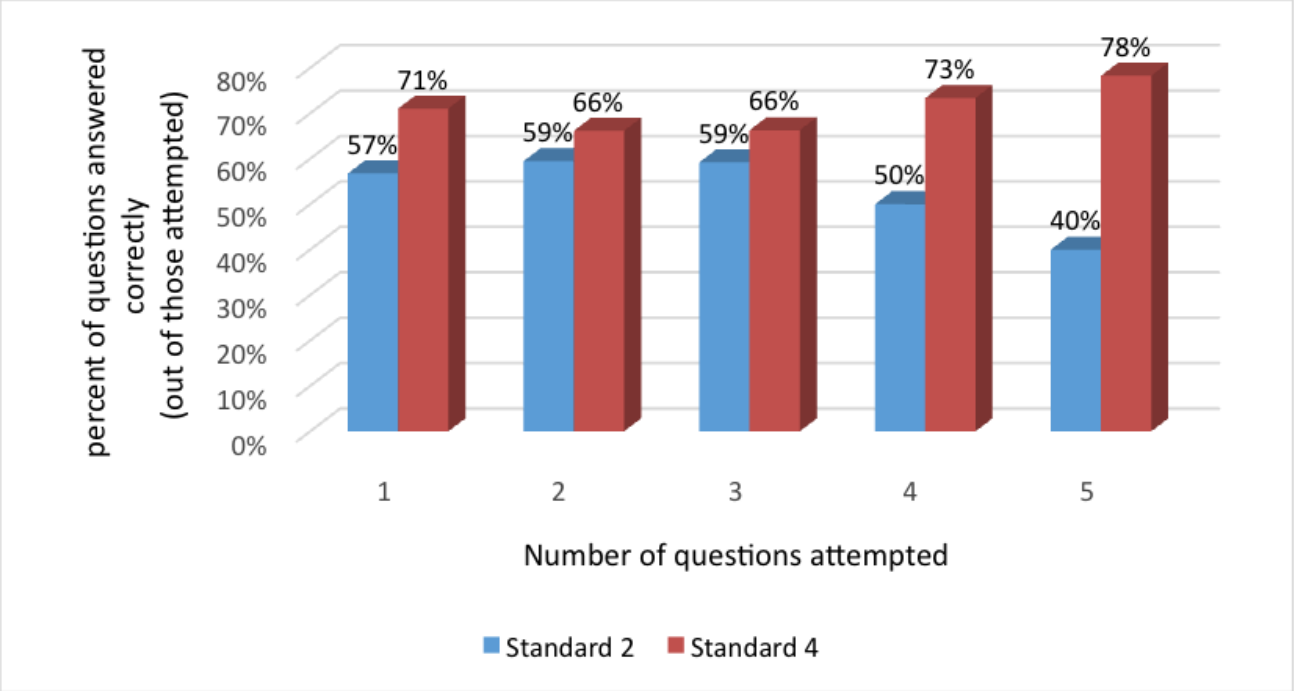
Reading Comprehension. The reading comprehension subtask measures the ability to answer several comprehension questions based on the passage read. There are a total of five questions for this subtask, consisting of direct, fact-based questions and at least one question requiring inference from the passage read. Learners were asked questions only up to the point where they had stopped reading. For instance, if the child read only half the text, he or she would be asked the questions that could be answered by half the passage (two or three

questions). Likewise, if learners did not read any of the text, they were not asked any questions and would score zero on reading comprehension. The reading comprehension score was derived by taking the percentage of questions answered correctly out of the total possible (five).

In Standard 2, almost 80 percent of the learners were unable to read enough of the passage to be asked even a single comprehension question, while of the almost 13 percent who read enough to be asked one question, only about 58 percent were able to answer that question correctly. Only about 2.5 percent read enough of the passage to be asked three to five questions, and of the almost 2 percent asked three questions, less than 30 percent were able to answer all three correctly.

For a more accurate measure, Figure 13 shows total percentage of questions answered correctly based on total attempted. Thus, this is a true measure of the children's ability to comprehend what they read according to their current fluency rate. For example, of the learners who attempted question number four, 50 percent in Standard 2 and 73 percent in Standard 4 answered the question correctly. The difference in scores between actual attempted and total possible is significant. The results in Figure 13 show that learners are reading at a slower pace than required to complete the entire 54-word passage, but they comprehend between 40 to 78 percent of what they read. Nonetheless, these rates are not sufficient to meet expected EGRA Coordinating Committee benchmarks and suggest that learners are lagging behind curricular goals.

Figure 13. Distribution of Reading Comprehension Scores by Number of Questions Attempted



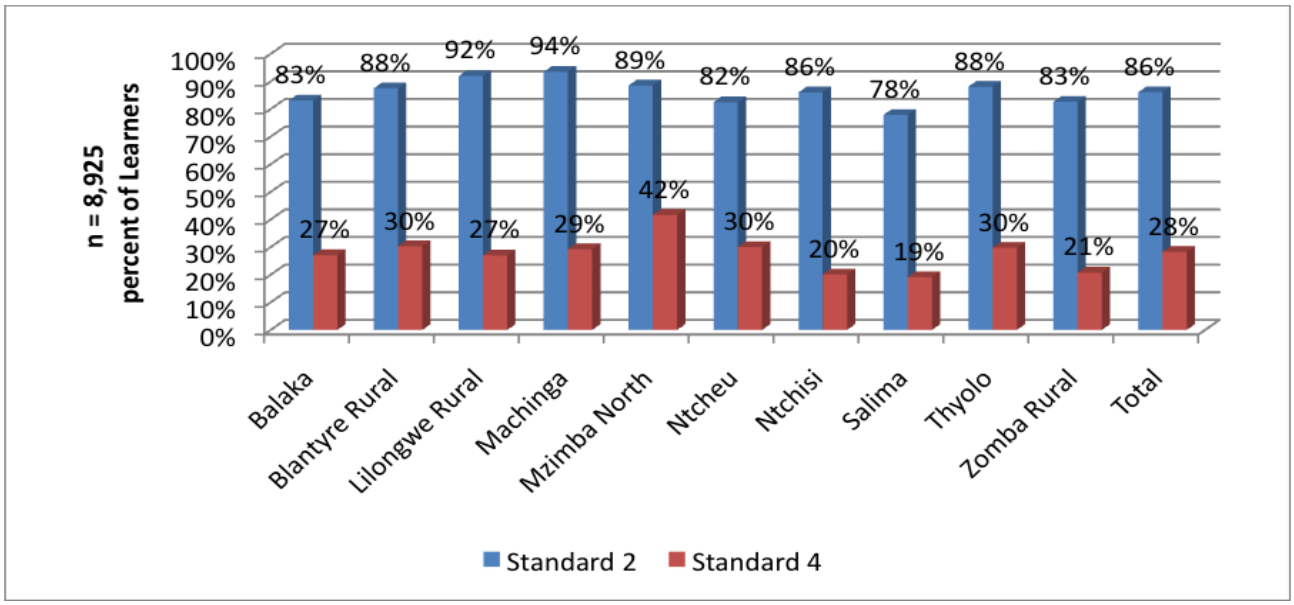
Reading comprehension EGRA Coordinating Committee benchmarks are based on mean score values out of total possible. As can be seen in Table 13, the percentage of learners attaining reading comprehension EGRA Coordinating Committee benchmarks is 3.4 percent in Standard 2 and 10.8 percent in Standard 4. This means that 97 percent of learners are unable to correctly respond to even two comprehension questions. In Standard 4, 89 percent of learners could not correctly answer four out of five basic comprehension questions. The combined mean score of 20.7 percent indicates the majority of learners in both standards are reading at half the expected level of a Standard 1 learner.

Figure 14 shows the proportion of pupils unable to answer one comprehension question. Eighty-six percent of learners across all districts sampled in Standard 2 showed no indication of having any reading comprehension skills. The highest percentages of zero scores were in Machinga (94 percent) and Lilongwe Rural (92 percent). At Standard 4, an average of 28 percent of learners scored zero. The district with the worst performance was Mzimba North, where 42 percent of Standard 4 learners tested could not correctly answer a single comprehension question.

Table 13. Percentage of Learners Reaching Reading Comprehension EGRA Coordinating Committee benchmarks

Standard	No. of Items	Mean Score (%)	EGRA Coordinating Committee Benchmark	% Reaching Std. 1 and Std. 3 Benchmark
2	5	4.6	40 (Std. 1)	3.4%
4	5	36.7	80 (Std. 3)	10.8%

Figure 14. Percentage of Learners Scoring Zero on Reading Comprehension



Gender and District Disaggregations

As shown in Figure 15, there was no significant difference in scores between boys and girls in Standard 2 for most subtasks. Boys' performance on listening comprehension was significantly, but marginally higher, than girls' scores (p-value < 0.01), and girls' performance was significantly higher on letter name knowledge (p-value < 0.05) than boys.

Similar results were found at the Standard 4 level (see Figure 16). Although there were slight differences in some subtasks, the only significant differences were in the boys' higher performance on listening comprehension (p-value < 0.01), and girls' higher performance on familiar-word reading (p-value < 0.0218), and oral reading (p-value < 0.05), although the overall scores for both sexes on oral reading were so low that significant differences here are not meaningful—under 10 percent in both standards.

Figure 15. Average Percent Correct by Subtask and Gender – Standard 2 Learners

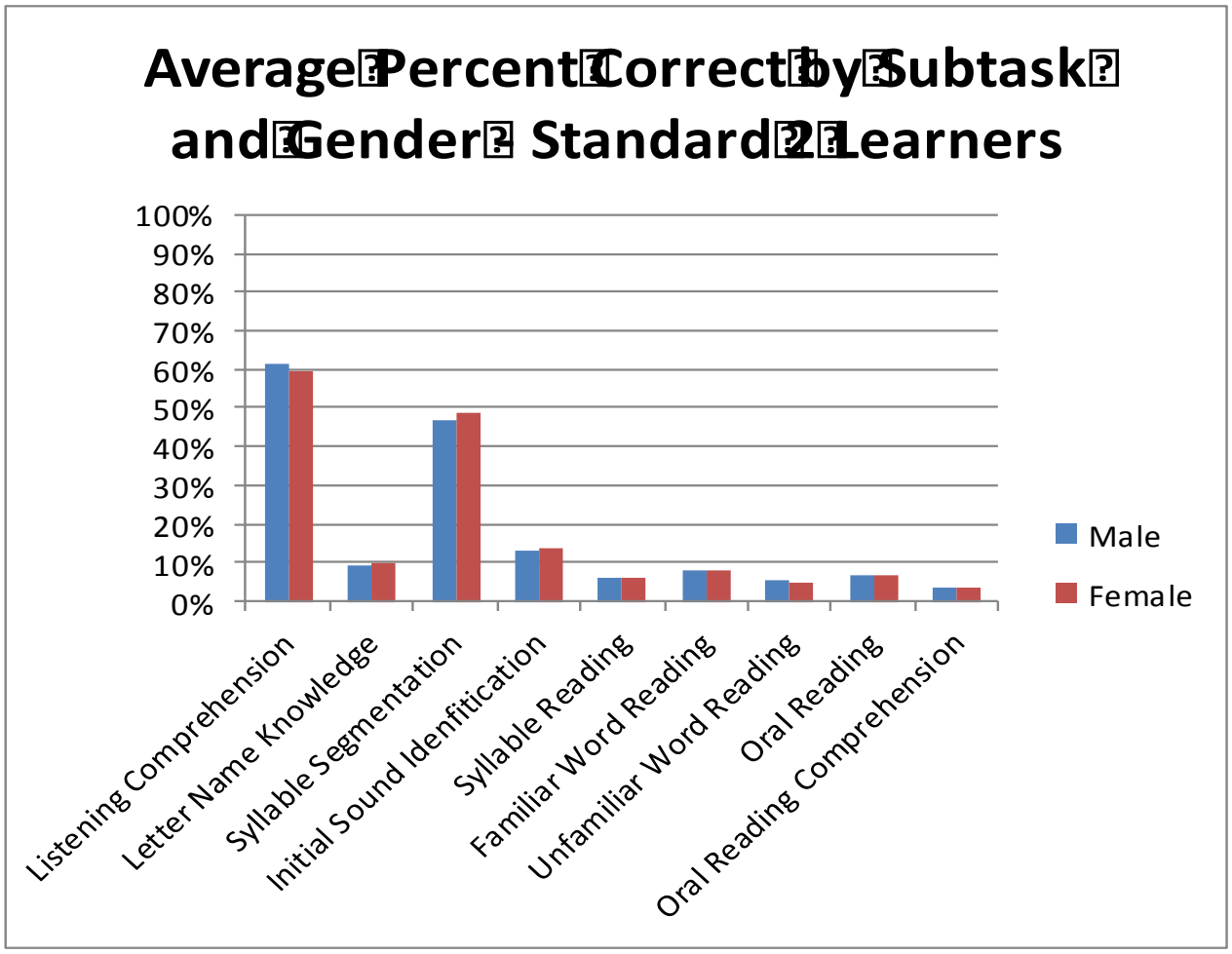
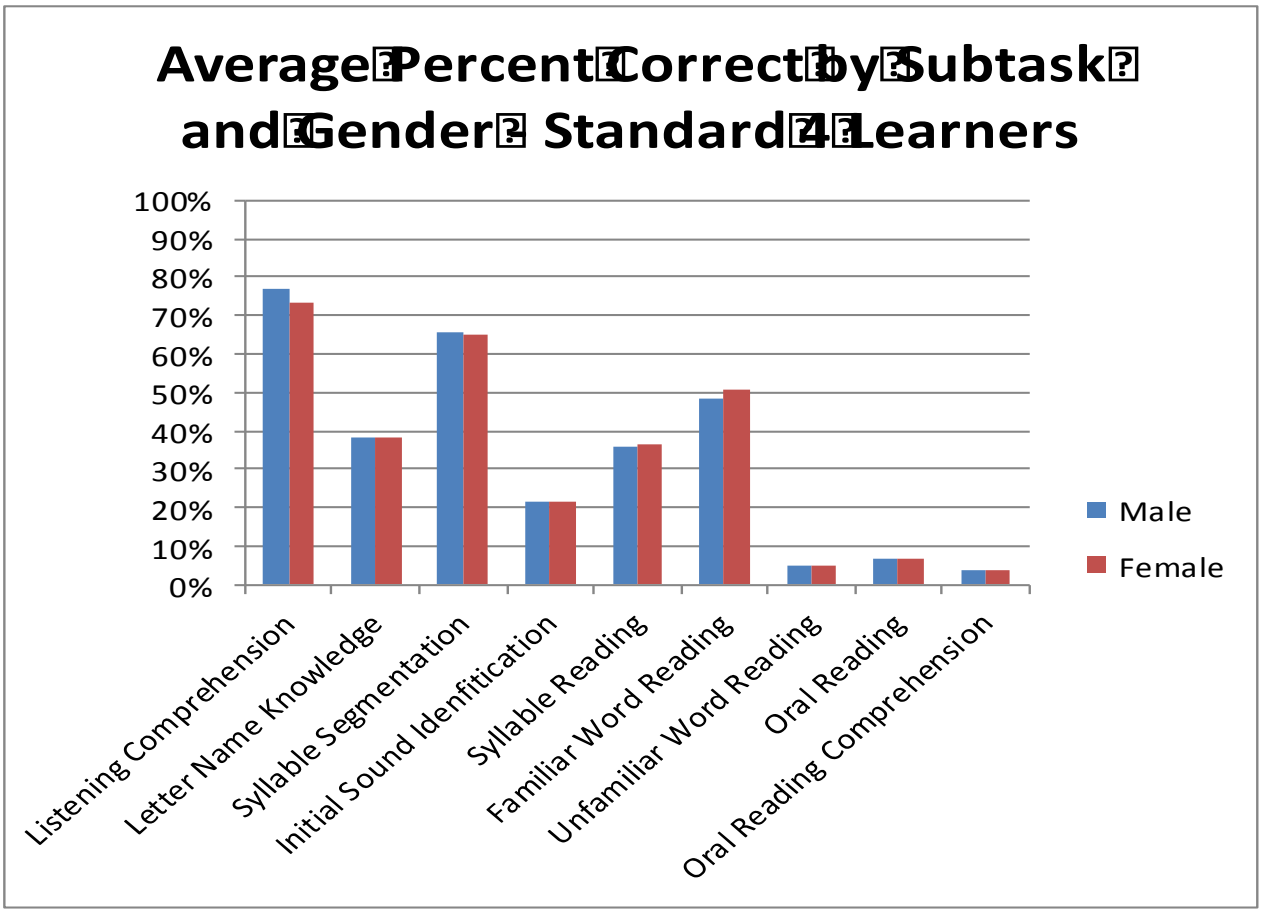


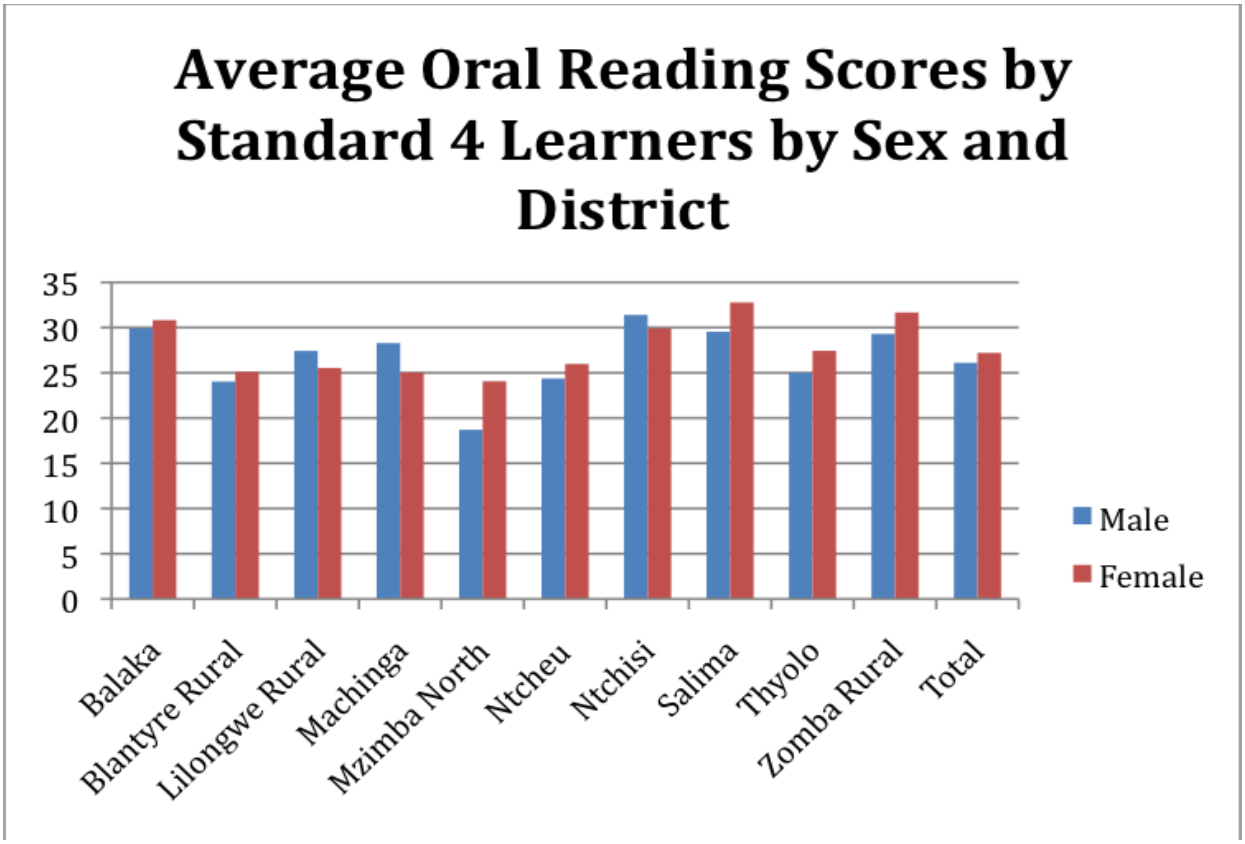
Figure 16. Average Percentage Correct by Subtask and Gender – Standard 4 Learners



Overall, it appears that boys did better on listening comprehension while girls did better on various other reading subtasks. It is unclear why girls did better on these particular subtasks, but it appears that sampled girls may be learning to read more quickly than boys.

To further explore the role of gender in test scores, the evaluation team also disaggregated gender scores within each district (see Figure 17). Before examining this, it is important to reiterate that the samples from each district are not representative of the district and the number of observations in each varies significantly. In some districts, girls do better than boys on all subtasks, in others results are mixed, and in yet others boys do better than girls on all subtasks. However, the overall scores are so consistently low that even statistically significant differences are not likely to be meaningful. In other words, substantial improvements are needed for almost all of the learners assessed.

Figure 17. Average Oral Reading Scores by Standard 4 Learners by Sex and District



English Reading Assessment Results

This section reports the results of the 994 learners tested in English. According to the MoEST policy, the language of classroom instruction transitions from Chichewa to English in Standard 5. This policy assumes that learners have mastered Chichewa and have achieved oral reading fluency in English so that they are fully prepared to learn in the English language by Standard 5. The objective of the English-language RA was to measure learners' ability in reading fluency and comprehension at the end of Standard 4 in order to determine their level of preparedness.

Research shows that mastering one language facilitates learning of a second language. Given the low level of Chichewa readers, it is not surprising that 99 percent of Standard 2 learners and 95 percent of Standard 4 learners could not read one word in the English language on the Oral Fluency Reading passage, as illustrated in Table 14. On average, Standard 2 learners could correctly read one word of an English story within one minute, but none could correctly answer one comprehension question. Standard 4 learners could read an average of 11 correct words per minute, but only 0.1 percent could answer one comprehension question. Virtually

100 percent of learners could not read with comprehension across both standards tested. Based on the high proportion of non-readers in English, it's not unreasonable to conclude the results are similar across all districts tested, and certainly are nowhere close to being sufficient to learn in English.

Table 14. English Subtask Mean Scores

Standard	Listening Compre-hension	Letter Naming	Initial Sound Identifi-cation	Familiar Word	Unfamiliar Word	Oral Reading Fluency	Reading Compre-hension
2	0.2	3.2	2.7	1.1	1.4	1	0
4	0.7	22.9	1.6	7.9	9.4	11.3	0.1

SELECTED SUMMARY STATISTICS OF THE SAMPLE POPULATION

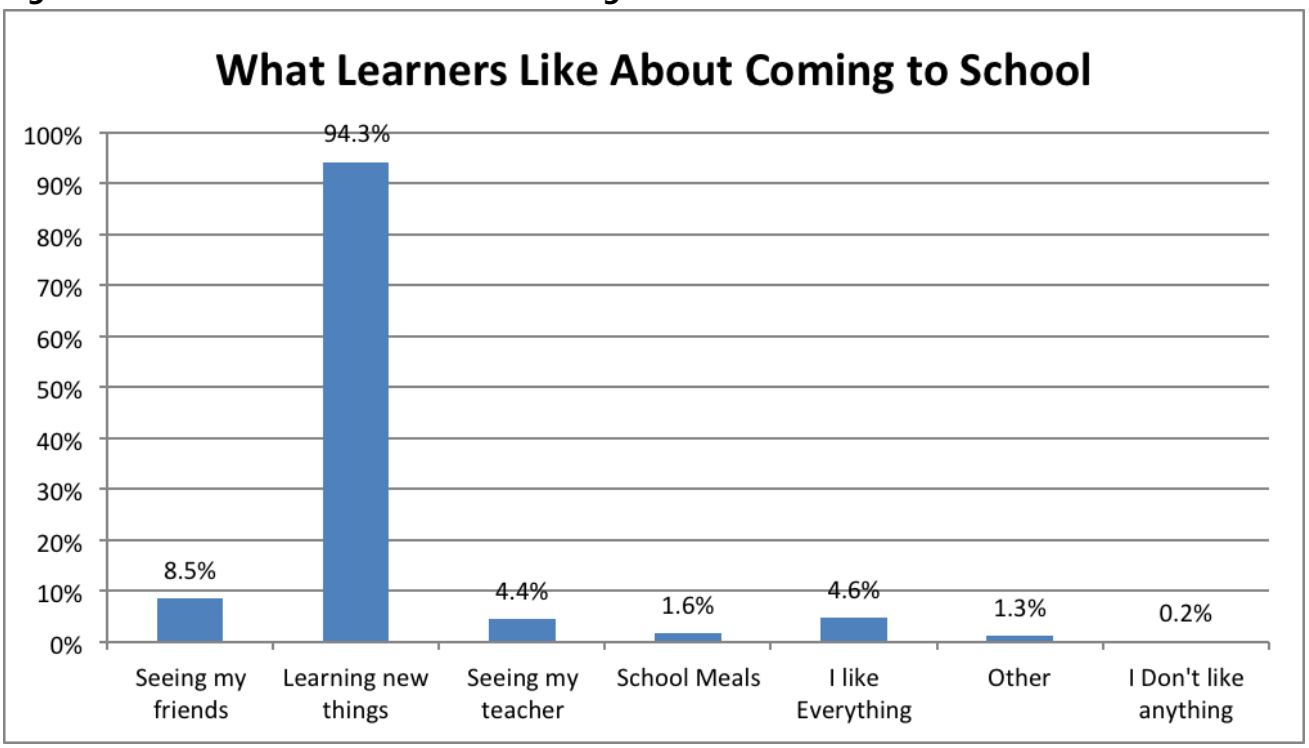
The next several sections provide summary statistics on factors that have been shown (in other studies, such as those described in the literature review above) to affect learner reading outcomes. This section does not attempt to tie these factors to learner reading outcomes from this study but instead presents a snapshot of the sample of learners and schools that this baseline study assessed to give reviewers an idea of what the study population looks like as a whole. The relationship between these factors and learning reading outcomes is explored in the section on Preliminary Responses to Task 2 Evaluation Questions beginning on page 81 of this report.

SELECTED FINDINGS FROM LEARNER QUESTIONNAIRE

Learner Attitudes toward School and Teachers. After completing the RA, each learner was asked to answer a collection of questions about his or her experience with and feelings about school. Even though the responses of children this young are generally less reliable than those of older children or adults, there was a high degree of internal consistency in their responses, giving evaluators greater confidence in the reliability of their responses overall. Overall, findings from the learner questionnaire suggest that most sampled learners have a positive view of their school and teacher. They like going to school, think they learn a lot, and they also like their teachers. Specifically, of the 8,060 learners interviewed, 98.5 percent reported that they like going to school. When asked an open-ended question about what they like about school, almost all sampled learners, (94.3 percent) said that they like learning new things. Some sampled learners also listed other things they like about school, but these

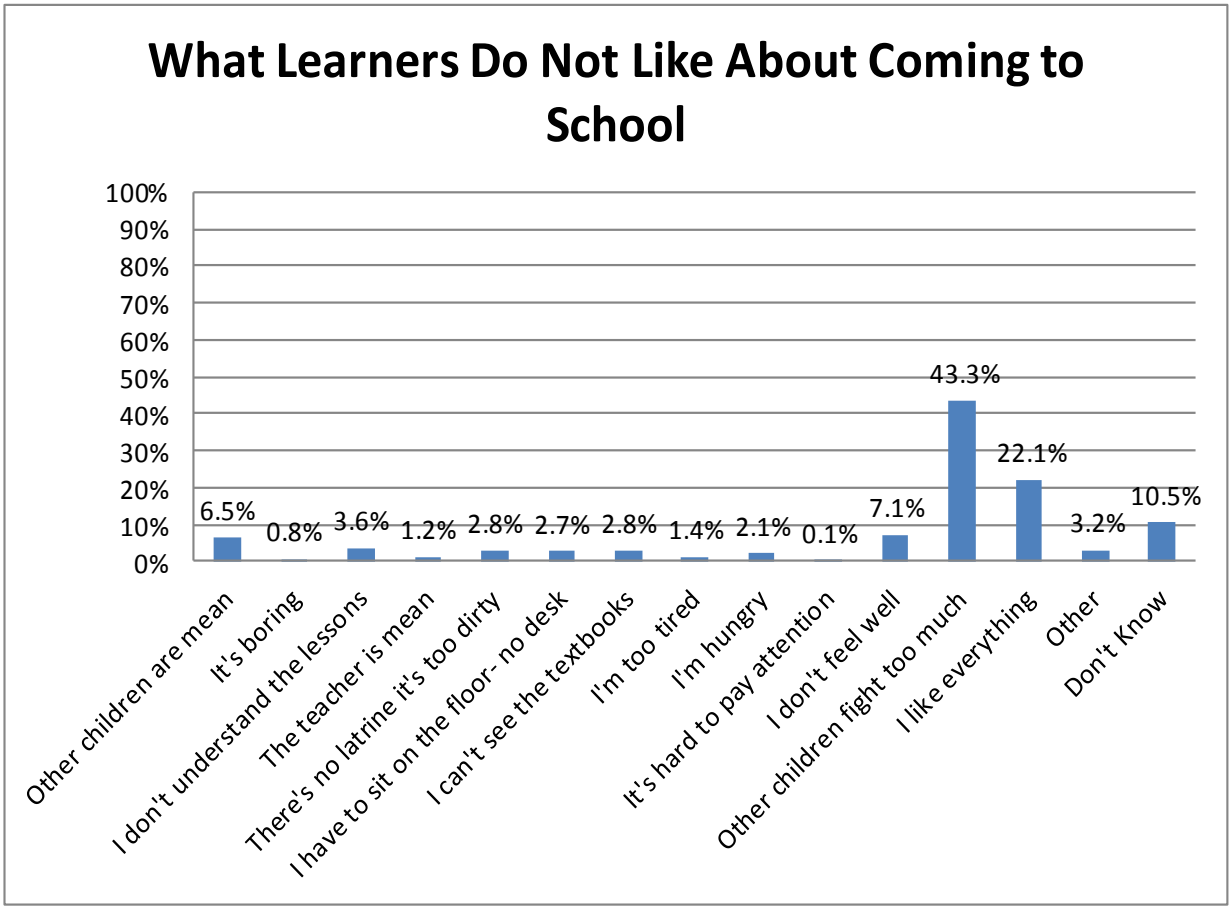
responses were far less common. For example, “seeing friends” was the second most popular response, with 8.5 percent expressing this sentiment. (See Figure 18 for more details.) There were very small differences between boys and girls with regard to these questions.

Figure 18. What Learners Like About Coming to School



In addition to asking what children like about going to school, enumerators also asked what they do not like about going to school (see Figure 19). The most common response was that “other children fight too much,” with 43.3 percent giving this response. The second most commonly expressed concern was that “other children are mean.” All totaled, 6.5 percent of sampled learners gave this answer, which is far less than the percentage of children who report getting “teased” at school (31 percent). A relatively small percentage of sampled learners, only 3.6 percent, reported that they do not like school because they do not understand the lessons. While not a category originally included on the questionnaire, a number of learners also said that they do not like when the teacher does not show up or does not teach them.

Figure 19. What Learners Do Not Like About Coming to School

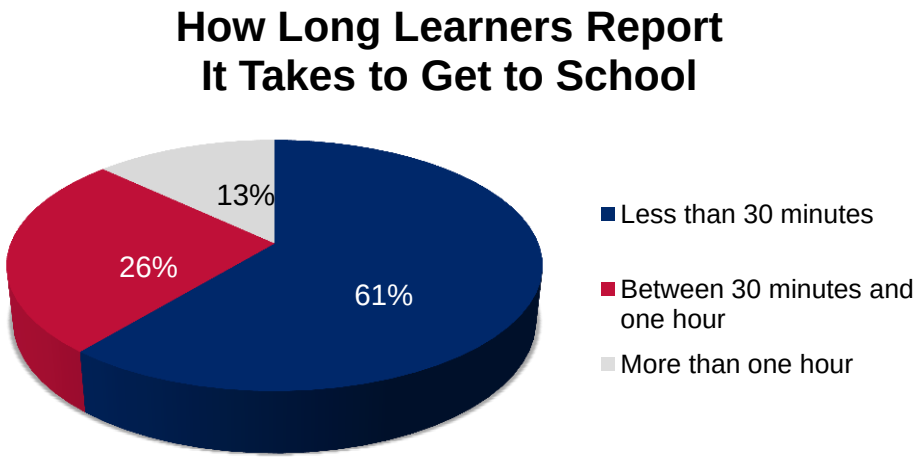


However, when asked *How much do you think you learn at school?*, 78.1 percent of sampled learners said that they learn "a lot." The majority of the remaining learners (just less than 11.5 percent) responded that they feel they either learn "not anything" or "not much."

When asked about their teachers, the majority of learners described them as "nice or happy." Another 11 percent said their teachers were "sometimes nice and sometimes not so nice," and about 1.2 percent described their teachers as mean. While few learners described their teachers as mean, about half of the sampled learners reported getting punished at school, with boys more likely to be punished than girls (52.4 percent versus 46.8 percent) and learners in Standard 4 more likely to be punished than learners in Standard 2 (60.7 percent compared to 38.7 percent). The main reason learners report getting punished is for coming to school late—23.2 percent of Standard 2 learners and 35.9 of Standard 4 learners of learners reported getting punished at school because they had arrived late. According to learners, the most likely punishment is "sweeping the classroom or school grounds," an answer give by 33 percent of Standard 2 learners who got punished and 48.5 percent of Standard 4 learners.

School Accessibility and Perceived Safety. Past education studies have shown that the more difficult or inconvenient it is for children to attend school, the less likely they are to do so. As such, another factor that evaluators considered is distance from school, measured by how long it takes learners to walk there. The evaluation team found that 60.9 percent of learners reported that their walk to school is less than 30 minutes (see Figure 20). Studies also show that children are less likely to attend school if they are not safe walking there. In fact, parents sometimes prevent children, and especially girls, from attending school if they live far from the school due to safety concerns. This is a problem especially in rural areas. Fortunately, findings from the learner questionnaire suggest that 89 percent of both boys and girls and in both grade levels feel safe walking to school. About 4.1 percent reported that the reason they do not feel safe is because of "bad men or boys" (*n* = 8,058).

Figure 20. How Long Learners Report It Takes to Get to School

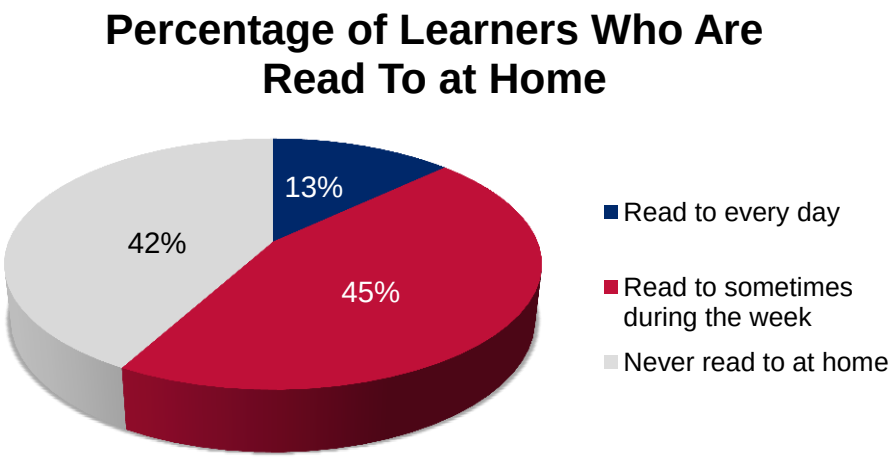


Almost three quarters of the learners feel comfortable using the latrine at school, with slightly more Standard 2 learners feeling comfortable compared with Standard 4 learners (72.1 percent versus 69.1 percent, *n* = 8,044). There was little difference between boys and girls. This is important because a number of studies have shown that a lack of clean and separate latrines for girls often deters them from going to school. The main reasons that learners gave for not being comfortable using them is that the latrines are dirty and smelly.

Reading and Homework Support. Generally, learners are more likely to perform better in school if they have books at home, are read to at home, and are helped with their homework. Among the learners interviewed for this study, 58.6 percent reported that someone in their household reads to them, with 12.4 percent saying they are read to every day, 44.3 percent saying that they are read to only sometimes or two to three times a week, and 41.7 percent saying they are never read to at home (*n* = 8,061). (See Figure 21.) Of those surveyed, 59.9

percent of learners said that they read on their own at home, with 81.6 percent of learners in Standard 4 reporting reading at home, compared with only 37.9 percent of Standard 2 learners reading at home on their own (*n* = 8,051). A much higher percentage of learners in Salima District reported reading on their own at home (67.1 percent) than did learners in overall, 59.9 percent, and particularly compared with a low of 53.0 percent in Lilongwe Rural (*n* = 8,051). Almost all of the Standard 4 learners reported liking to read at home (91.7 percent) compared with only 67.4 percent of Standard 2 learners (*n* = 8,038). This may be because they can read better by Standard 4.

Figure 21. Percentage of Learners Who Are Read To at Home



Slightly more girls reported that someone in their household helps them with homework (56.2 percent versus 52.6 percent for boys) and substantially more Standard 4 learners reported someone in the household helping them with homework (64.3 percent) than did Standard 2 learners (44.0 percent), although it is likely that fewer Standard 2 learners are assigned homework. The differences across districts was fairly substantial, with 62.3 percent of the learners from Mzimba North reporting that they receive help with homework, compared with lows of 47.4 percent and 45.9 percent for Lilongwe Rural and Machinga districts, respectively (*n* = 8,051). The average was 54.2 percent across all districts.

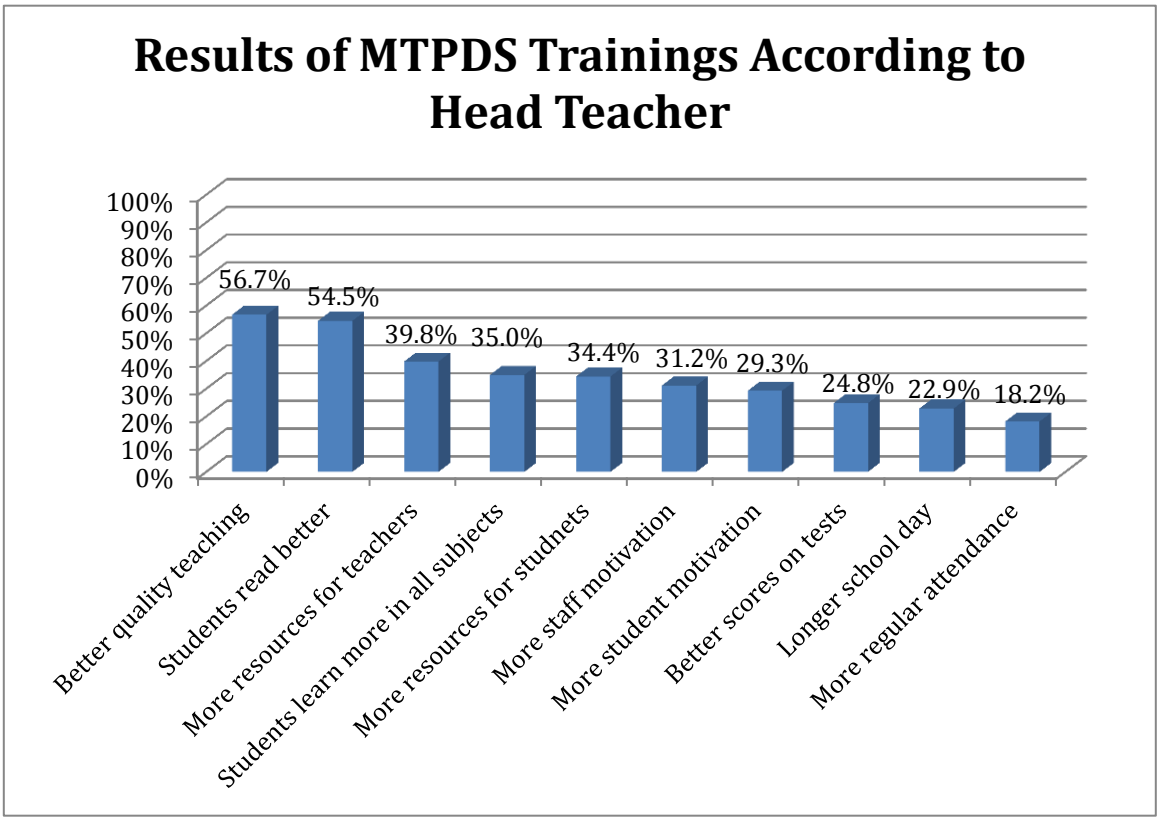
Health and Nutrition. Of learners surveyed, 78.9 percent reported eating breakfast every day and 95.7 percent said they ate lunch every day, normally at home (*n* = 8,053). About 4.4 percent of the learners reported not eating any food on three days of the week, and 68 learners reported not eating any food on six to seven days of the week. About 15 percent said they are hungry at school a few times a week, while 18 percent said they are hungry at school every day (*n* = 8,002). Of learners surveyed, 16.8 percent reported that they miss school because they are sick "a lot" (*n* = 7,875).

SELECTED FINDINGS FROM TEACHER AND HEAD TEACHER QUESTIONNAIRE

During the school visits, an interview was conducted with the Head Teacher (HT) or, when the HT was not available, with the Deputy Head Teacher (14 percent of the cases). Most of those interviewed were men (89 percent) and only about 11 percent were women. Only two of the HTs had certifications higher than the Malawi Schools Certificate of Excellence (MSCE), while 84 percent reported having acquired MSCE.

Training and Coaching. Almost 87 percent of the schools reported having received support from MTPDS, with the most frequently provided support being training of head teachers (65 percent) and teachers (58.6 percent), followed by 48.7 percent having received textbooks for use in classrooms. The outcomes resulting from MTPDS support as reported by head teachers are shown in Figure 22.

Figure 22. Results of MTPDS Trainings According to Head Teacher



Almost 70 percent of the HTs/DHTs reported having participated in training in how to provide instructional support to teachers and in how to teach reading using the phonics approach. On average, 48.9 percent of teachers in each school have had MTPDS training interventions. On average, HTs report that 46.9 percent of teachers in their school are confident about using the MTPDS intervention methods in their teaching and. Head teachers also on average suggested that 64 percent of teachers in each school need more training in these methods for teaching reading (*n* = 285 for this section).

Eighty percent of the Standard 2 and 4 teachers interviewed reported that they had participated in an MTPDS training intervention, but of those who did, 36 percent had participated in only one to three days of training, 34 percent in four to nine days and 11 percent in more than 10 days of training. However, some teachers had participated in training in teaching reading that was not provided by MTPDS. Thirty-two percent had participated in one to three days of reading training provided by another organization, 20 percent in four to nine days of training, and only 6 percent participated for more than 10 days.

It is important to remember that these findings represent the views of HTs and not any kind of objectively measured outcomes.

School Resources. HTs reported that, on average, only 13 percent of learners have the prescribed number of textbooks, and only 38 percent of the teachers felt that they have sufficient teaching and learning resources.

The following is the percentage of HTs or teachers who reported that their schools have the following basic resources:

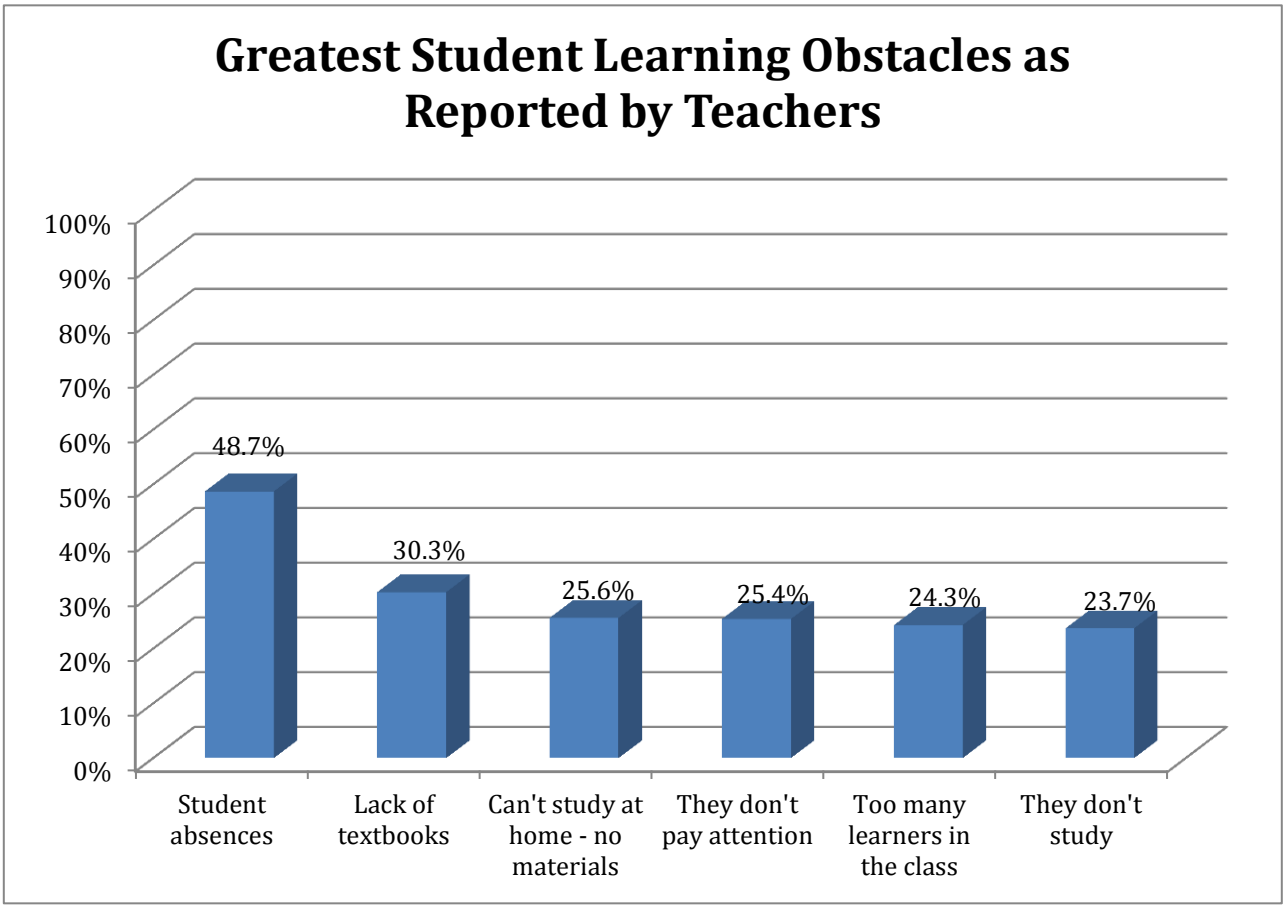
- Clean, safe water supply 73.7 percent
- Electricity 10.6 percent
- School feeding program 19.9 percent
- School libraries 7.3 percent
- Classroom libraries 5.1 percent

Teacher Tardiness and Absences. On the day of the school visit, 50 percent of HTs reported that at least some teachers often arrive late, with 33 percent reporting that at least one to three teachers often arrive late and 5.3 percent of HTs saying that four to five teachers often arrive late. Moreover, 57.7 percent of the HTs interviewed reported that at least one teacher was absent from school on the previous day, 33.4 percent reported that between one and three teachers were absent on the previous day, and 6 percent reported that four to seven teachers were absent. It can be difficult to encourage learners to arrive at school on time if teachers arrive late or not all.

Perceived Obstacles to Learning. As shown in Figure 23, teachers reported the greatest obstacles to learner learning to be the following:

- Learner absences 48.7 percent
- Lack of textbooks 30.3 percent
- Can't study at home; no materials 25.6 percent
- They don't pay attention 25.4 percent
- Too many learners in the class 24.3 percent
- They don't study 23.7 percent

Figure 23. Greatest Student Learning Obstacles as Reported by Teachers



Very few teachers viewed the following factors to be important obstacles to learning: 1) not enough time in the school day and 2) some learners being too young, and 4) learners cannot study at home because of lack of electricity.

SELECTED FINDINGS FROM CLASSROOM OBSERVATIONS

At each school, one enumerator was assigned responsibility for observing one Standard 2 classroom and one Standard 4 classroom, for approximately 90 minutes each, and preferably observing at least a reading class or language class, although this may not have always been possible. The enumerators tasked with these observations were those with expertise in teacher training and knowledge of good teaching practices. Observers were asked to indicate whether they observed: 1) the opposite of the behavior described (all behaviors were stated in a positive frame, such as "uses a lesson plan" or "manages instructional time effectively," 2) the behavior exhibited sometimes or partially correct, or 3) the behavior exhibited very well and consistently where appropriate.

Of classrooms surveyed, there were 304 observations of Standard 2 Chichewa reading classes and 298 Standard 4 classes. Out of behaviors observed (see classroom observations study in Annex 6, only one behavior was observed in more than 80 percent of Chichewa classes. This task was that the teacher "applies multiple methods to support comprehension, including games, group work, etc.," which was observed in 88.16 percent and 89.26 percent of classes in Standards 2 and 4, respectively. Enumerators observed that 71.71 percent of Standard 2 teachers observed and 70.13 percent of Standard 4 learners observed were provided with a variety of methods for learners to establish good writing skills.

Of classes observed, enumerators observed that teachers in 77.6 percent of Standard 2 classes and 76.8 percent of Standard 4 classes engage learners in reading activities or games appropriate to their reading level. However, they also found that 38.2 percent of Standard 2 teachers have individual learners read out loud. As more than 75 percent of Standard 2 learners could not read any words on the oral reading subtask, it seems unlikely that this activity would engage learners in reading activities appropriate to their grade. A much higher percentage of teachers in Standard 4 classes, or 75.17 percent, have learners read out loud. Based on test scores, this seems more viable but it also seems likely that many learners would have struggled with this as well.

Some tasks varied by grade. Teachers asked 41.78 percent of Standard 2 Chichewa classes and only 18.46 percent of Standard 4 learners to recognize letters and say letter names and/or sounds. This is somewhat proportional to what would be expected at the different standards, although test scores indicate that Standard 4 learners have not mastered letter name knowledge and still may need more work on this. In Standard 2 classes, 59.2 percent of teachers observed provide instructions on how to decode syllables and words, while 80.2 percent of teachers in Standard 4 do so. Again, few learners effectively met this benchmark.

There are also a number of recommended teaching behaviors that were not commonly observed. Teachers only asked 10.2 percent of Standard 2 classes and 5.37 percent of Standard 4 classes to recite the alphabet. While the alphabet may be too basic for Standard 4 learners, Standard 2 learners could benefit from reciting the alphabet. Teachers in only about 15 percent of both Standard 2 and Standard 4 classes "asked pre-reading questions such as *What do you think the story will be about based on the pictures and/or title of the book?*" Only teachers in 10.7 percent of Standard 2 classes and 17.45 percent of Standard 4 classes had learners retell a story that they or the teacher had read.

A number of other behaviors that are less specific to reading are also important to learner success. Enumerators observed teachers doing the following behaviors in either Chichewa reading or non-reading classes frequently (in 80 to100 percent of classes) and were consistently identified in the third category: "see the behavior exhibited very well and consistently where appropriate" were as follows (n = 1,368):

- Demonstrates effective classroom management skills
- Treats all learners equally/fairly
- Assesses pupil learning
- Avoids using gender-biased language
- Avoids using abusive language
- Provides positive, encouraging feedback
- Does not allow learners to use abusive language
- Most learners are paying attention
- Most are actively engaged in the lesson
- Learners appear to understand what the teacher is saying

SELECTED FINDINGS FROM SCHOOL CLIMATE CHECKLIST

While learners were being assessed at the schools, the data-collection supervisor went around the school grounds and checked on a number of factors that can influence the "climate" of the school. School climate has been shown to be a factor in how successful a school is in supporting learners' learning and social development. A copy of the School Climate Checklist used in this study is in Annex 6. All of the ratings on these factors are the judgments of the data-collection supervisors and are inevitably subject to differences in interpretation.

The factors that showed up most striking are as follows (n = 316 unless otherwise specified):

- More than one third of the 316 schools visited did not have well-maintained school grounds or still had litter on the ground.
- Almost half of the schools did not have functioning locks for the classrooms or other buildings on the school grounds.
- Of schools observed, 37 percent had broken windows.
- Almost one quarter (24.12 percent) of schools did not have classrooms with sufficient light, although 90 percent of classrooms were deemed to have sufficient ventilation.
- Eighteen schools observed did not have a latrine for learners.
- Almost all of the schools visited (88 percent) had separate latrines for girls, which is often considered important to increase female enrollment in school.
- Just more than two thirds (or 69.93 percent) of the classrooms did not have enough desks for all of the learners.

- Of schools observed, only 20.5 percent of schools had a library (*n* = 274).
- In 32 percent of the schools, enumerators observed bullying or fighting among the learners.
- Most textbooks were distributed to learners in only 23.8 percent of classes.

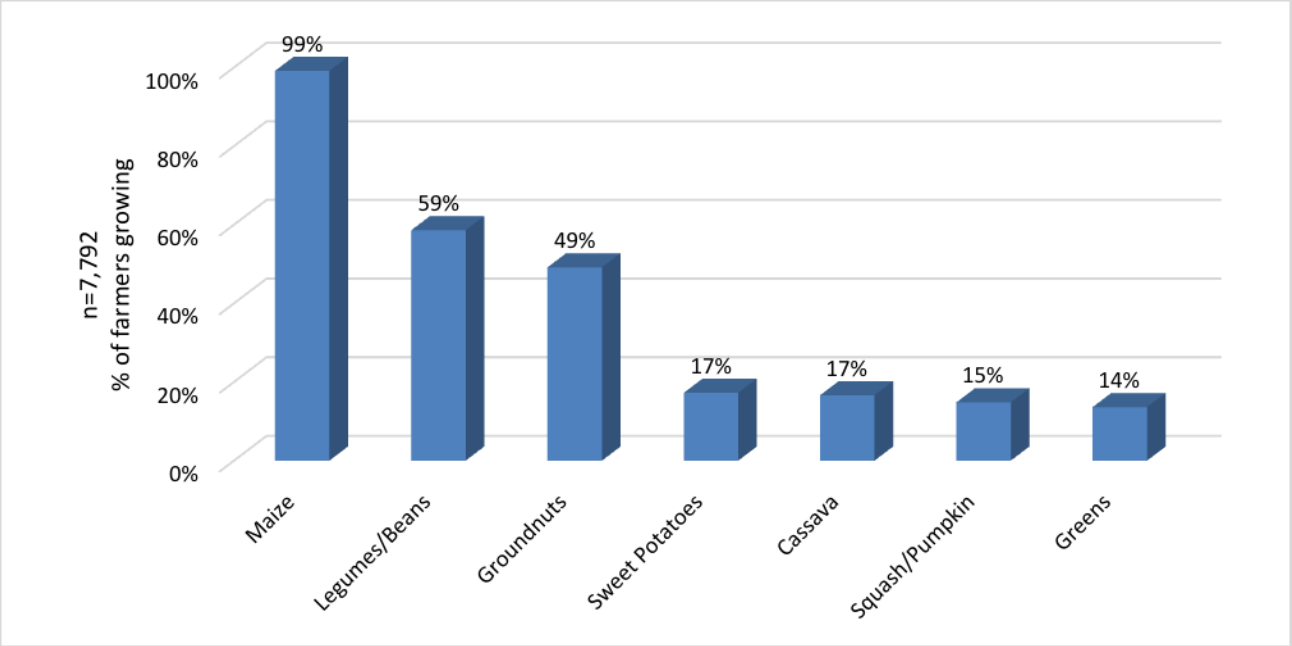
SELECTED FINDINGS FROM HOUSEHOLD SURVEY

Basic Household Characteristics. Two thirds (66.4 percent) of the households surveyed were headed by the father of the learner (*n* = 8,888), with another 15.3 percent headed by the mother and 10.5 percent by a grandparent of the learner. Most households (62.9 percent) included two parents, while 23.5 percent had only one parent, and 13.6 percent had no parents living in the household (*n* = 8,847). The average size of the households surveyed was almost seven people; the largest had 13. On average, each household had 3.9 children under the age of eighteen (*n* = 8,847). The average age of Standard 2 learners was just more than nine years old (*n* = 4,350), while Standard 4 learners averaged almost 12 years old (*n* = 4,464).

Access to Services and Development Support. When asked how they travel to visit a clinic, 72.3 percent of HHS respondents indicated that they walked, 23.1 percent traveled by bicycle, and 3.9 percent took some type of public transportation. Using those means of transportation, the average amount of time required to reach the closest clinic was 1.9 hours, 59.7 percent of respondents claimed they did not have access to finance or a loan if they were to need it (*n* = 8,796), and 10.3 percent of households reported having received support from a donor organization in the past three years, with 12.1 percent receiving assistance through INVC or FtF, 5.1 percent through SSDI, 10.5 percent through WALA, and 5.2 percent through Save the Children.

Farming. Of households surveyed, 87.9 percent own land, and an even greater proportion, 89.0 percent, farm (*n* = 8,847). Of those farming, the average amount of land farmed is 2.4 acres (*n* = 7,792). The most commonly grown crops are shown in Figure 24.

Figure 24. Most Commonly Grown Crops among Households Surveyed



School Dropouts. Of households surveyed, 32.2 percent reported at least one youth (under the age of 26) in the household who had dropped out of school (*n* = 8,847). The likelihood of either a male or female youth dropping out was approximately the same: 19.1 percent for females (*n* = 18,420 female youth) versus 20.3 percent for males (*n* = 18,773 male youth). According to the respondents, the most likely reason for dropping out was that the learner was not motivated or did not value education, cited by 14.2 percent of respondents. Other, less common responses included financial reasons (6.1 percent), pregnancy (2.7 percent), sickness (2.5 percent), and marriage (2.3 percent).

PRELIMINARY RESPONSES TO TASK 2 EVALUATION QUESTIONS

The Task 2 evaluation questions for this study were derived to help USAID, MoEST, and other stakeholders gain an understanding of what factors most affect learner reading outcomes, dropouts, and repetition in Malawi. This will help stakeholders determine how the EGRA Project and other development interventions might affect those outcomes. To answer these questions, reading outcomes were correlated with a number of the findings from the household, head teacher, teacher, and learner questionnaires as well as the school climate and classroom observation protocols. This was done using ordinary least squares (OLS) regression models, which reveal the isolated effects of various factors on reading scores while controlling for other factors.

The evaluations team used OLS regression to answer Task 2 Evaluation Questions 1, 3, and 4. Question 1 asks about factors that affect learner reading scores; Question 3 asks whether FtF

and GHI are affecting learner reading outcomes; and Question 4 asks about how factors similar to those considered for Evaluation Question 1 are affecting learner dropout and repetition rates. Since all of these questions seek to determine the effect of certain characteristics on others, regression analysis is the most appropriate statistical tool. Evaluation Question 2, however, is slightly different, as it simply asks about what community-level resources are dedicated to education. This question is easily answered through the use of summary statistics.

It is important to note that the answers to these evaluation questions are preliminary and, thus, not conclusive at this stage of the study. Baseline assessments are usually intended to provide a snapshot of *what is* and lack the benefit of tracking results over time. Moreover, they are less effective at explaining *why* conditions exist. Nonetheless, results are presented here in the hope that they can be used to provide some background knowledge and guidance to USAID, MoEST, and RTI in the implementation of the EGRA Project. All regressions will be rerun upon collection of midterm and endline data. At that point, evaluators will be able to look at the changes that have occurred within factors (or variables) over time to see if and how those *changes* affect outcomes. For instance, rather than just looking at a snapshot, evaluators will be able to look at whether (and if so, how) a 50 percent increase in the number of books available to learners has affected learner reading scores. If scores have not changed, it might suggest that this factor is not as significant as evaluators have estimated in this baseline report. Additionally, evaluators will be able to compare changes between a treatment (beneficiary) group and a control (non-beneficiary) group, which will allow evaluators to determine whether the changes identified (a 50 percent increase in books, in this example) are due to the EGRA Project.

Preliminary findings for each of the four evaluation questions are provided below.

Task 2 Evaluation Question 1: What household and community factors relate to learner reading outcomes?

To explore what factors relate to reading outcomes, evaluators specified several different regression models that looked at the association of learner scores on the oral reading fluency subtask of the RA with a variety of learner, household, school, teacher, and community factors (or independent variables). The team selected the independent variables based on the research summarized in the literature review presented earlier and created subgroups according to the following characteristics:

- Household Resources
- Household Support/Involvement

- Household Education Levels
- Learner Health and Nutrition
- Learner Attitude toward School
- School Resources
- Classroom Resources
- Teacher Experience, Training, and Use of Best Practices in Teaching
- Community Involvement in the School
- School Support from Outside Organizations

These subgroup regressions help to explain the extent to which each of the groupings is related to oral reading fluency scores. For instance, from these regressions, evaluators found that learner attitude toward school seems to explain more of the variation in reading outcomes than any of the other major factor subgroups. This is followed by household support and involvement in school and school resources, respectively. All of the subgroup regressions are included in Annex 9 of this report.[60]

In order to look at factors from each subgroup together, the evaluation team also specified regressions using the most important factors from each subgroup. An observation must be dropped if any variable in the regression is missing; as such, the more variables included in a regression, the more likely it is that observations will need to be dropped due to missing data. Accordingly, the evaluation team conducted three regressions to test if results were robust when including additional variables, which decreased the sample size. However, the team found consistency in significance and coefficients of variables and that the composition of the sample along key characteristics (sex and district) does not change across the regression specifications, providing evidence that the results are robust. As such, results are presented here only for the larger regression. Nonetheless, the two smaller regressions (with more records) are included in Annex 9 to show this process and the robustness of the findings.

The evaluation team conducted regressions separately for Standards 2 and 4 and found some variation in the results. This variation might be explained by the possibility that some factors affect learners of different ages or different levels of reading fluency differently. For instance, the highest level of education of any household member may matter more for older learners who start to surpass other household members in years of schooling, and teacher disposition

[60] Variables that did not show high degrees of correlation with the outcome were not included in the subgroup regressions.

may affect younger, potentially more tenuous learners more. See Tables 15 and 16 below for regression results by standard, followed by an examination of these results.

Table 15. Summary of Factors Affecting Standard 2 Oral Reading Fluency Scores

Independent Variable	Coefficient	Standard Error	*t*-Statistic	*p*-Value	[95% Confidence Interval]	
Access to reading materials at Home	2.152385***	0.4670141	4.61	0	1.236499	3.068272
Household member reads to learner	1.133336**	0.4489708	2.52	0.012	0.2528355	2.013837
Household member helps learner with homework	0.8108339*	0.448923	1.81	0.071	-0.0695729	1.691241
Household member volunteers at least one hour per month in learner's school	-0.1814292	0.3889693	-0.47	0.641	-0.9442576	0.5813993
At least one household member has graduated from Senior Level 2	1.459181***	0.5486679	2.66	0.008	0.3831588	2.535202
Number of days the learner did not eat anything this week	-0.336623	0.406999	-0.83	0.408	-1.13481	0.4615644
Learner gets tired at school	-0.9795245**	0.3861015	-2.54	0.011	-1.736729	-0.2223203
Learner likes to read	3.606555***	0.4324824	8.34	0	2.75839	4.454719
School resources (PCA score)	0.1392299	0.148107	0.94	0.347	-0.1512305	0.4296903
Separate latrines for girls	2.102013***	0.6385635	3.29	0.001	0.8496925	3.354334
Learner-to-teacher ratio	-0.0084686*	0.0046687	-1.81	0.07	-0.0176247	0.0006874

Teachers report having sufficient teaching resources	0.4289117	0.4042849	1.06	0.289	-0.3639529	1.221776
Number of reading materials on classroom walls	0.0393458	0.0245035	1.61	0.108	-0.0087093	0.0874008
Years of teaching experience	0.0570094**	0.0242432	2.35	0.019	0.0094647	0.1045541
Teacher use of best practices in teaching reading (PCA score)	0.4290836***	0.1014053	4.23	0	0.2302123	0.6279549
PTA meets at least monthly	1.034341***	0.4012062	2.58	0.01	0.2475138	1.821168
Number of times teacher meets with groups of parents yearly	-0.3841561***	0.1409916	-2.72	0.006	-0.6606623	-0.10765
Teacher invites HH involvement	1.166285**	0.5015333	2.33	0.02	0.1827013	2.149869
MTPDS support	0.8407176	0.5913775	1.42	0.155	-0.3190643	2.000499
Sex (0 = male, 1 = female)	-0.6076308	0.382942	-1.59	0.113	-1.358639	0.1433772
Constant	-2.750443	1.08403	-2.54	0.011	-4.876393	-0.6244934

Observations	2,017
***R*-squared**	0.1249
Adjusted *R*-squared	0.1161

Table 16. Summary of Factors Affecting Standard 4 Oral Reading Fluency Scores

Independent Variable	Coefficient	Standard Error	*t*-Statistic	*p*-Value	[95% Confidence Interval]	
Access to reading materials at home	1.846295**	0.7540464	2.45	0.014	0.3675683	3.325022
Household member reads to learner	3.885589***	0.8510401	4.57	0	2.216652	5.554526
Household member helps learner with homework	-2.427167***	0.8425437	-2.88	0.004	-4.079442	-0.7748927
Household member volunteers at least one hour per month in learner's school	1.605428**	0.7419193	2.16	0.031	0.1504828	3.060372
At least one household member has graduated from Senior Level 2	3.203274***	1.008979	3.17	0.002	1.224611	5.181937
Number of days the learner did not eat anything this week	-1.345949*	0.7879031	-1.71	0.088	-2.89107	0.1991729
Learner gets tired at school	-0.3285472	0.7905496	-0.42	0.678	-1.878859	1.221764
Learner likes to read	17.42218***	1.344002	12.96	0	14.78651	20.05784
School resources (PCA score)	0.8930867***	0.288751	3.09	0.002	0.32683	1.459343
Separate latrines for girls	-1.326891	1.268623	-1.05	0.296	-3.814731	1.160948
Learner-to-teacher ratio	-0.0269137**	0.0132527	-2.03	0.042	-0.0529029	-0.0009245

Teachers report having sufficient teaching resources	2.916964***	0.7690359	3.79	0	1.408842	4.425086
Number of reading materials on classroom walls	0.0795572*	0.0462373	1.72	0.085	-0.0111167	0.1702311
Years of teaching experience	-0.1217528***	0.0419449	-2.9	0.004	-0.2040091	-0.0394966
Teacher use of best practices in teaching reading (PCA score)	0.4092773**	0.1757957	2.33	0.02	0.0645323	0.7540224
PTA meets at least monthly	1.993275**	0.786215	2.54	0.011	0.4514637	3.535086
Number of times teacher meets with groups of parents every year	0.2846675	0.2421959	1.18	0.24	-0.1902921	0.7596271
Teacher invites HH involvement	-1.243736	0.8185236	-1.52	0.129	-2.848906	0.3614339
MTPDS support	4.227051***	1.087432	3.89	0	2.094536	6.359565
Sex (0 = male, 1 = female)	2.098303***	0.7259465	2.89	0.004	0.6746821	3.521925
Constant	5.266103	2.291099	2.3	0.022	0.7731325	9.759074

Observations	2,196
***R*-squared**	0.1405
Adjusted *R*-squared	0.1325

Household Resources. Household resources, including household wealth or socioeconomic status, amount of money spent on the learner's education in the past year, and availability of reading materials in the household help to explain between 1 and 2 percent of the variation in oral reading fluency scores in the assessment areas.[61,62] The factor that appears to have the greatest correlation with oral reading fluency scores is access to reading materials (including books, magazines, newspapers, etc.) at home, as reported by household members. As shown in Tables 15 and 16, Standard 2 learners with access to these materials at home scored an average of 2.2 cwpm higher on their reading fluency test than did Standard 2 learners who did not have access to such materials (p-value < 0.01), and Standard 4 learners with access to reading materials at home scored 1.8 cwpm higher than their peers (p-value < 0.01). Approximately 30 percent of learners assessed have access to reading materials in their homes, and this proportion is relatively equal for learners of different sexes, though Standard 4 learners were somewhat more likely to have access to reading materials at home than Standard 2 learners (37 percent of Standard 4 learners had access to reading materials at home while only 23 percent of Standard 2 learners did). Also, while there was a positive correlation between household wealth (socioeconomic status as measured by household assets) and oral reading fluency, this result was not statistically significant.

Home Support/Involvement. Home support and family involvement in school helps to explain about 2 percent of the variation in oral reading fluency scores for Standard 4 learners and almost 4 percent of the variation in reading fluency scores for Standard 2 learners in the reading assessment treatment and control areas.[63] As found in many other studies, including the PIRLS study described in this evaluation's literature review, learners who are read to at home were more likely to score higher on the RA reading fluency subtask than learners who are not. In fact, as shown in Tables 15 and 16, Standard 2 learners who report being read to at home were able to read an average of 1.1 cwpm faster than their peers who are not read to (p-value < 0.05), and for Standard 4 learners, that number was 3.9 cwpm (p-value < 0.01). Approximately 43 percent of sampled Standard 2 learners and 63 percent of sampled Standard 4 learners report being read to at home at least sometimes, and boys and girls were equally as likely to be read to at home.

---

[61] Household resources were determined through household surveys. Socioeconomic status is derived from conducting PCA with reported household assets, and school expenditures were derived by adding up all of the expenses households reported incurring for the learner to attend school in the past academic year. All of the variables included in household resources are shown in Annex 6.

[62] This is evidenced by the adjusted R-squared figure in both figures.

[63] Home support and family involvement in school was determined through household surveys (number of hours someone from the household volunteers in the learner's school and for what period of time the learner attended preschool) and learner questionnaires (whether anyone at home helps the learner with his or her homework or reads to him or her).

Moreover, whether or not a household member helps the learner with his or her homework appears to be positively correlated with reading scores for Standard 2 learners at a statistically significant level (*p*-value < 0.1). However, help with homework is negatively correlated with higher oral reading scores for Standard 4 learners, which could be because Standard 4 learners may only need help with their homework if they are struggling more in school, whereas Standard 2 learners are newer to school and, thus, *all* may require more help. Standard 4 learners also appear to be affected by whether or not their household members are active in their school (*p*-value < 0.05). More than 44 percent of sampled Standard 2 learners report receiving help on their homework from someone in their household (which is less than the percentage who receive help in Standard 4). Also, girls are statistically slightly more likely to receive help on their homework than boys (at 56 and 53 percent, respectively). In addition, 56 percent of Standard 4 households report that someone from that household volunteers at least once per month at the learner's school. The percentages for volunteer support are similar across standards and by sex.

Household Education. While, overall, household education levels help to explain less than 1 percent of the variation in oral reading fluency scores for both Standard 2 learners and Standard 4 learners in the reading assessment treatment and control areas, the correlation between household education levels and learner reading scores is quite high.[64] For each additional level of education completed by someone in a learner's household, the learner's reading fluency scores increased. Those learners who live in a household where at least one member of the household has graduated from Senior Level 2 (secondary school) scored an average of 1.5 to 3.2 cwpm higher on their reading fluency subtask than did learners who lived in households where the highest level of education was less than secondary (*p*-value < 0.01). Approximately 14 percent of learners fit this description, with the percentages equal across standards and sexes.

There could be several reasons for the strong correlation between household education and learner reading ability, as discussed in the literature review section of this report. Two main reasons, though, are that: 1) Better-educated households are likely better able to provide homework support to learners, and 2) Better-educated households are more likely to model the importance of a good education and, therefore, likely have higher educational expectations of learners than do other households. Baseline findings from EGRA treatment

[64] Household education levels were determined through gathering information about each of the members of a learner's household through completion of a household roster during a survey conducted with the head of the household. Evaluators then took the highest level of education completed for each of the household members and determined which one was the highest. From there, the team created variables for households where at least one member had completed certain levels of school. These levels are commutative and non-exclusive. Thus, they include everyone who completed the designated level, even if that person also completed a higher level of education.

and control areas support both of these hypotheses. Evaluators find that households where at least one member has graduated from Senior Level 2 (secondary school) are statistically more likely to read to learners, help learners with their homework, and demonstrate characteristics indicating that they may place a higher value on education. With regard to the latter, evaluators found that higher-educated households were more likely to both send learners to preschool for at least two years and to volunteer at least one hour per month in a learner's school.

Learner Health and Nutrition. Learner health and nutrition does not appear to explain much of the variation in oral reading fluency scores for either Standard 2 learners or Standard 4 learners (at 1/3 percent or less).[65] The one factor that appears to consistently be correlated with Standard 2 oral reading fluency scores among the sample population is whether the learner reports feeling tired at school. Those Standard 2 learners who report feeling tired at school scored an average of 1 cwpm lower on their reading fluency test than do those who do not, and 46 percent of Standard 2 learners reported sometimes feeling tired at school (that number was 35 percent for Standard 4 learners, and sleepiness did not vary by sex).

While a few of the health and nutrition variables appear to consistently show a relationship with oral reading fluency scores for Standard 4 learners, none of those variables is consistently statistically significant. This does not mean that these variables do not matter but only that the evaluation team is less confident in their relationship with learner reading scores. The evaluation team did find that in some specifications of the regression model, the number of days in the past week a learner has not eaten, whether the child's family takes him or her to the doctor when he or she is sick (those who do go to the doctor score higher), and the number of weeks the child was sick in the past month (the more sick weeks, the lower the scores) appeared to be matter. For instance, in the regression model above, for each day that a Standard 4 learner did not eat in the past week, his or her score decreased by an average of 1.3 cwpm. However, when other variables were added into the regressions or taken out, these factors became less significant. This could be due to the strong correlation between household education and health, as described in the literature review section of this report and/or the strong relationship between nutrition and health. Follow-up studies for this evaluation will look more closely at this relationship.

Many of the factors that were examined in this subgroup are often used as indicators of a child's level of health and nutrition. While most of these factors do not have a statistically significant relationship with learner reading scores at this stage of the evaluation, that does

[65] Nutrition and health information was largely gathered through the household survey using food security and nutrition modules from the FtF Population Survey instrument as well as questions related to learner health. The one exception to this rule is the variable about the number of days a learner did not eat in the week prior to the survey, which is derived from self-reported data from learners gathered during the RA in schools.

not mean they are not important. Rather, it means that the evaluation team is unable to determine with confidence whether these factors are correlated with reading outcomes. As such, changes in health and nutrition due to the GHI and FtF interventions, which focus on improving access to and quality of healthcare, and agricultural production and nutrition among beneficiary households, respectively, may help the evaluation team to better determine relationships among these factors. The evaluation team will track these indicators as the GHI and FtF interventions are rolled out to determine what relationships exist between health, nutrition, and reading ability.

Learner Attitude toward School. Learner attitude about school appears to explain more of the variation in oral reading fluency scores than any of the other subgroups. For Standard 2 learners, it explains almost 6 percent of the variation in oral reading fluency scores, and for Standard 4 learners it appears to explain more than 10 percent of the variation in scores. The main reason for this is the strong correlation between oral reading scores and whether or not a learner likes to read. As described in the literature review, children usually like to do things they are good at. So, while there is a strong relationship here, causality is particularly difficult to determine—because it is not clear which comes first—liking to read or being good at it. Regardless, Standard 2 reading fluency scores were an average of 3.6 cwpm higher for those learners who like to read than they were for those who don't (p-value < 0.01). And, for Standard 4, that number was 17.4 cwpm higher, which is almost one full standard deviation (18.1 cwpm) from the average oral reading fluency score (15.2 cwpm) at that standard level (p-value < 0.01).

Approximately 68 percent of Standard 2 learners and 92 percent of Standard 4 from the evaluation sample reported that they like to read, which also supports the hypothesis that learners like reading when they are good at it (since more Standard 4 learners are better at reading than Standard 2 learners).

School Resources. While, overall, school resource levels help to explain less than 1 percent of the variation in oral reading fluency scores for both Standard 2 learners and Standard 4 learners in the reading assessment treatment and control areas, several statistically significant relationships exist.[66] As shown in Tables 15 and 16, both Standard 2 and Standard 4 learner reading fluency levels appear to be negatively correlated with the learner-to-teacher ratio (p-

[66] School resource levels were determined through enumerator observation of the state of school infrastructure through the use of the school climate protocol, included in Annex 4 as well as through HT-reported resources such as electricity, clean water, and the learner-to-teacher ratio. All of the variables used are shown in the regression table for school resources in Annex 6. Following that regression, the evaluation team created a principal component for school resources that included all of the items from that regression with the exception of girls' latrines and the learner-to-teacher ratio, which were kept separate due to their oft-reported significance in learner success in school.

value < 0.1 and *p*-value < 0.05, respectively). Additionally, Standard 2 learner scores seem to be highly correlated with whether or not learners have access to a separate girls' latrine at the school (*p*-value <0.01), with learners scoring an average of 2.1 cwpm higher on their oral reading fluency test if they do have access to a girls' latrine. On the other hand, Standard 4 learner scores appear to be correlated with overall school resources as observed and elaborated by enumerators and HTs (*p*-value <0.01).

The average learner-to-teacher ratio at sampled schools was 94 to 1 for Standard 2 and 65 to 1 for Standard 4, and 80 percent of schools have a separate girls' latrine. Access to a separate girls' latrine varies significantly by district, however, with only 21 percent of learners sampled in Balaka having access to a girls' latrine. Zomba Rural had the next lowest percentage of learners without access to a girls' latrine, at 64 percent. At least 75 percent of the learners from all of the other sampled districts had access to girls' latrine at school.

Classroom Resources. Classroom resources only appeared to explain between 0.75 and 1 percent of the variation in oral reading fluency scores among sampled learners. The main variables that seemed to matter here were whether or not the teacher reported having sufficient resources and the number of reading materials posted on classroom walls, which is used as a proxy for indicating overall classroom resources. Both of these variables appear to help predict scores for Standard 4 learners but not necessarily for Standard 2 learners. The evaluation team found that approximately 38 percent of learners assessed came from classes where the teachers reported having access to sufficient resources to teach, with Standard 2 teachers slightly more likely to feel satisfied than Standard 4 teachers. Also, on average, classrooms had approximately six different reading items on their walls, which did not vary by standard.

Teacher Experience, Training, and Use of Best Practices in Teaching. Interestingly, teacher experience, training, and use of best practices in teaching reading appears to explain quite a bit more of the variation in learner reading scores for Standard 2 learners than it does for Standard 4 learners (at almost 3 percent and less than 0.5 percent, respectively). This is likely because Standard 4 teachers from the evaluation sample are statistically more likely to use best practices in teaching reading than are Standard 2 teachers, and, likewise, Standard 4 teachers are more likely to receive more hours of coaching and to have said that they believe they have received enough coaching to practice the methods they have been taught.

Standard 4 learners in classrooms where teachers report believing they have sufficient access to resources scored an average of 2.9 cwpm higher on their oral reading fluency test than did learners from classrooms where the teachers did not feel this way (*p*-value < 0.01). Teacher use of best practices in teaching, as defined through teacher adherence to the practices defined in this evaluation's classroom observation protocol (see Annex 6 for the protocol), showed a statistically significant positive correlation with learner reading fluency for both

Standard 2 and Standard 4 learners; however, it is difficult to determine to what extent test scores were affected by teacher use of best practices. The reason for this is that it is difficult to interpret the coefficient for best practices since this variable was created through PCA, and the PCA score by itself does not have much meaning. It is simply a number that allows evaluators to rank or sort records, and, thus, only holds meaning when compared with other PCA scores. Nonetheless, teacher use of best practices is positive and robust across regression models; the evaluation team thus feels confident that it is important.

Community Involvement in the School. Since a great deal of the focus of the EGRA, FtF, and GHI Projects is on community involvement, which also plays a significant role in sustainability of these projects, the evaluation team also looked at community involvement in schools to determine how much of an effect it had on learner reading ability. The team found that community involvement in school, as measured by frequency of Parent Teacher Association (PTA) meetings, teacher meetings with groups of parents outside of PTA, whether or not the teacher invites parents to be involved in school, and other community support to the school, only appears to explain between 0.25 and 0.75 percent of the variation in learner reading scores. Despite this fact, evaluators found that learners from schools where the PTA meets at least monthly performed between 1 and 2 cwpm higher on their reading fluency test than did learners from schools where either the PTA meets less frequently or no PTA exists (p-value < 0.01 for Standard 2 and p-value < 0.05 for Standard 4). Yet, only 35 percent of sampled learners attend a school where the PTA meets at least once per month.

Standard 2 learners were also more likely to score higher (1.2 cwpm on average) on their reading fluency subtask if their teacher reported having invited their parents to be involved in school (p-value < 0.01). Approximately 72 percent of Standard 2 learners and 63 percent of Standard 4 learners' teachers reported inviting parents to be involved in the school. And, although 86 percent of HTs interviewed reported that their school received other community support outside of the PTA and parental involvement, the evaluation team did not find such support to be statistically significantly correlated with learner reading fluency across regression models.

School Support from Outside Organizations. Although it appears to explain less than 1 percent of the variation in reading fluency scores, school support from outside organizations does seem to affect learner reading ability. Standard 4 learners from MTPDS-beneficiary schools scored 4.2 cwpm higher on their oral reading fluency test than did learners from schools not supported by MTPDS. Approximately 86 percent of sampled learners attend schools where MTPDS provided some sort of support.

District Variation. In addition to looking at various household, classroom, school, and community variables and their effects on learner reading outcomes, the evaluation team also specified a regression model that included districts. All districts were compared to Lilongwe

Rural, which was chosen as the comparison district because average reading scores were closest to the overall average and there was a high number of observations in Lilongwe Rural in comparison to other districts with similar average scores. There was only one district where learner reading scores were statistically higher than those for learners from Lilongwe Rural for both Standards 2 and 4 (by 2.6 and 4.7 cwpm, respectively), and that district was Salima. Standard 2 learners from Blantyre Rural also scored statistically higher than Lilongwe Standard 2 learners (by 1.7 cwpm), and Standard 4 learners from Machinga, Ntchisi, and Zomba Rural scored statistically higher than Lilongwe Rural Standard 4 learners (by 3.4, 5.3, and 6.8 cwpm, respectively). Finally, Standard 4 Mzimba North learners scored significantly lower than Lilongwe Rural learners and the average across all Standard 4 learners (by approximately 6.6 cwpm).

Task 2 Evaluation Question #2: How have health and agricultural interventions at the household and community level affected schooling and reading outcomes?

Determining the impact of the GHI and FtF interventions on schooling and reading outcomes is one of the central goals of this project. As this report has data from only one point in time (a snapshot of the situation in 2013) it is impossible to answer the question of how these two interventions have affected the outcomes of interest. Truly answering this question requires the difference-in-differences comparisons described above in the methodology section, which will only begin to be possible after the collection of midline data in 2015.

At this stage, after baseline data has been collected, it is possible to look for correlations between early exposure to these interventions (knowing that neither of these interventions has been in place for more than a year and are, thus, not likely to have acheived results yet) at the household level and schooling and reading outcomes. In other words, the analysis of the current data can show whether or not those who have received either of these interventions are more or less likely to have youth in their households who have dropped out of school. Examining the baseline RA data can also show whether there is a correlation between a household's participation in the GHI or FtF interventions and the reading scores of that household's learner. It is important to note, however, that even if a correlation currently exists, such a correlation would not imply a causal relationship between the two. For example, if households who have received support from FtF have learners with lower average reading fluency scores, this finding could not be interpreted to say that FtF causes learners' scores to be lower. As an alternative explanation, it is possible that FtF is active in areas with greater food insecurity, and it is this food insecurity that is leading to lower reading scores, not the FtF program.

The HHS included one question that asked the respondent whether the household had received support from the INVC/FtF Program, and another question that asked whether the household had received support from the SSDI/GHI program. The variables representing the

answers to these two questions were included in all of the regression models constructed for all of the Task 2 Evaluation Questions. There were no statistically significant correlations between GHI and any of the outcomes of interest (dropout rates, repetition, reading fluency), and similarly, no statistically significant correlations between support to the household from FtF and any of these outcomes. This finding was true across all of the regression models described in the preceding and following sections, including for Standard 2 and Standard 4 learners, for male and female youth, and for repetition at all standards.

Again, these findings cannot be interpreted as an evaluation of the effectiveness of either intervention, nor even of their potential impact on schooling and reading outcomes. They simply state that, in the snapshot captured by this evaluation's baseline report, no statistically significant relationships were observed.

Task 2 Evaluation Question #3: What levels of household and community resources/factors are dedicated to schooling and reading?

Individual skill and motivation are important factors for academic success, and teacher and school quality are crucial, but the amount of support from the family and the broader community are also integral to a learner's success. This section summarizes the findings related to the level of household and community investment in schooling, and in reading in particular. The data presented here draws primarily from the HHS, combined with the responses to several questions on the teacher and head teacher questionnaires.

At the community level, HHS respondents were asked whether their learner's school had a PTA or school committee. Almost all respondents (99.4 percent) to this question (n = 8,119) indicated that their school had a PTA (5.5 percent), a school committee (60.1 percent), or both (33.8 percent), and 98.7 percent of head teachers (n = 314) surveyed indicated that their school did have a PTA, implying that many members of the learners' households may not be aware of the presence of the PTA or understand what exactly it is. According to the head teachers, most PTAs met between once a month and once every two to three months, as shown in Figure 25.

Figure 25. Frequency of PTA Meetings for Each School

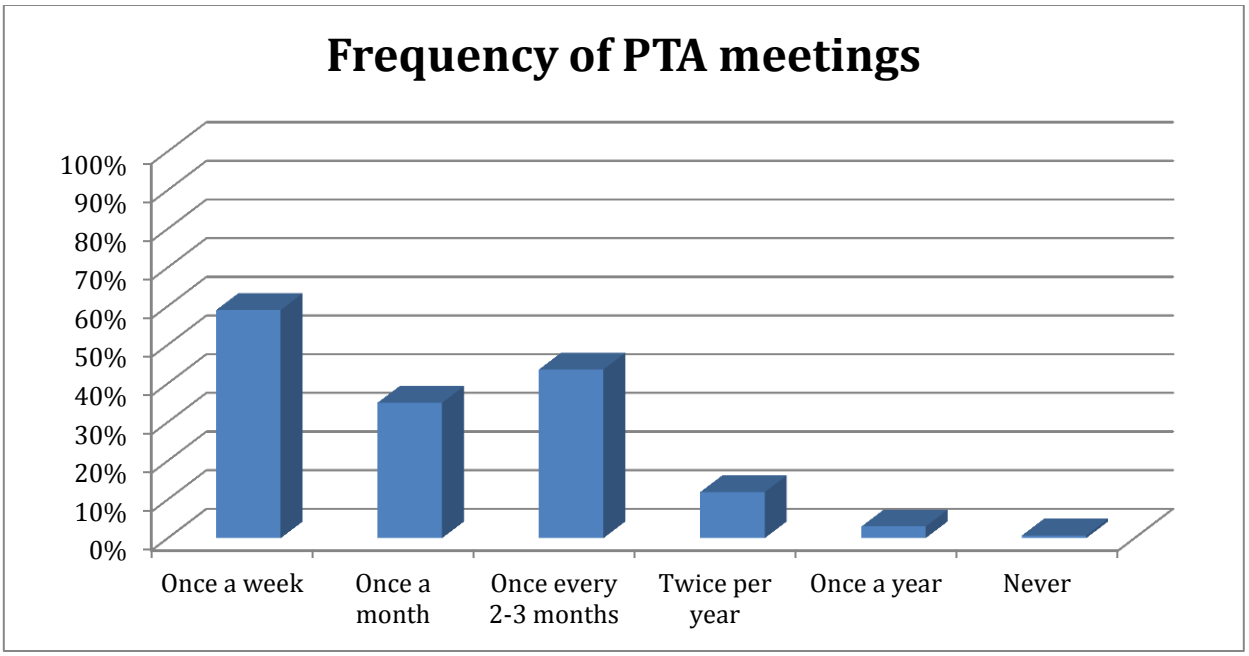
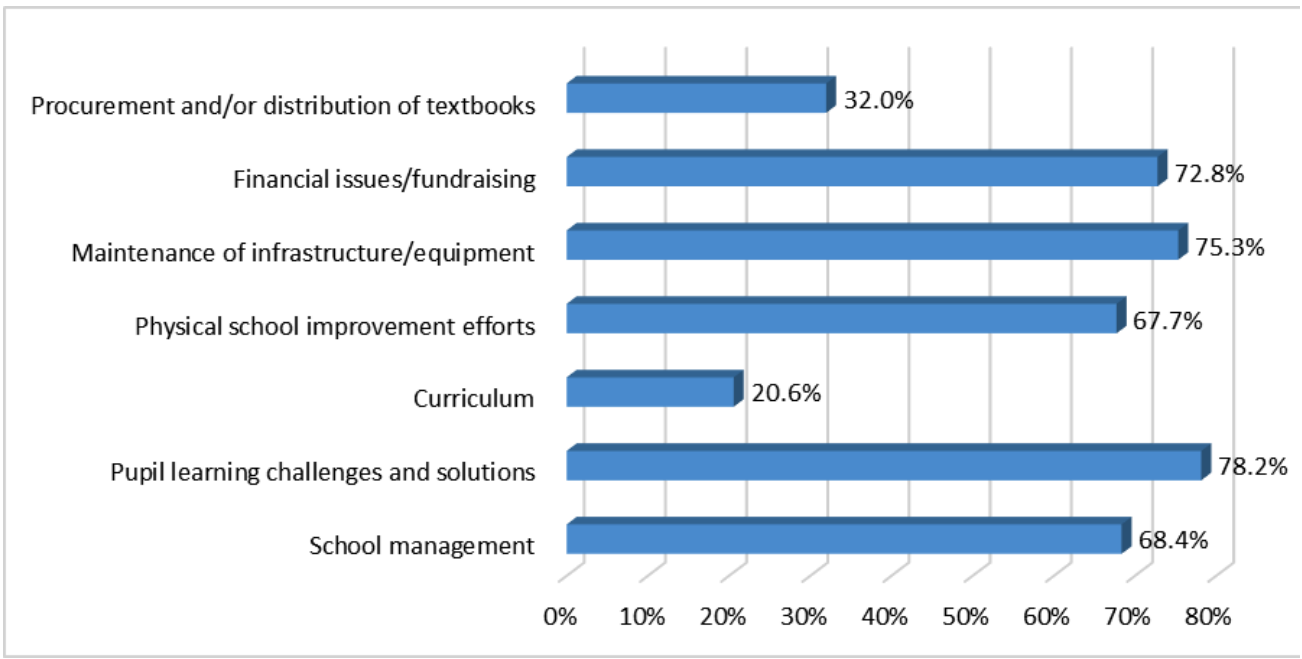


Figure 26 shows the various responsibilities that HTs reported their PTAs having *vis-à-vis* the schools, including contributing to fundraising (72.8 percent), maintaining the physical infrastructure of the school (75.3 percent), and strategizing solutions to pupil learning challenges (78.2 percent). Approximately one third (32.0 percent) of PTAs were said to be responsible for the procurement or distribution of textbooks, and only 20.6 percent held some authority over the curriculum. Approximately three quarters (78.7 percent) of head teachers (n = 301) reported inviting parents to participate in their learner's classroom or engage in extracurricular activities, and 84.9 percent of HTs (n = 312) said that the community, even besides parents and the PTA, was also involved in supporting the school and learner learning. However, 73.4 percent of HTs (n = 304) felt that community involvement had decreased over the past three years.

Figure 26. Responsibilities of PTAs



At the household level, 93 percent of respondents (*n* = 8,252) felt welcome in the learner's school, and 75 percent (*n* = 8,247) said that they or other members of their household were currently involved in the learner's school. As shown in Figure 27, of those that indicated that they or another household member were involved in the learner's school, the most common estimate for the amount of time spent each month participating in school activities was two hours (24.3 percent), and the majority spent between two and four hours a month involved with school activities.

Figure 27. Amount of Monthly Time Spent Participating in School Activities

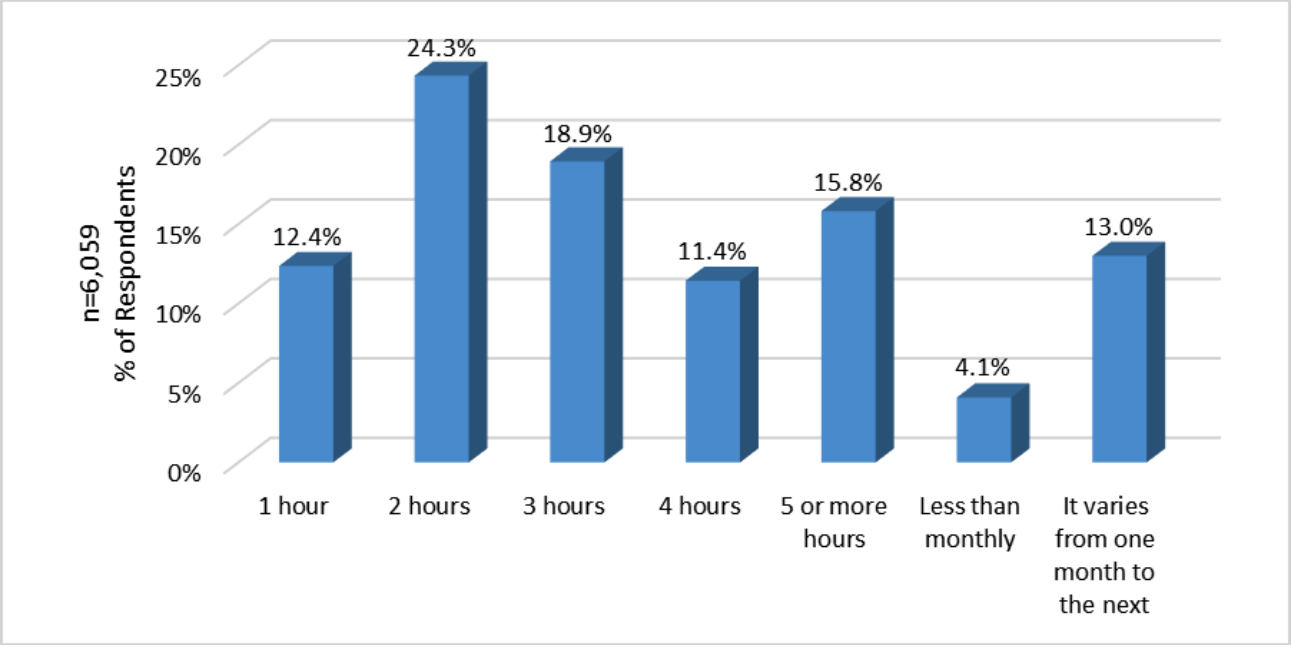
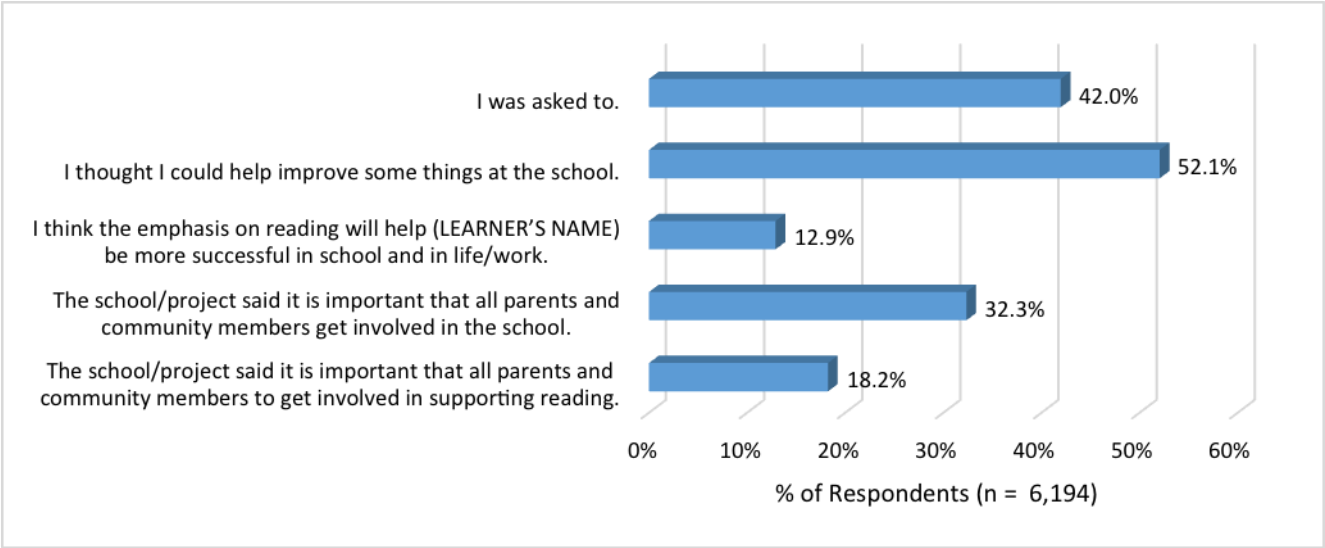


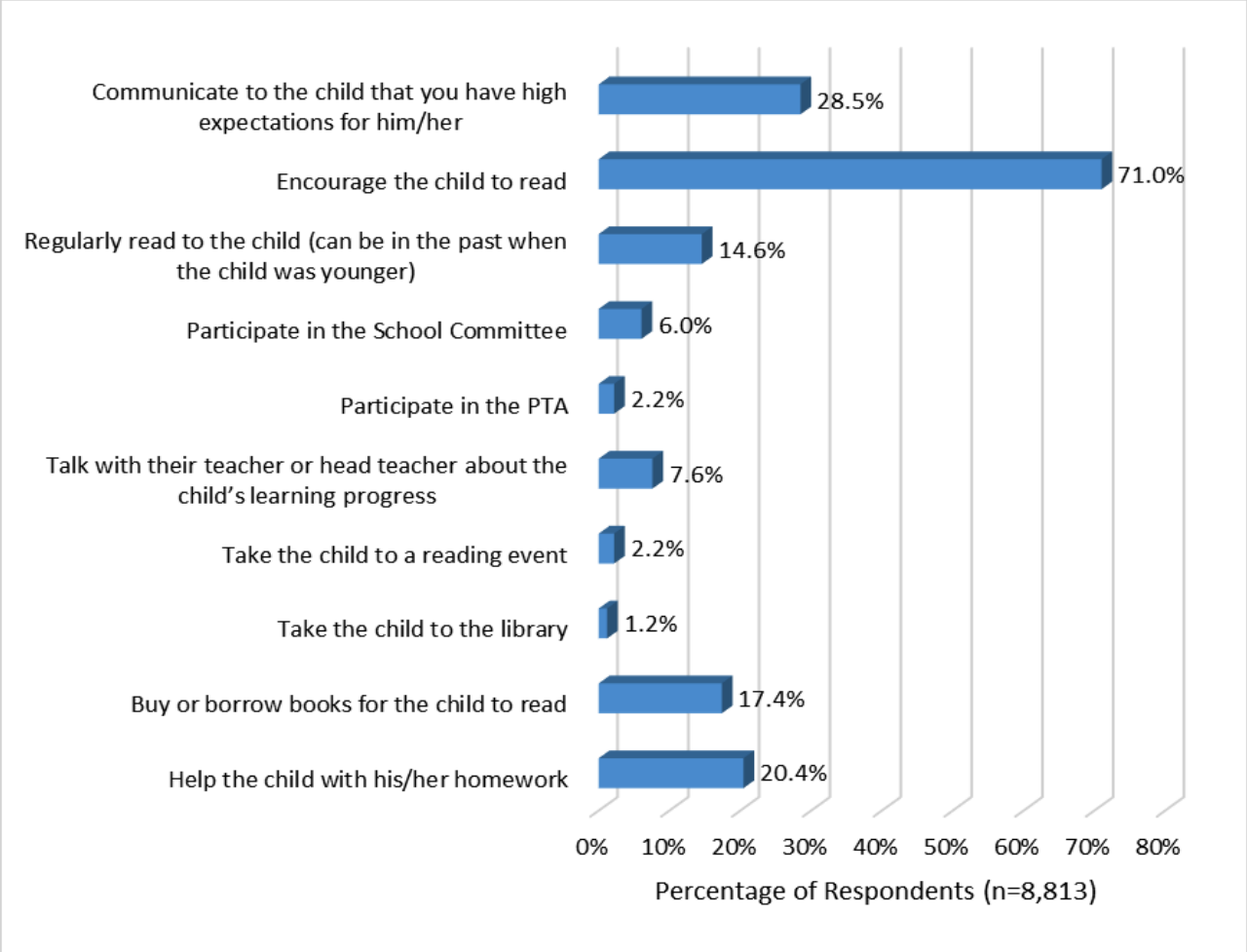
Figure 28. Motivation for Household Member Participation in School Activities



When asked about the motivation for their participation, more than half (52.1 percent) expressed the idea that they would be able to help improve things at the school, while others cited requests or encouragement from the school or project, and only 12.9 percent said that they thought the increased emphasis on reading would help their learner to be more successful. (See Figure 28.)

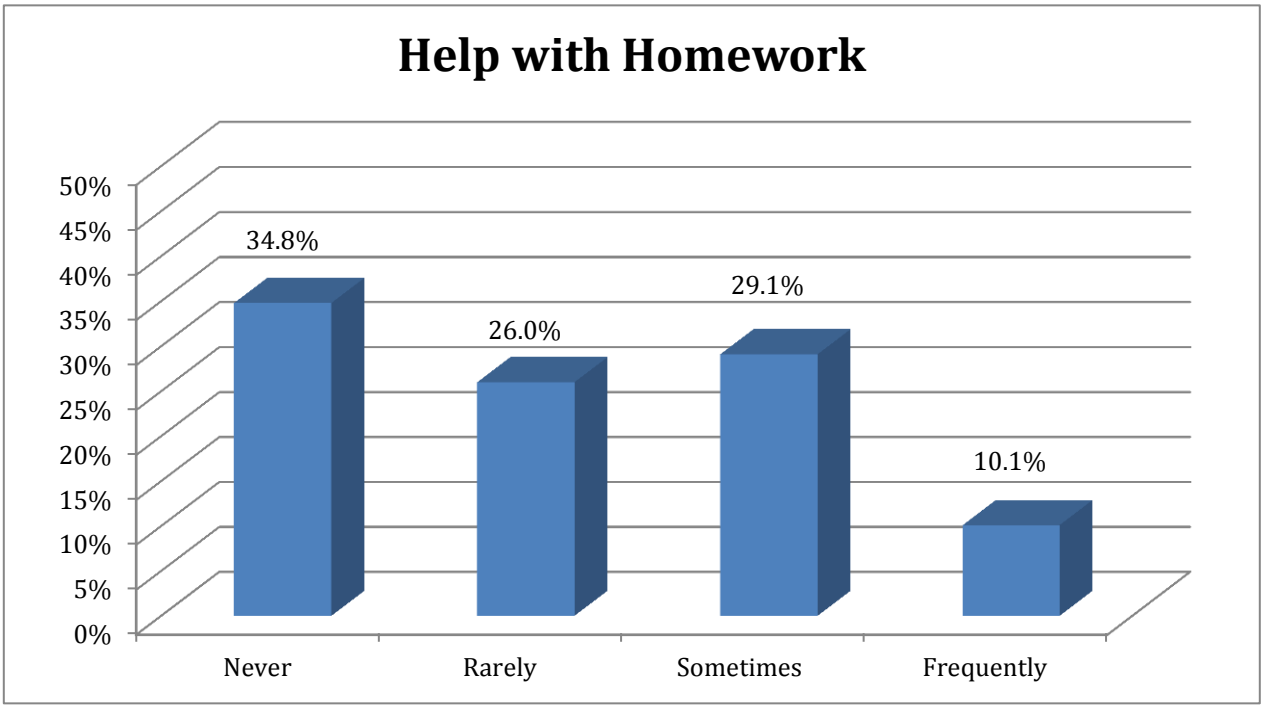
Respondents (*n* = 8,813) were asked in what ways they helped the learner to learn, and the most common response by far was that they encouraged her or him to read (71.0 percent), followed by communicating high expectations (28.5 percent) and helping her or him with homework (20.4 percent). Only 17.4 percent of households bought or borrowed books for the learner to read, and only 14.6 percent reported reading to the child regularly. (See Figure 29.)

Figure 29. Ways to Help a Learner Learn



Another question on the household survey asked whether the learner ever worked on homework outside of school. According to HHS respondents, 42.5 percent of Standard 2 learners and 80.2 percent of Standard 4 learners did homework outside of school. Of those whose learners did do homework outside of school, the majority, 65.1 percent, said that someone from the household has helped the learner with his or her homework, with 10.1 percent saying they did so frequently, 29.1 percent sometimes, and 26.1 percent saying they did so only rarely. (See Figure 30.)

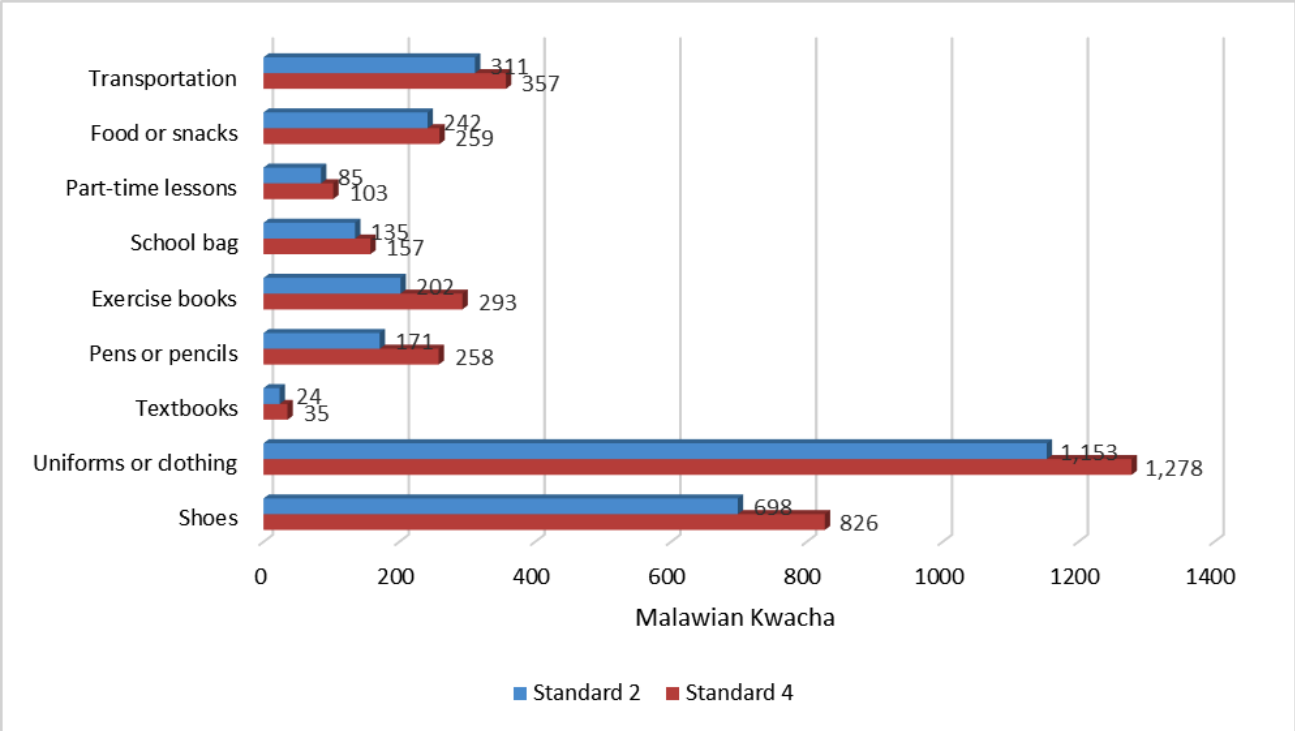
Figure 30. Learner-Reported Frequency of Help with Homework from Someone in their Household



For the figure above: n=4,967

In addition to investments of time and encouragement, families also invested money in their learners' education. The largest expenditures were for clothes and shoes, with food, transportation and supplies also identified as significant school expenses. (See Figure 31.) On average, households spent more on Standard 4 learners than on Standard 2 learners for every type of expense.

Figure 31. Average Yearly Expenditure on Learner's Education



Task 2 Evaluation Question #4: What factors at the household and community level have been identified that relate to repetition and dropout, and are there sex differences at the household level?

Because different factors might influence dropping out of school and repeating a standard in different ways, the evaluation team developed separate models for repetition and dropouts. Sex is integrated into both models.

Dropout Model. Since this is the baseline study, the evaluation team could not look at factors that influence dropouts over time. For instance, the team could not identify whether changes in learner reading outcomes were associated with dropouts given that data only exists at one point in time (a snapshot). Also, all learners assessed in the baseline are enrolled in school. Thus, dropouts had to be looked at for a snapshot and for either all learners in a school or all leaners in a household. In order to incorporate household factors into the analysis of dropouts, the evaluation team decided to use household dropouts as the unit of analysis. The evaluation team defined "dropping out" as not finishing Standard 8 for the purposes of this study. The reason for this is the study's focus on primary school education outcomes. Standard 8 is the last standard of primary school, and, thus, analysis of those who remained in school through this standard allows evaluators to separate primary-school graduates from primary-school dropouts. Household dropout rates were calculated for females, males, and all youth. For the purposes of this analysis, "youth" was limited only to

those between 13 and 25 living in each household, as 13 is the earliest age at which a learner can graduate from Standard 8. Since dropout rates were measured for youth in household questionnaires, the models are limited to household and community factors. Since household and community questionnaires reflect current conditions and dropout rates may include youth who dropped out nearly 20 years prior, the evaluation team does not expect the model to fully explain dropout. Further analysis will be conducted on dropout rates at the school level during the midline and endline.

Table 17. Individual and Household Predictions on Youth Dropout Rates Per Household

Independent Variable	Coefficient	Standard Error	*t*-Statistic	*p*-Value	[95% Confidence Interval]	
Access to reading materials at home	-0.0059211 **	0.0025367	-2.33	0.02	-0.0108938	-0.0009483
Socioeconomic status (Asset Index)	-0.0025839 ***	0.0005165	-5	0	-0.0035964	-0.0015715
Amount spent on school expenses	8.55E-08	2.56E-07	3.30E-01	0.738	-4.16E-07	5.87E-07
Household member volunteers at least one hour per month in learner's school	-0.0037741 *	0.0022893	-1.65	0.099	-0.0082619	0.0007137
Learner attended preschool for two or more years	0.00336	0.0029224	1.15	0.25	-0.0023688	0.0090889
At least one household member has graduated from primary school	-0.0080839 ***	0.0027762	-2.91	0.004	-0.013526	-0.0026418
At least one household member has graduated from Junior Level 2	-0.0215689 ***	0.0037185	-5.8	0	-0.0288583	-0.0142795

At least one household member has graduated from Senior Level 2	-0.0106766 **	0.0041842	-2.55	0.011	-0.0188789	-0.0024743
Learner's diet includes protein	0.0012863	0.0013229	0.97	0.331	-0.0013071	0.0038797
Distance to a clinic	0.0004442	0.0009266	0.48	0.632	-0.0013723	0.0022606
Number of youth in household	0.0070387 ***	0.0007135	9.87	0	0.0056401	0.0084374
Percentage of youth in household who are male	0.2084602 ***	0.0081605	25.55	0	0.1924633	0.2244571
Constant	0.0068226	0.0047639	1.43	0.152	-0.002516	0.0161613

*p-value < 0.1 **p-value < 0.05 ***p-value <0.01

Observations	7061
***R*-squared**	0.1359
Adjusted *R*-squared	0.1345

Dropout Results. As can be viewed above in Table 17, a number of household and learner factors appear to have a statistically significant relationship with the household dropout rate of individuals between the ages of 14 and 25. The evaluation team found that living in a household with reading materials predicts a decrease in dropout rates by 0.006 percentage points (*p*-value 0.02). Likewise, higher socioeconomic status significantly predicts a decrease in youth dropout rates per household (*p*-value 0.000). Dropout rates decrease 0.004 percentage points if someone from the learner's family reports volunteering at school for at least one hour a month (*p*-value 0.099). Highest household education is also associated with dropout rates. Dropout rates decrease by 0.008 percentage points if at least one person has graduated from primary school in the household (*p*-value 0.004) and decrease by 0.022 percentage points if someone has graduated from Junior Level 2 or higher. The number of youth between 14 and 25 increases the dropout rates in the household by 0.007 percentage points for each additional youth. Household dropout rates also increase as the percentage of males in the household increases. Differences in dropout rates by sex are examined more thoroughly in the regression in Table 18.

Table 18. Individual and Household Effects on Dropout Rates Per Household

	Girl Dropout Rates				Boy Dropout Rates			
Independent Variable	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**
Access to reading materials at home	-0.0457806 ***	0.0176362	-2.6	0.009	-0.0415339 ***	0.0156539	-2.65	0.008
Socioeconomic status (Asset Index)	-0.0138392 ***	0.0035833	-3.86	0	-0.0089016 ***	0.0032276	-2.76	0.006
Amount spent on school expenses	-2.28E-06	1.97E-06	-1.16	0.247	-1.37E-07	1.69E-06	-0.08	0.936
Household member volunteers at least one hour per month in learner's school	-0.0351062 **	0.0161025	-2.18	0.029	0.0038603	0.0145393	0.27	0.791
Learner attended preschool for two or more years	0.0146519	0.0203622	0.72	0.472	0.0042201	0.0191113	0.22	0.825

At least one household member has graduated from primary school	-0.1224529 ***	0.0199321	-6.14	0	-0.0973162 ***	0.0178155	-5.46	0
At least one household member has graduated from Junior Level 2	-0.1093307 ***	0.0243804	-4.48	0	-0.1103297 ***	0.0227261	-4.85	0
At least one household member has graduated from Senior Level 2	-0.0395657	0.0269721	-1.47	0.143	-0.043524 *	0.0250952	-1.73	0.083
Learner's diet includes protein	0.0299953 ***	0.0093345	3.21	0.001	-0.0001867	0.0082405	-0.02	0.982
Distance to clinic	-0.0011849	0.0066522	-0.18	0.859	0.0057663	0.0058585	0.98	0.325
Number of youth in household	0.0141154 ***	0.0051935	2.72	0.007	0.0168095 ***	0.0046556	3.61	0
% of HH youth who are male	-0.0314517	0.0671365	-0.47	0.639	0.1734229 **	0.0681284	2.55	0.011
Constant	0.3669915 ***	0.0371014	9.89	0	0.2058058 ***	0.0406305	5.07	0

*p-value < 0.1 **p-value < 0.05 ***p-value < 0.01

Girl Dropout Rates

Observations	2855
R-squared	0.1
Adjusted R-squared	0.09

Boy Dropout Rates

Observations	2917
R-squared	0.08
Adjusted R-squared	0.08

While several factors are predictive of dropout rates for both boys and girls, there were also factors that are predictive of dropouts rates for one sex but not the other. Living in a household where there was access to reading materials at home predicts a decrease in girl dropout rates (*p*-value 0.009) and in boy dropout rates (*p*-value 0.008). Higher socioeconomic status predicts a decrease in dropout rates for both girls (*p*-value 0.000) and boys (0.006). The number of youth in a household predicts a significant increase in the percentage of dropouts for both girls (*p*-value 0.007) and boys (*p*-value 0.000). Dropout rates decrease for girls (*p*-value 0.000) and for boys (*p*-value 0.000) if at least one person has graduated from primary school in the household. Rates also drop for girls (*p*-value 0.000) and boys (*p*-value 0.0000) if someone has graduated from Junior Level 2 or higher.

Of factors explored, four are significant for one sex and not the second. Dropout rates for boys also decrease by 4.3 percent if at least one family member has graduated from Senior Level 2 (*p*-value 0.083) but do not predict a significant difference for girls. Girls dropout rates increase by 2.9 percent if the family's diet includes protein (*p*-value 0.001), but there is no correlation between protein and boys' scores. If a family member volunteers at least one hour a month at school, it predicts a 3.5 percent decrease in girls' dropout rates (*p*-value 0.028) but does not significantly predict changes dropout rates for boys. The percentage of boys in the household significantly predicts higher dropout rates for boys (*p*-value 0.011) but does not predict dropout rates for girls.

Repetition Model. Repetition rates were available for the learners who took the reading assessment, so the repetition correlations take into account individual, household, and school factors. Five repetition models with the same independent variables but different dependent variables were used to find out what predicts repetition rates for different populations. The first regression uses whether a learner has ever repeated a grade as the dependent variable to see what factors predict any repetition. More than half of learners had repeated at least one standard. The second and third models looked at repetition of Standards 1 and 3. Because learners who had repeated the current standard were eliminated from the study, Standards 1 and 3 were selected. Different factors may predict repetition in different standards because of both societal expectations and roles and because of curriculum. Learners who repeat Standard 1 also may be more likely to drop out, so it is important to examine Standards 1 and 3 separately. This study also seeks to examine differences in repetition rates between girls and boys, so the final two models examine what factors predict repetition for boys and for girls separately. Table 19 below shows a breakdown of the number of respondents that repeated and did not repeat by Standard repeated and sex of repeaters.

Table 19. Repetitions of Standards by Subgroup

	Any Repetition	Repetition of Standard 1	Repetition of Standard 3	Girls	Boys
Repeated	4793	2665	1163	2157	2635
Did not repeat	3395	4847	2644	1614	1714
Total	**8188**	**7512**	**3807**	**3771**	**4349**

Not all learners reported what standards had been repeated. Because the current standard could not have been repeated, all repetitions had to have happened at least a year before the survey was conducted, so some variables measuring the current situation in households or schools may not accurately reflect the situation of learners when they repeated. These models operate on the assumption that most of these factors have not changed significantly. Also, caretakers were asked what standards had been repeated, and not all caretakers knew, reducing the number of observations for the repetition of Standards 1 and 3 in particular. The midline and endline reports will further explore what predicts and what causes repetition and will be able to incorporate more timely data into the analysis.

Finally, the *R*-squared and adjusted *R*-squareds are relatively low, indicating that the factors included in the regression only explain a small percentage of the dependent variables. See Table 20 below for the results of the repetition rate model.

Table 20. Community, Household, and School Factors Predicting Repetition Rates

Independent Variable	Coefficient	Standard Error	*t*-Statistic	*p*-Value	[95% Confidence Interval]	
Standard	0.0540882***	0.0072183	7.49	0	0.0399368	0.0682396
Access to reading materials at home	-0.0088062	0.0160705	-0.55	0.584	-0.0403124	0.0226999
Household member volunteers at least one hour per month in learner's school	-0.0037908	0.0146913	-0.26	0.796	-0.0325931	0.0250114
At least one household member has graduated from Junior Level 2	-0.0777164***	0.0165947	-4.68	0	-0.1102501	-0.0451827
Number of days the learner did not eat anything in the last week	0.0070625	0.0154916	0.46	0.648	-0.0233086	0.0374336
Learner gets tired at school	-0.0135437	0.0148012	-0.92	0.36	-0.0425613	0.015474
Number of weeks the learner was sick in the last month	0.0117865	0.0121287	0.97	0.331	-0.0119917	0.0355647
Learner goes to the doctor when sick	-0.014439	0.0562733	-0.26	0.798	-0.1247622	0.0958842
Level of food insecurity (principal component)	0.0020949	0.0042475	0.49	0.622	-0.0062323	0.0104222
Dietary diversity	-0.003338	0.004699	-0.71	0.478	-0.0125503	0.0058744

Learner's diet includes protein	0.0035991	0.0112944	0.32	0.75	-0.0185434	0.0257417
School resources	0.0161908***	0.0055826	2.9	0.004	0.0052462	0.0271354
Separate latrines for girls	-0.0233009	0.0232954	-1	0.317	-0.0689714	0.0223695
PTA meets at least monthly	0.0084827	0.0146428	0.58	0.562	-0.0202243	0.0371897
Amount spent on school expenses	-0.0000017	1.77E-06	-0.96	0.339	-5.17E-06	1.78E-06
Socioeconomic status (Asset Index)	-0.0194834***	0.0060453	-3.22	0.001	-0.0313352	-0.0076315
Learner attended preschool for two or more years	-0.0144105	1.89E-02	-0.76	0.447	-5.15E-02	2.27E-02
Sex	-0.0332588**	0.0143369	-2.32	0.02	-0.0613662	-0.0051514
Constant	0.56255	0.0678469	8.29	0	0.4295366	0.6955633

Observations	4520
***R*-squared**	0.0297
Adjusted *R*-squared	0.0258

Based on this model, learners in Standard 4 were .054 percentage points more likely to repeat a grade than Standard 4 learners (*p*-value 0.000), as would be expected considering that Standard 4 learners have had more opportunities to repeat a standard. Learners were also 0.078 percentage points less likely to repeat if someone in their household had graduated from Junior Level 2 or had further education (*p*-value 0.000). Socioeconomic status had a negative correlation with repetition (*p*-value = 0.001).

An increase in school resources predicts an increase in repetition (*p*-value 0.004). This may be because teachers with more resources are better able to evaluate the progress of learners and may be more likely to hold a learner back so that they repeat a standard.

It appears that sex is a significant predictor of repetition and that boys are 0.033 percentage points more likely to repeat a grade than girls (*p*-value 0.019). Repetition rates of boys and girls will be examined more extensively after the models for Standard 1 and Standard 3 repetition are reviewed in Table 21 below.

Table 21. Community, Household, and School Factors Predicting Repetition Rates for Standards 1 and 3

	Standard 1 Repetition				Standard 3 Repetition			
Independent Variable	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**
Standard	-0.0571381***	0.0071744	-7.96	0	NA			
Access to reading materials at home	-0.0051039	0.015967	-0.32	0.749	-0.0368863*	0.020506	-1.8	0.072
Family investment in learner education (Asset Index)	-0.0051276	0.0145993	-0.35	0.725	0.0036461	0.0197836	0.18	0.854
At least one household member has graduated from Junior Level 2	-0.0812356***	0.0164859	-4.93	0	-0.0180384	0.0223789	-0.81	0.42
Number of days the learner did not eat anything in the last week	-0.0015634	0.0153903	-0.1	0.919	0.0121161	0.0211018	0.57	0.566
Learner gets tired at school	0.0088354	0.0147085	0.6	0.548	-0.0116606	0.0204985	-0.57	0.57
Number of weeks the learner was sick in the last month	0.007204	0.012085	0.6	0.551	0.0104709	0.0163479	0.64	0.522

Learner goes to the doctor when sick	-0.0292138	0.0558551	-0.52	0.601	0.0260165	0.0756023	0.34	0.731
Level of food insecurity (principal component)	0.0040823	0.0042233	0.97	0.334	-0.0012169	0.0058957	-0.21	0.836
Dietary diversity	0.0016466	0.0046711	0.35	0.724	0.0012171	0.0064567	0.19	0.851
Learner's diet includes protein	-0.0210562*	0.0112432	-1.87	0.061	-0.0023136	0.0154975	-0.15	0.881
School resources	0.012675***	0.005545	2.29	0.022	0.0040922	0.0075773	0.54	0.589
Separate latrines for girls	-0.0007169	0.0231263	-0.03	0.975	0.0371851	0.0314661	1.18	0.237
PTA meets at least monthly	0.0089882	0.0145507	0.62	0.537	-0.0386034*	0.0198305	-1.95	0.052
Amount spent on school expenses	0.000000801	1.76E-06	0.45	0.649	-0.000000937	2.19E-06	-0.43	0.668
Socioeconomic status (Asset Index)	-0.0160637***	0.0060117	-2.67	0.008	-0.022454***	0.0082667	-2.72	0.007
Learner attended preschool for 2 or more years	-0.0255464	0.0188146	-1.36	0.175	0.0065168	0.0258595	0.25	0.801
Sex	-0.0235045*	0.014248	-1.65	0.099	0.0053598	0.0193264	0.28	0.782
Constant	0.6035155	0.0673607	8.96	0	0.2686707	0.0861314	3.12	0.002

Observations	4510
***R*-squared**	0.0313
Adjusted *R*-squared	0.027

Observations	2322
***R*-squared**	0.013
Adjusted *R*-squared	0.0048

It appears that there is little overlap between factors that predict repetition of Standard 1 and repetition of Standard 3. The only factor that predicts repetitions of both Standards 1 and 3 is socioeconomic status (*p*-values of 0.008 and 0.007, respectively). Current standard is statistically significant for predicting Standard 1 repetition. Learners in Standard 4 are 0.057 percentage points less likely to have repeated Standard 1 than learners in Standard 2 (*p*-value 0.000). This may be because learners who repeat Standard 1 are less likely to stay in school until Standard 4. Learners were 0.081 percentage points less likely to repeat Standard 1 if someone in their household had graduated from Junior Level 2 or higher in school (*p*-value 0.000). However, highest household education does not predict repetition for Standard 3. A learner that has protein in his or her diet is 0.021 percentage points less likely to repeat Standard 1 (*p*-value 0.061) but no less likely to repeat Standard 3.

For Standard 4 learners, only three factors in the model predicted repetition. Learners with access to reading materials at home were 0.036 percentage points less likely to repeat Standard 3 (*p*-value 0.072). Learners were also 0.039 percentage points less likely to repeat if they attend a school where the PTA meets at least monthly (*p*-value 0.05). The third factor was socioeconomic status.

Males were 0.024 percentage points more likely than females to repeat Standard 1 (*p*-value 0.099) but sex did not predict repetition for Standard 3. In Table 22, differences in why girls and boys repeat are examined.

Table 22. Community, Household, and School Factors Predicting Repetition Rates For Girls and Boys

	Girl Repetition				Boy Repetition			
Independent Variable	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**	**Coefficient**	**Standard Error**	***t*-Statistic**	***p*-Value**
Standard	0.0671907***	0.0101918	6.59	0	0.0489298***	0.0092118	5.31	0
Access to reading materials at home	-0.0005483	0.0215321	-0.03	0.98	-0.035465*	0.0196241	-1.81	0.071
Family investment in learner education (Asset Index)	-0.0123214	0.0193476	-0.64	0.524	-0.0029827	0.0175674	-0.17	0.865
At least one household member has graduated from primary school	-0.0080887	0.0233785	-0.35	0.729	-0.0034882	0.0208496	-0.17	0.867
At least one household member has graduated from Junior Level 2	-0.0535822*	0.0313321	-1.71	0.087	-0.0888768***	0.0286335	-3.1	0.002
At least one household member has graduated from Senior Level 2	-0.0475707	0.0350458	-1.36	0.175	0.0191142	0.032178	0.59	0.553

Number of days the learner did not eat anything in the last week	0.0072404	0.0209239	0.35	0.729	0.0210549	0.0181201	1.16	0.245
Learner gets tired at school	-0.020309	0.0196353	-1.03	0.301	-0.001382	0.0177347	-0.08	0.938
Number of weeks the learner was sick in the last month	0.0085038	0.0152045	0.56	0.576	0.0124345	0.0145215	0.86	0.392
Learner goes to the doctor when sick	0.0832342	0.0699469	1.19	0.234	-0.1615725**	0.0680808	-2.37	0.018
Level of food insecurity (principal component)	0.002144	0.0052738	0.41	0.684	0.008932*	0.0046944	1.9	0.057
Dietary diversity	-0.0161948***	0.0061926	-2.62	0.009	0.0098215*	0.005642	1.74	0.082
Learner's diet includes protein	0.0186157	0.0151757	1.23	0.22	-0.0140592	0.0135859	-1.03	0.301
School resources	0.0164227**	0.0071314	2.3	0.021	0.0160842**	0.00669	2.4	0.016
Separate latrines for girls	-0.0626312*	0.0339685	-1.84	0.065	-0.0097715	0.0269511	-0.36	0.717
Learner-to-teacher ratio	0.0006562**	0.0002645	2.48	0.013	0.000021	0.0002279	0.09	0.926

PTA meets at least monthly	0.0156155	0.0193607	0.81	0.42	0.0160325	0.0177654	0.9	0.367
Amount spent on school expenses	-0.00000166	2.28E-06	-0.72	0.469	-0.00000529***	1.67E-06	-3.18	0.002
Constant	0.4328236***	0.0904498	4.79	0	0.6701663***	0.0855271	7.84	0

Observations	2608
***R*-squared**	0.0326
Adjusted *R*-squared	0.0259

Observations	3102
***R*-squared**	0.0277
Adjusted *R*-squared	0.022

It appears that there is little difference between what factors predict repetition rates for both boys and girls. Female Standard 4 learners were 0.034 [67]percentage points more likely to have repeated a standard than Standard 2 learners (*p*-value = 0.000) and male Standard 4 learners were 0.021 percentage points [68]more likely to have repeated a standard than Standard 2 learners (*p*-value = 0.000). Girls were 0.069 percentage points less likely to repeat if at least one household member had graduated Junior Level 2 (*p*-value = 0.006) and boys were 0.084 percentage points less likely (*p*-value = 0.000).

Socioeconomic status predicts a significant decrease in repetition rate for girls (*p*-value = 0.000) but does not predict change in repetition rates for boys. On the other hand, amount spent on school expenses predicts a significant decrease in repetition for boys (*p*-value 0.084) but no decrease for girls.

Overall, many of the same factors predict repetition rates and dropout rates of girls and boys. However, there are some factors that only predict rates for one sex. Sex is also a significant predictor of repetition when included in the model.

[67] Note: because Standard increase by two points the coefficient was divided by two.

[68] Note: because Standard increase by two points the coefficient was divided by two.

VI. COMPARISON OF TREATMENT AND CONTROL GROUPS

As detailed above, evaluators employed propensity score matching (PSM) to reduce the effect of selection bias, or initial differences between the treatment and comparison group, on impact estimates. To estimate propensity scores, evaluators used baseline data to build a model of selection into the program and then match each treatment unit with comparison units with similar propensity scores. To build the model, evaluators included all baseline variables that are hypothesized to be related to both selection into treatment and the outcomes of interest. The evaluation team then tested to see if there are significant differences between treatment and control groups after matching. For the most part, the evaluation team found there were not significant differences of note.

TREATMENT LEVELS

Since this evaluation will compare multiple treatment and comparison groups, evaluators estimated separate propensity scores and identified separate matched comparisons for several different treatment and control groups. Each level of treatment was compared to control observations in the same level, where feasible, and then to the "pure" control, where applicable. These levels were then elaborated upon in the background section of the report. The "pure" control group includes control households from Levels 3 and 4 because, while controls from Level 1 will also receive the GHI intervention, these controls will not receive any interventions. PSM models have been identified to balance each of the following treatment and control group comparisons:

- Level 1 Treatment vs. Level 1 Control
- Level 3 Treatment vs. Level 3 Control
- Level 4 Treatment vs. Level 4 Control
- Level 1 Treatment vs. Pure Control
- Level 2 Treatment vs. Pure Control
- Level 3 Treatment vs. Pure Control

Because it is important to match treatment learners with other learners in their standard, learners were disaggregated by standard for each level. The evaluation team therefore created two sets of matches (one for each standard) for each of the six levels described above.

PSM VARIABLES

For this propensity score matching model, learners were matched based on baseline values of variables from four categories: individual demographics, family demographics, family investment in learner, and learner EGRA scores. The propensity score matching did not include information at a standard or school level from the teacher, head teacher, classroom observation, or school climate data because the level of variance is much lower for these variables, as many learners within the same treatment or control group receive the same scores for each of these factors. Some of the key variables from the class level and school level data are included in the balancing tables to review whether the PSM model reduces bias in these areas as well.

INDIVIDUAL DEMOGRAPHICS

1. **Sex**. This can have an important effect on reading scores and is also one of the key components this report will examine.
2. **Whether a learner is older than most in his or her standard**. Older learners may have experienced various setbacks that held them back in school, and thus their learning experience may be very different. The cutoff was 12 for Standard 2 learners and 14 for Standard 4 learners.
3. **Whether Chichewa is a primary language.** If Chichewa is not a learner's first language, it may be significantly more difficult to learn to read. This is only included where applicable because some learners do not speak Chichewa as a primary language.
4. **The number of weeks a learner has stayed home from school for over a week in the past month due to illness**. While this measure may be temporal, it is used as a general measure of the learner's health, which will help with evaluation of the GHI intervention when in conjunction with the GHI intervention.

FAMILY DEMOGRAPHICS

5. **Whether a family farms**. One of the interventions that is supposed to be measured in combination with the EGRA intervention is FtF, an agricultural intervention, so the evaluation team wanted to match those who would be receiving this treatment against those who would not.
6. **Whether anyone in the household had graduated from primary school.** Often, the education of parents or elders sets the expectations for the next generation of

learners. Analysis has shown that whether anyone in the household has graduated from primary school is a significant predictor of higher test scores.

7. **The index for principle component analysis of socio-economic status**. It is generally agreed that those benefiting from higher socioeconomic status and more affluence do better in school because of the many advantages that they have.

FAMILY INVESTMENT IN LEARNER

8. **The total amount of expenditure on a learner's schooling within a year**. This signifies both socioeconomic status and the amount that caregivers and parents are willing and able to invest financially in a learner's education.
9. **Whether or not a learner attended two or more years of preschool**. The evaluation team's regressions show that having two or more years of preschool helped learners do significantly better in school than having one or less years of schooling.
10. **Whether a learner receives help with homework at home**. This is important because it shows both the ability and willingness of parents to actively engage in their learner's schooling.

LEARNER READING SCORES

11. **Whether a learner met EGRA Coordinating Committee benchmarks for the listening comprehension subtask of EGRA.** Listening comprehension was the most basic skill of the EGRA subtasks and many learners met these EGRA Coordinating Committee benchmarks.
12. **Whether a learner met EGRA Coordinating Committee benchmarks for the oral reading fluency subtask of EGRA.** Few learners met the oral reading fluency benchmark, but it is being used as the measure of learner reading fluency and success in the regressions and will be used at the endline to show success, so its inclusion is important. Because of low numbers in Standard 2, this is only included for Standard 4 learners.
13. **Whether a learner received a zero score for the oral reading fluency subtask of EGRA.** Because so few learners met the oral reading benchmark and so many learners received zero scores, this seemed important to be able to suitably match learners by ability.

ESTIMATION OF PROPENSITY SCORE

The evaluation team used a probit regression to determine how the variables listed above were related to whether a learner attends a treatment school. Based on that regression, the evaluation team estimated how likely a given learner is to attend a treatment school based solely on his or her baseline characteristics. This likelihood is the propensity score.

A probit regression was estimated for each of the treatment versus comparison groups, and the evaluation team finds that many of the variables significantly predict selection into the treatment and comparison groups. This indicates that there was likely selection bias and confirms the need to use propensity score matching in order to reduce selection bias. There are also a few instances in which variables significantly predict treatment in one model and do not in another.

PSM BALANCE

After the propensity scores were estimated, balance tables were produced to see whether the PSM model effectively reduces bias. These balance tables include both the variables included in the model and other variables thought to be important that either were not considered predictors of treatment or did not balance.[69] Overlap was also checked to see whether there were observations outside the area of common support so that they could be dropped.

The balances tables, which are displayed in Annex 10, show that the matching reduced bias for most variables, including both those that were included in the model and those that were only added to the balancing tables. As with the probit regression, there was variation across the different models. A variable may have reduced bias in one model and not another. However, there was a reduction in bias for most variables included in all of the models. The areas of common support also overlapped for the vast majority of variables, so few variables were dropped because they were outside the area of common support.

LIMITATIONS

While the PSM model reduces bias, it does not fully eliminate bias. The PSM model also takes into account observed variables only. It did not effectively incorporate head teacher, classroom, or teacher data as discussed previously. It is important to note that this is a preliminary match based on the current data and is subject to change during the midline and endline and may include school-level variables in the future. The matching is included in the

[69] This includes a number of variables at the school level that were unable to balance because of high inter-school variance and no intra-school variance.

baseline report in order to demonstrate that bias that was introduced because random sampling was not used can be significantly reduced for the treatment and control groups used in this study.