



## **Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E)**

### **Early Grade Reading Assessment (EGRA) in Three Mother Tongues (Aff Somali, Sidamu Affoo and Tigrigna) Data Analytic Report**

Contract No. AID-663-C-15-00001  
Revised April 13, 2016

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

Reading for Ethiopia's Achievement Developed  
Monitoring and Evaluation (READ M&E)

**Early Grade Reading Assessment (EGRA) in Three Mother Tongues  
(Aff Somali, Sidamu Affoo and Tigrigna)  
Data Analytic Report**

**Submitted to:  
U.S. Agency for International Development/Ethiopia  
Marc Bonnenfant, Contracting Officer's Representative (COR)**

**Prepared by:  
American Institute for Research  
1000 Thomas Jefferson Street, NW  
Washington DC 20007**

**Contract No. AID-663-C-15-00001**

March, 2016  
Revised April 13, 2016

This study of early grade reading assessment is made possible by the support of the American people through the United States Agency for International Development (USAID). The contents of this report are the sole responsibility of American Institutes for Research (AIR) and do not necessarily reflect the views of USAID or the United States Government.

## Table of Contents

1	Executive Summary .....	1
2	Introduction.....	2
3	Background .....	3
3.1	Why Test Early Grade Reading?.....	3
3.2	Purpose of EGRA.....	3
3.3	What EGRA Measures .....	3
3.4	Limits of EGRA .....	4
4	Development of the 2015 EGRA tool.....	5
5	Design and Methodology .....	6
5.1	The Sample.....	6
5.2	Data Collection.....	6
5.2.1	Data Collection Training.....	6
5.2.2	Data collection methodology .....	6
5.3	Instruments .....	7
6	Findings.....	9
6.1	Discontinuity Rate.....	9
6.1.1	Discontinuity Rate in Aff Somali .....	9
6.1.2	Discontinuity Rate in Sidamu Affo.....	9
6.1.3	Discontinuity Rate in Tigrigna.....	10
6.2	Proportions Meeting Benchmarks .....	10
6.2.1	Proportions Meeting Benchmarks in Aff Somali by Grade.....	11
6.2.2	Proportions Meeting Benchmarks in Sidamu Affo by Grade.....	11
6.3	Mean Fluency Scores .....	12
6.3.1	Mean Fluency Scores in Aff Somali by Grade .....	13
6.4	Mean Fluency Scores by Gender .....	14
6.4.1	Mean Fluency Scores in Aff Somali by Gender .....	14
6.4.2	Mean Fluency Scores in Sidamu Affo by Gender .....	15
6.5	Mean Scores of Untimed Tasks by Language and Grade .....	16
6.5.1	Mean Scores of Untimed Tasks in Aff Somali by Grade .....	16
6.5.2	Mean Scores of Untimed Tasks in Sidamu Affo by Grade .....	17
6.5.3	Mean Scores of Untimed Tasks in Tigrigna by Grade .....	17
6.6	Mean Scores of Untimed Tasks by Language and Gender .....	18
6.6.1	Mean Scores of Untimed Tasks in Aff Somali by Gender .....	18
6.6.2	Mean Scores of Untimed Tasks in Sidamu Affo by Gender .....	18

6.6.3	Mean Scores of Untimed Tasks in Tigrigna by Grade .....	19
6.6.4	Comparisons between RC (%) and LC (%) Scores .....	19
7	Conclusion .....	21

**List of Tables**

Table 1.	Sample Distribution by Grade and Language.....	6
----------	------------------------------------------------	---

**List of Figures**

Figure 1.	Timed Subtasks Proportion of Discontinuity in Aff Somali by Grade .....	9
Figure 2.	Timed Subtasks Proportion of Discontinuity in Sidamu Affo by Grade.....	10
Figure 3.	Timed Subtasks Proportion of Discontinuity in Tigrigna by Grade.....	10
Figure 4.	Proportion at Various Levels of Reading Proficiency in Aff Somali by Grade.....	11
Figure 5.	Proportion at Various Levels of Reading Proficiency in Sidamu Affo by Grade ....	12
Figure 6.	Proportion at Various Levels of Reading Proficiency in Tigrigna by Grade .....	12
Figure 7.	Mean Fluency Scores in Aff Somali by Grade.....	13
Figure 8.	Mean Fluency Scores in Sidamu Affo by Grade.....	13
Figure 9.	Mean Fluency Scores in Tigrigna by Grade.....	14
Figure 10.	Mean Fluency Scores in Aff Somali by Gender.....	15
Figure 11.	Mean Fluency Scores in Sidamu Affo by Gender.....	15
Figure 12.	Mean Fluency Scores in Tigrigna by Gender.....	16
Figure 13.	Mean Scores of the Untimed Tasks in Aff Somali by Grade .....	16
Figure 14.	Mean Scores of the Untimed Tasks in Sidamu Affo by Grade .....	17
Figure 15.	Mean Scores of the Untimed Tasks in Tigrigna by Grade .....	17
Figure 16.	Mean Scores of the Untimed Tasks in Aff Somali by Gender .....	18
Figure 17.	Mean Scores of the Untimed Tasks in Sidamu Affo by Gender .....	19
Figure 18.	Mean Scores of the Untimed Tasks in Tigrigna by Gender .....	19
Figure 19.	Reading (RC) and Listening Comprehension (LC) Scores by Language .....	20
Figure 20.	Reading (RC) and Listening Comprehension (LC) Scores by Language .....	21

**Acronyms**

AIR	American Institutes for Research
EGRA	Early Grade Reading Assessment
LC	Listening Comprehension
PA	Phonemic Awareness
RC	Reading Comprehension
MOE	Ministry of Education
NEAEA	National Educational Assessment and Examinations Agency
NLA	National Learning Assessment
ORF	Oral Reading Frequency
NR	Non Readers
RWLC	Reading With Limited Comprehension
RWC	Reading With Comprehension
FRWC	Fluent Readers with Comprehension
READ M&E	Reading for Ethiopia's Achievement Developed Monitoring and Evaluation
RSEB	Regional State Education Bureau
USAID	United States Agency for International Development

## 1 Executive Summary

This Mini EGRA (Early Grade Reading Assessment) is a snapshot of the reading fluency and comprehension levels of Grades 2 and 3 students in three selected languages: *Aff Somali*, *Sidamu Affo* and *Tigrigna*. Data was collected from 45 schools (15 from each language). A total of 1,719 Grade 2 and 3 students were the subjects of the study. About equal number of girls and boys were selected. Electronic data was collected using Nexus Tablets with Tangerine Software. The Mini EGRA was conducted parallel with the pilot study whose main purpose was field testing e-data collection.

The data collecting tool for each language was composed of timed (*letter name knowledge, familiar words reading, invented words reading and passage words reading*) and untimed subtasks (*phonemic awareness, reading comprehension and listening comprehension*). Each tool was presented to the child one-by-one orally. Each timed task is given 60 seconds and fluency levels as correct letters per minute or correct words per minute. The following presents the major findings of the data analysis. Comparisons are made by grade levels and gender for each language.

**Proportions Discontinued:** In all the three languages, the sub task that showed the lowest percentage of early stop cases (zero score) was letter name knowledge. In all case the proportions of discontinuity were found larger in Grades 2 than Grade 3.

In Aff Somali, the largest proportions of early stop cases appeared in invented words reading (65.1%) followed by familiar words reading (61%). In Sidamu Affu, the largest proportions of early stop cases appeared in familiar words reading (49.8%) followed by invented words reading (46.2%). In Tigrigna, the proportions of early stop cases in familiar words reading, invented words and oral reading passage were more or less equal. In each of the three subtasks about a quarter of students were found unable to read a single word within a minute.

In Grade 2, not a single student achieved the highest levels of benchmark. In the same way with the other two languages the proportion of ‘nonreaders’ has decreased while the other levels increased in Grade 3.

**Proportions Meeting Benchmarks:** Based on a combination of the oral reading and reading comprehension scores the students were categorized as non-readers, reading with limited comprehension, reading with comprehension, and reading fluently with comprehension. In all three languages the proportion of ‘non-reader’ and those ‘reading with limited comprehension’ were very high. Only very few students (less than 1 out of 10) were able to read with comprehension. In Tigrigna, more than 98% of students fall in ‘nonreader’ and ‘reading with limited comprehension’ categories. Looking at the grade level comparisons, in all the three languages, the proportion of ‘nonreaders’ decreased from grade 2 to grade 3, while the other levels increased in Grade 3. This shows that most students are improving their reading skills from grade 2 to grade 3.

**Mean Fluency Scores:** In Aff Somali, on the average, students were able to read very few correct familiar words per minute or correct invented words per minute. They were able to read only about seven correct words per minute in each case. In Sidamu Affo, the least mean score was observed in invented words (10.7) while the highest in letter sound (50.4). In Tigrigna, the least mean score was observed in invented words (12) while the highest in letter name (28.2).

In all the three languages the mean fluency scores of Grade 3 students were found higher than that of Grade 2. In each subtask Grade 3 students were able to read better than Grade 2 indicating grade gain however overall reading skills are still very low. In Aff Somali, boys performed better than girls and the mean differences were statistically significant except in oral reading fluency. In Sidamu Affo and Tigrigna, girls performed slightly better than boys.

**Mean Scores of the Non-Fluency (Untimed) Tasks by Language and Grade:** The untimed subtasks are: Phonemic Awareness (PA), Reading Comprehension (RC) and Listening Comprehension (LC). In all the three languages, the mean differences between the two grade levels were found statistically significant in favor of Grade 3.

In Aff Somali, in all the three subtasks, boys performed better than girls. The widest gap was observed in phonemic awareness (16.2%) followed by reading comprehension (7%). In Sidamu Affu in reading comprehension girls performed slightly better than boys but in the other two subtasks no major differences are observed. In Tigrigna, girls performed slightly better than boys in reading comprehension while in the other two subtasks boys performed slightly better than girls but the mean differences are negligible.

**Comparisons between Reading Comprehension and Listening Comprehension Scores:** In all the three languages there exist very wide gap between the two in favor of listening comprehension. About 70% of the students obtained a zero score on the Reading Comprehension subtask, reflecting the fact that nearly seven out of ten students were not able to answer even one comprehension question correctly. The accepted benchmark for comprehension is 80 percent or more correct on the Reading Comprehension Task. Average student scores fell far below this threshold.

**Conclusions:** The ultimate goal of reading is constructing meaning. The findings suggest that although some students were able to respond correctly to the Reading Comprehension subtask questions, they were very few in number. There is still a high percentage of students who cannot read enough words with in one minute to develop an understanding of what they read. Students appear to understand their mother tongue as reflected by relatively higher scores on the Listening Comprehension subtask.

## 2 Introduction

The Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E) is one of the four projects on early grade reading launched by USAID Ethiopia. READ M&E focuses on monitoring and evaluation of early grade reading and writing as well as the M&E needs of the overall USAID READ program. The project closely works in partnership with the Ethiopian Federal Ministry of Education (MoE) and the National Education Assessment and Examinations Agency (NEAEA). The project also works in collaboration with Regional State Education Bureaus (RSEBs).

One of the specific objectives of the READ M&E project is to gather data to monitor reading skill for seven local languages and English. In this study three languages namely Aff Somali, Sidamu Affo and Tigrigna were selected to conduct mini EGRA parallel with the Midterm EGRA Pilot. The main purpose of the study was to describe the reading fluency and comprehension levels of the sample students in each language. The sample is not representative of each language and we do not make any generalizations.

This report describes the EGRA, what it attempts to measure, and presents the results of reading performance in the three mother tongue languages.

### **3 Background**

#### **3.1 Why Test Early Grade Reading?**

The ability to read and understand simple text is one of the most fundamental skills that a child can learn. Without basic literacy, there is little chance that a child can escape the intergenerational cycle of poverty. Yet in many countries, students enrolled in school for as many as six years are unable to read and understand simple text. Recent evidence indicates that learning to read both *early* and at a sufficient *rate* are essential for learning to read well. Acquiring literacy becomes more difficult as students grow older. Children who do not learn to read in the first few grades are more likely to repeat grades and eventually drop out, leading to the gap between early readers and non-readers increasing over time.

When students are first learning to read, they must learn the letters of a language, the forms of those letters, and the sounds associated with each letter and then apply this knowledge to decode (or “sound out”) new words. At the same time, they are gaining familiarity, or automaticity, with words that they can then read by sight, without having to decode them. By the end of this first phase of reading development, students on a normal development trajectory develop sufficient speed and accuracy in decoding and word recognition to be able to read connected text easily enough to allow focus to shift from identifying individual words to comprehending the meaning of words, phrases, sentences, and eventually passages. As students are able to read text faster and with greater ease, they begin to read orally with speed and expression similar to their speech.

#### **3.2 Purpose of EGRA**

Prior to EGRA development and use, there had been very little information about student learning in the early grades in low-income countries. EGRA was developed to provide a way to measure a child’s initial reading skills. More specifically, EGRA was constructed to assess the reading and language skills identified to be critical for students to become fluent readers who comprehend what they read. By assessing student’s knowledge of the alphabetic principle, decoding skills, oral reading fluency (ORF), and comprehension of written text and oral language, EGRA may inform Ministries of Education, donors, teachers, and parents about students’ reading skills in the early grades. Because of EGRA’s direct links with the skills critical for successful reading achievement, the assessment may assist education systems in setting standards and planning curricula to best meet children’s needs in learning to read.

#### **3.3 What EGRA Measures**

The EGRA instrument consists of a variety of subtasks designed to assess foundational reading skills crucial to becoming a fluent reader. EGRA is designed to be a method-independent approach to assessment (i.e., the instrument does not reflect a particular method of reading instruction). Instead, EGRA measures the basic skills that a child must possess to eventually be able to read fluently and with comprehension—the ultimate goal of reading. EGRA subtasks are based on research regarding a comprehensive approach to reading acquisition across languages. These skills are phonological awareness, decoding, reading fluency, reading comprehension, and listening comprehension.



### **3.4 Limits of EGRA**

EGRA is not intended to be a high-stakes accountability measure to determine student grade promotion or to evaluate individual teachers. EGRA is designed to complement, rather than replace, existing curriculum-based pencil-and-paper assessments. EGRA is made up of a set of subtasks that measure foundational skills that have been found to be predictive of later reading success. However, due to the constraints imposed by children's limited attention span and stamina, neither EGRA nor any other single instrument is capable of measuring all skills required for students to read with comprehension. EGRA is not intended to be an instructional program, but rather is capable of informing instructional programs. EGRA cannot fully determine background or literacy behaviors that could impact a student's ability to read.

## 4 Development of the 2015 EGRA tool

Development of EGRA in Ethiopia began in 2006 when USAID, through the EdData II project, contracted RTI to develop an instrument for assessing early grade reading. In 2007, RTI piloted the draft instrument in several languages and countries. In 2010, the EGRA instrument was adapted for use in Ethiopia in six local languages in grades 2 and 3: Afan Oromo, Af-Somali, Amharic, Tigrigna, Sidamo Affoo, and Harari (approximately 90% of the population speaks at least one of these languages).

The first national level EGRA was conducted in Ethiopia through a joint effort of RTI and USAID /IQPEP project. The IQPEP project conducted two additional large scale EGRAs in five local languages during 2012 and 2014. In 2014 READ TA carried out baseline EGRA in additional two languages namely Wolayittatto, and Hadiyissa. During January 30-31, 2015 RTI and USAID conducted a benchmarking exercise for the 2014 EGRA in seven mother tongue languages. USAID hired an international consultant, Joe DeStephano of RTI for the purpose and convened a national workshop where all MoE, RSEBs, and Implementing partners came together in Adama. Amy Todd from AIR home office and three technical staff from the READ M&E field office were in attendance.

To revise the tool, AIR hired local mother tongue language experts who had worked on the development of the 2010 EGRA tool. These experts each constructed two EGRA tools: one tool A) had completely new items and the other tool B) had between 20-30% new items depending on the language. In November, to create the EGRA 2015, READ M&E brought together representatives from the Ministry of Education (MoE), Regional State Education Bureaus (RSEB), National Examination Assessment and Evaluation Agency (NEAEA), USAID, and 2-3 local curriculum and assessment specialists per language. At this workshop, it was decided not to engage with the completely new tool (Tool A) but to focus solely on the revised version (Tool B).

Each item of the EGRA (Tool B) was reviewed by language experts, MoE, and USAID until consensus could be reached on each language's tool. The workshop compared items with the 2010 baseline and checked for compatibility with the newly developed mother tongue curriculum.

The workshop took into consideration that alpha syllabaries, such as Amharic and Tigrigna, are written with symbols called fidels, which are represented as syllables (consonant and vowel), and not at the phoneme level as in alphabetic languages such as English. However, there is direct fidel-sound correspondence and children must learn the fidels and their corresponding sounds to learn to read. Thus, it is important that the EGRA in Ethiopian languages test for phonemic awareness as well as syllabic awareness. Thus, the revised EGRA 2015 measures: phonemic awareness, syllabic awareness, letter sound fluency, word naming fluency, unfamiliar word naming fluency, oral reading fluency, reading comprehension, and listening comprehension.

During the revision process the level of difficulty of each letter in both tools, position and distribution of the letters and words within the test in both tools, the nature of comprehension passage in terms of number of words and structure between both tools were systematically reviewed.

## 5 Design and Methodology

### 5.1 The Sample

The 2015 Ethiopian min-EGRA assessed Grade 2 and Grade 3 students in three mother tongue languages (Aff Somali, Sidamu Affo and Tigrigna). Data was collected in November, 2015. Table 1 shows the sampling distribution by language and grade. It was planned to collect data from a total of 1800 students and the achieved sample size was 1719 from 45 schools. In Aff Somali, there were not enough number of students on the days of data collection due to scarcity of water in area.

**Table 1. Sample Distribution by Grade and Language**

<b>Language</b>	<b>Schools</b>	<b>Grade 2 students</b>	<b>Grade 3 students</b>	<b>Total</b>
Aff Somali	15	268	251	519
Sidamu Affo	15	300	300	600
Tigrigna	15	300	300	600
<b>Total</b>	<b>45</b>	<b>868</b>	<b>851</b>	<b>1719</b>

### 5.2 Data Collection

#### 5.2.1 Data Collection Training

Data collection training was conducted from November 16 – 20, 2016. EGRA assessors were selected from Colleges of Teacher Education (CTEs), Universities, the MoE, Regional State Education Bureaus (RSEBs), National Education Assessment and Examinations Agency (NEAEA), Zone Education Departments (ZEDs) and preparatory schools. The minimum qualification was a M.Ed. or MA. Assessors received training on how to administer each subtask of the EGRA, how to interact with students during the assessment as well as school staff, and how to properly implement the protocol for visiting the sampled schools and randomly selecting the students to be assessed. Assessors learned how to operate handheld electronic tablet devices loaded with the Tangerine® software. This software, designed for education survey data collection, allowed assessors to gather all the data on the tablets rather than on paper, thus streamlining the data collection and data cleaning process. Assessors were trained in how to administer the tests as well as how to save and upload the finished assessments.

#### 5.2.2 Data collection methodology

The Mini EGRA was administered in three languages in three regions of Ethiopia; Aff Somali, Sidamu Affoo and Tigrigna from 21 November to 01 December 2015. The test administration required test administrators who were deployed in teams of four people to collect data in sixty-five sample schools in five regions in seven languages (15 schools each from Tigrigna, Aff- Somali and Sidama-Afu languages and 5 schools each from Amharic, Afan Oromo, Hadiyissa and Wolaytaato languages).

Data collectors were provided with the required materials; the tangerine tablet, pupil stimuli, sampling sheets, school visit summary sheets, EGRA protocol, pencils, envelopes, clipboards, and folders and a vehicle was assigned for each team. Most of the visited schools' managements were very cooperative and, where possible, empty classrooms were made available for the assessment. At the end of each school visit teams met and discuss issues that were noted during the data collection with their team leaders. Issues noted during the meetings were addressed immediately. The daily review meetings also provided a forum

for sharing the day's experiences. This process proved very useful by providing the enumerators with opportunities to learn from each other and correct any data collection misunderstandings they may have had. READ M&E staff managed to visit each data collection team at their assigned school.

The selection of schools was not predetermined as there is no approved list of schools available in Ethiopia. Instead, data collector supervisors consulted with regions and zones to select the woredas. The woredas were selected in a random lottery method. From the woreda offices, the data collector supervisor randomly selected using the lottery method the needed number of schools for each woreda.

**5.3 The assessors systematically selected students from each selected sample school. In each of the sampled schools, 20 pupils of Grade 2 and another 20 of Grade 3 will be selected to participate in the assessment. Where there are 20 or fewer children in a given class (which is highly unlikely in lower primary), all children in that class will be assessed. In each of these classes, an equal number of girls and boys were selected. Instruments**

The data collecting instruments were reading tests meant to measure basic literacy skills targeted at Grade 2 level. Grade 3 students take the same EGRA to measure growth by grade level. The tasks measuring basic literacy skills in each case were incremental in their complexity. Each task was presented to the child on a one-to-one basis. EGRA is an orally administered assessment targeted at measuring the pre-reading and reading skills foundational to later reading and academic success. EGRA takes approximately 15 minutes to administer and is often combined with a questionnaire measuring a variety of student background variables to assist in explaining some of the reading outcome findings. A brief description of the subtasks is given below:

**Letter-naming fluency:** ability to read the letters of the alphabet without hesitation and naturally. This is a timed test that assesses automaticity and fluency of letter names. It is timed to 1 minute, which saves time and also prevents children having to spend time on something that is difficult for them. It is timed to 60 seconds and is discontinued if none of the letters in the first (i.e., 10 letters) is read correctly.

**Phonemic awareness:** awareness of how sounds work with words. This is generally considered a pre-reading skill, and can be assessed in a variety of ways. In some Ethiopian languages, this task might be designed to determine whether children could differentiate the first syllable (or fidel) in a word, or whether they could identify all of the fidels in a word.

**Initial sound identification:** Measures the ability to discriminate beginning sounds. Three words are presented and the aim is to identify the word that begins with a different sound from the other two. It is oral and has 10 sets of words. It is discontinued if no points are earned in the first five items.

**Segmentation (phoneme or syllables):** Measures the ability to segment a word into individual phonemes or syllables. This subtask is oral and has 10 items. It is discontinued if no points are earned in the first five items.

**Familiar word fluency:** ability to read high-frequency words. This assesses whether children can process words quickly. The lists of words were derived from the most frequently used words in Grade 2 textbooks. It is timed to one minute. If the student could not read any of the first five words, the test was discontinued.

**Invented or none-sense word fluency:** ability to process words that could exist in the language in question, but do not. The words were derived from the list of familiar words and follow the common patterns of the language. This component assesses a child's ability to

“decode” words fluently. It is timed to 1 minute. If the student could not read any of the first five invented words, the test was discontinued.

**Connected text oral reading fluency:** Measures the ability to read a grade-level passage of approximately 60 words. It is scored for accuracy and rate. It is timed to 60 seconds and is discontinued if none of the words in the first line (i.e., about 10 words) is read correctly.

**Comprehension in connected text:** ability to answer comprehension questions based on the passage read. There were five questions. It is untimed and does not have a discontinuation rule.

**Listening comprehension:** being able to follow and understand a simple oral story. This assesses a child’s ability to concentrate and focus to understand a very simple story, assessed by asking simple non-inferential (factual) questions. It is untimed and does not have a discontinuation rule.

## 6 Findings

### 6.1 Discontinuity Rate

For the timed EGRA components (letter name knowledge, familiar words reading, invented words reading and passage words reading), an auto stop rule was implemented to discontinue the test if students could not correctly respond to a certain number of items within a specified time limit. This rule was established to relieve stress among students and to ensure that students who did not understand a task would not be tested on it. The auto-stop rule is not applied to the untimed tasks namely phonemic awareness, reading comprehension and listening comprehension.

#### 6.1.1 Discontinuity Rate in Aff Somali

Figure 1 below shows the percentages of students that discontinued each timed component. The largest proportions of early stop cases appeared in the invented words (65.1%) followed by familiar words (61%). The sub task that showed the lowest percentage of early stop cases was letter name knowledge (22.9%). In all case the proportions of discontinuity were found larger in Grade 2.

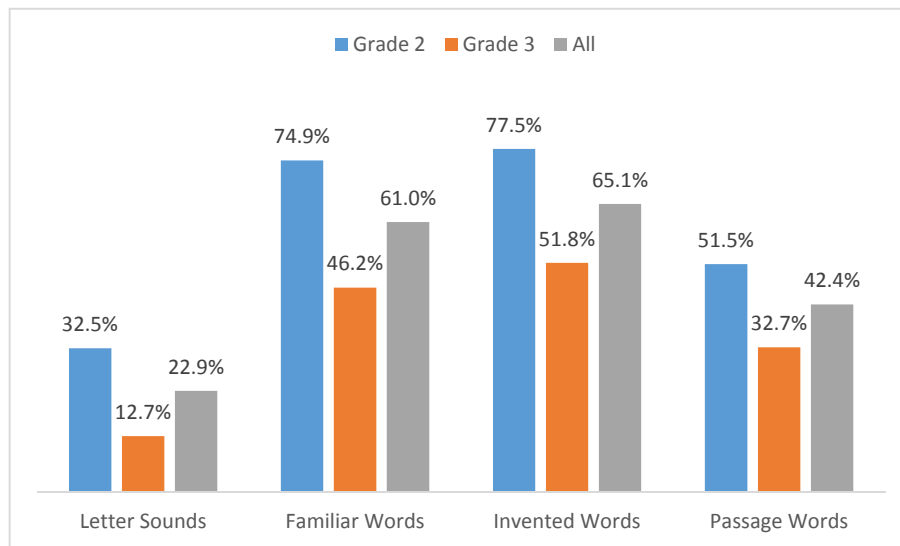
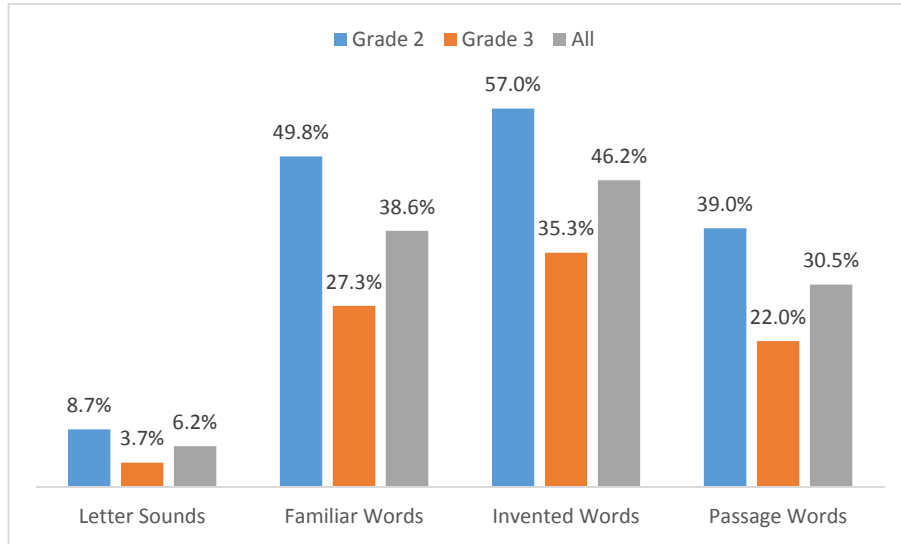


Figure 1. Timed Subtasks Proportion of Discontinuity in Aff Somali by Grade

#### 6.1.2 Discontinuity Rate in Sidamu Affo

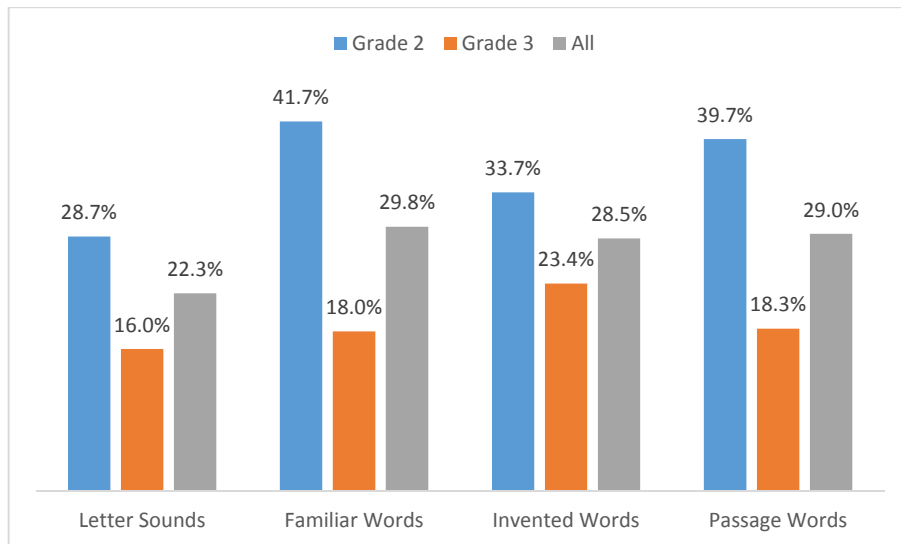
Figure 2 shows the percentages of students that discontinued each timed component. The largest proportions of early stop cases appeared in the invented words (46.2%) followed by familiar words (49.8%). The subtask that showed the lowest percentage of early stop cases was letter name knowledge (6.2%). In all cases the proportions of discontinuity were found larger in Grade 2. This was found similar with that of Aff Somali.



**Figure 2. Timed Subtasks Proportion of Discontinuity in Sidamu Affo by Grade**

### 6.1.3 Discontinuity Rate in Tigrigna

Figure 3 shows the percentages of students that discontinued each timed component. The proportions of early stop cases in familiar words reading, invented words and oral reading passage were more or less equal. In each of the three subtasks about a quarter of students were found unable to read a single word. In letter name knowledge too, a large proportion of students were found zero readers (22.3%). Similar with the other two languages Grade 2 students were found with higher proportion of discontinuity in all cases.



**Figure 3. Timed Subtasks Proportion of Discontinuity in Tigrigna by Grade**

## 6.2 Proportions Meeting Benchmarks

This section presents the proportions of students for each of the three languages by grade who can be categorized as non-readers (NR), reading with limited comprehension (RWLC), reading with comprehension (RWC), and reading fluently with comprehension (RWC). These categories represent a combination of the oral reading and reading

comprehension subtasks. READ M&E uses the benchmarks validated in the 2015 January workshop held by USAID and the MoE, facilitated by RTI. Students who scored zero on the oral reading fluency (ORF) portion of the EGRA were categorized as “non-readers”; students who scored more than zero on the ORF portion, but less than 60% on the reading comprehension were categorized as “reading with limited comprehension”; students who scored between 60 and 80% on the reading comprehension subtask were categorized as “reading with comprehension”; and students with a reading comprehension score above 80% were categorized as “reading fluently with comprehension.”

### 6.2.1 Proportions Meeting Benchmarks in Aff Somali by Grade

Figure 4 shows that the proportion of students at the ‘nonreader’ and ‘reading with limited comprehension’ covers more than 90%. Only very few students achieved the highest levels. Looking at grade levels the proportion of ‘nonreaders’ has decreased while the other levels increased in Grade 3. The proportion of reading with fluency and comprehension increased by 8%.

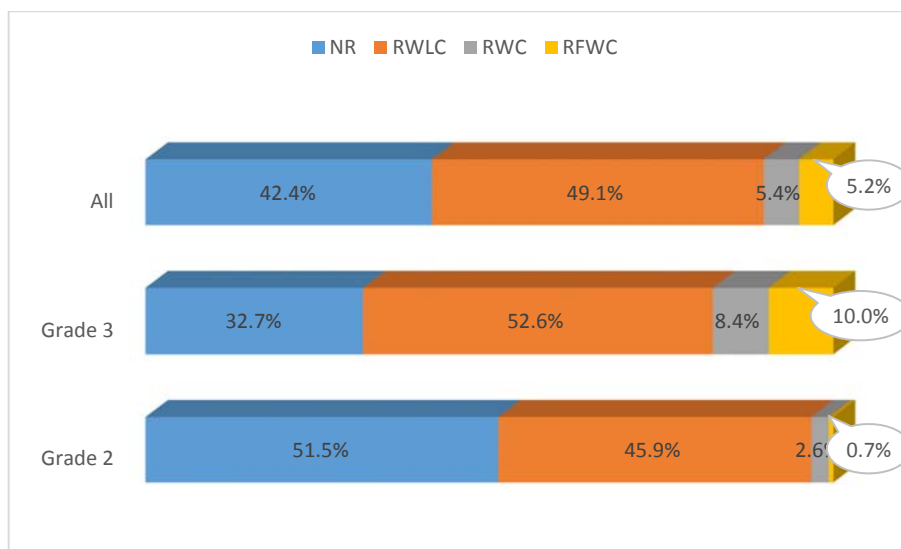
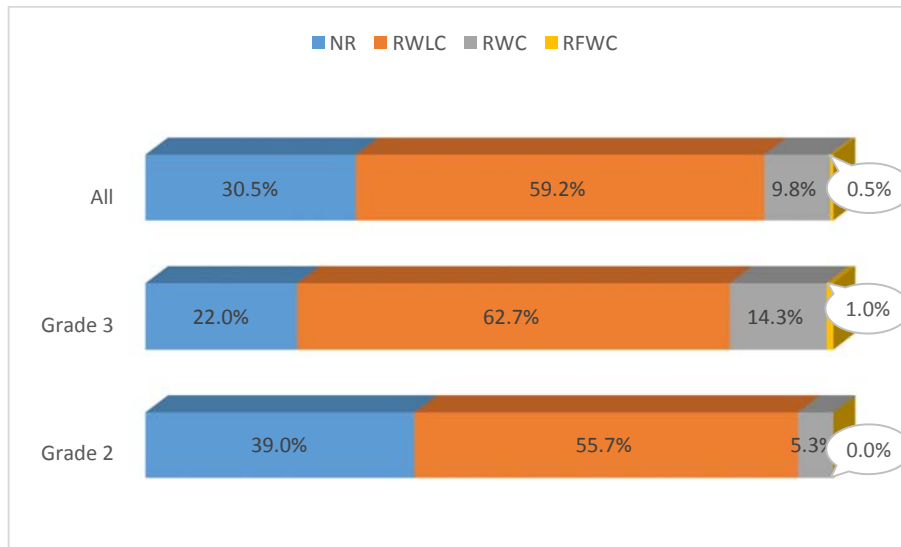


Figure 4. Proportion at Various Levels of Reading Proficiency in Aff Somali by Grade

### 6.2.2 Proportions Meeting Benchmarks in Sidamu Affo by Grade

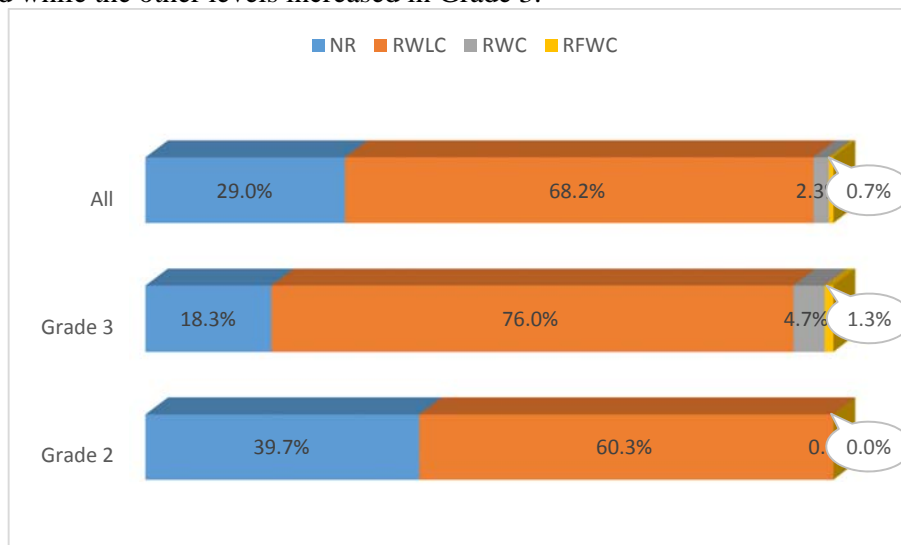
The proportion of students at the ‘nonreader’ and ‘reading with limited comprehension’ covers more than 90%. Proportions who achieved the highest levels of benchmark were found negligible. Looking at grade levels the proportion of ‘nonreaders’ has decreased while the other levels increased in Grade 3.





**Figure 5. Proportion at Various Levels of Reading Proficiency in Sidamu Affo by Grade**

Figure 6 shows that more than 98% of students fall in ‘nonreader’ and ‘reading with limited comprehension’ categories. In Grade 2, not a single student achieved the highest levels of benchmark. In the same way with the other two languages the proportion of ‘nonreaders’ has decreased while the other levels increased in Grade 3.



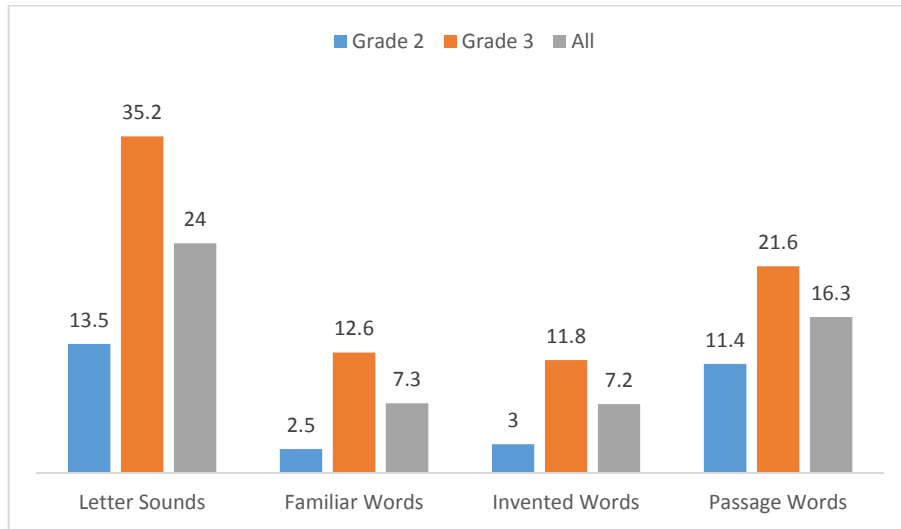
**Figure 6. Proportion at Various Levels of Reading Proficiency in Tigrigna by Grade**

### 6.3 Mean Fluency Scores

This section presents the mean fluency scores of the timed tasks by grade for each of the three languages. Independent samples t-tests were carried out taking each task as dependent variable (DV) and grade as independent variable (IV) for each language. In all the three languages, the mean differences between the two grade levels in the fluency tasks were found statistically significant in favor of Grade 3. *Cohenn’s d* was also computed to see the effect size and in Aff Somali the effects were found small in oral reading fluency, moderate in familiar and invented words and large in letter sound (Annex I shows details on the statistics).

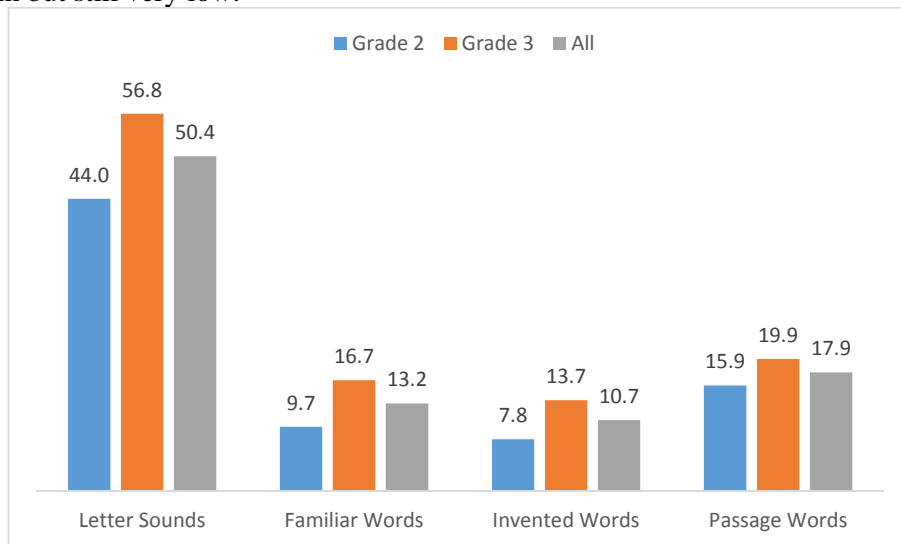
### 6.3.1 Mean Fluency Scores in Aff Somali by Grade

Figure 7 shows the mean fluency scores of the timed tasks in Aff Somali by grade. On the average students were able to read very few correct familiar words per minute or correct invented words per minute only about seven in each case. In all cases Grade 3 students were able to read better than Grade 2 indicating grade gain. Independent samples t-tests were carried out to detect the presence of statistically significant differences between grades. The mean differences range from 8.7 in invented words to 21.7 in letter sound.



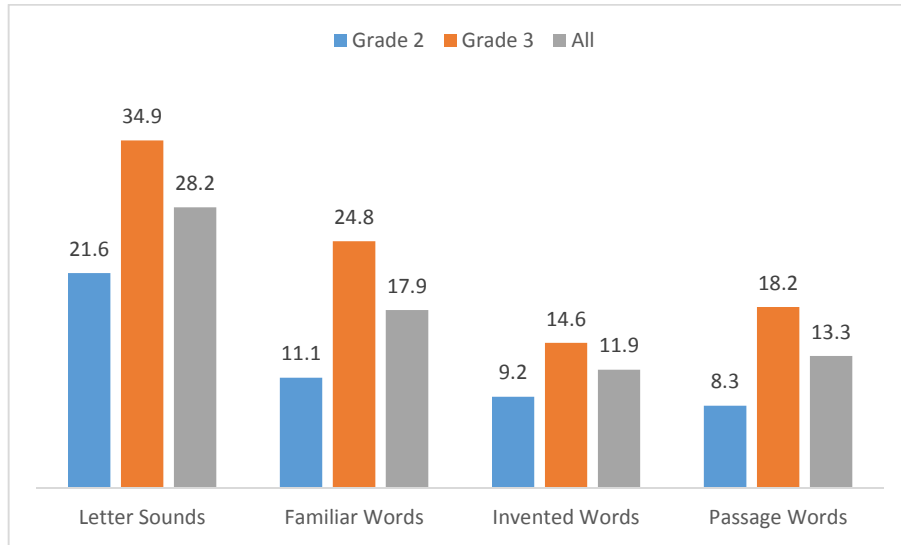
**Figure 7. Mean Fluency Scores in Aff Somali by Grade**

Figure 8 shows the mean fluency scores of the timed tasks in Sidamu Affo by grade. The least mean score was observed in invented words (10.7) while the highest in letter sound (50.4). In each subtask Grade 3 students were able to read better than Grade 2 indicating grade gain but still very low.



**Figure 8. Mean Fluency Scores in Sidamu Affo by Grade**

Figure 9 shows the mean fluency scores of the timed tasks in Tigrigna by grade. The least mean score was observed in invented words (12) while the highest in letter name (28.2). In each subtask Grade 3 students were able to read better than Grade 2 indicating grade gain but still very low.



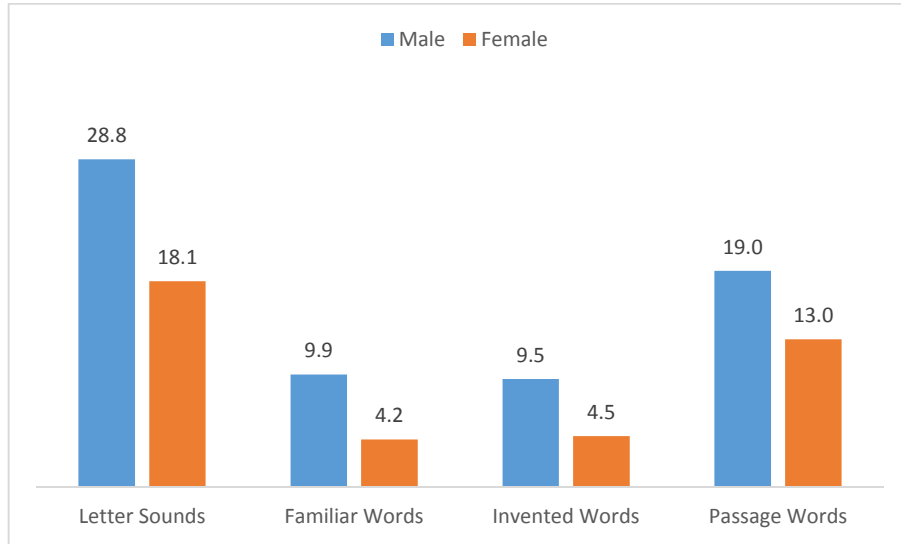
**Figure 9. Mean Fluency Scores in Tigrigna by Grade**

#### **6.4 Mean Fluency Scores by Gender**

Independent samples t-tests were carried out taking each subtask as dependent variable (DV) and gender as independent variable (IV) for each language. In Aff Somali, boys performed better than girls and the mean differences were statistically significant except in oral reading fluency. However the effect sizes were found small in all cases. In the other two languages girls performed better than boys but the mean differences were not found statistically significant (Annex III shows detail statistics).

##### **6.4.1 Mean Fluency Scores in Aff Somali by Gender**

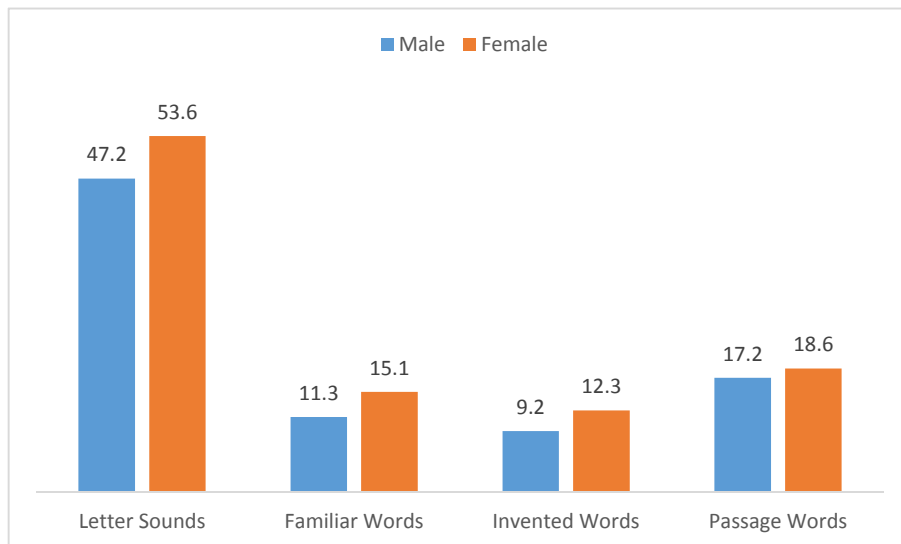
Figure 10 shows the mean fluency scores by gender in Aff Somali. In all the subtasks boys performed better than girls. In passage reading boys were able to read additional six words per minute.



**Figure 10. Mean Fluency Scores in Aff Somali by Gender**

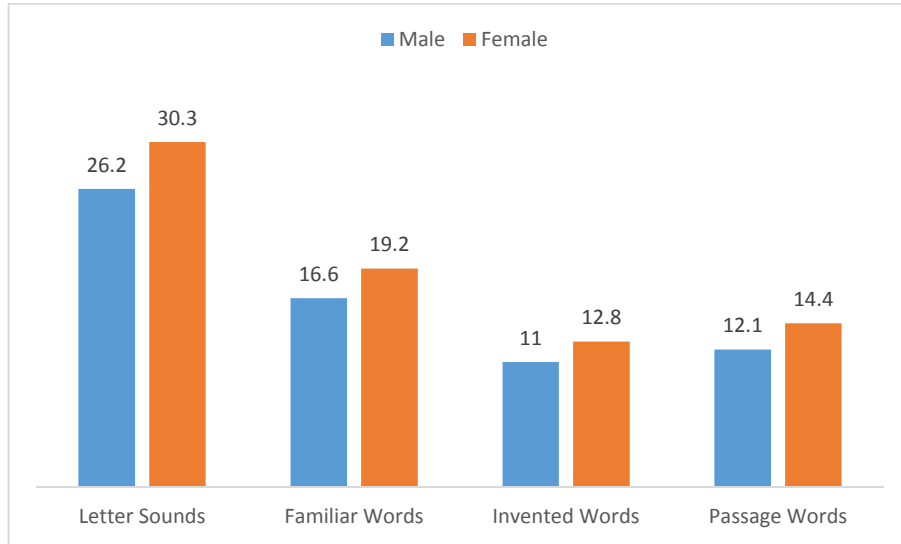
#### 6.4.2 Mean Fluency Scores in Sidamu Affo by Gender

Figure 11 shows the mean fluency scores in Sidamu Affo by gender. In all the subtasks girls performed slightly better than boys. The widest gap was observed in letter sounds where girls were able to identify additional six letters.



**Figure 11. Mean Fluency Scores in Sidamu Affo by Gender**

Figure 12 shows the mean fluency scores in Tigrigna by gender. In all the subtasks girls performed slightly better than boys. The widest gap was observed in letter sounds where girls were able to identify additional four letters.



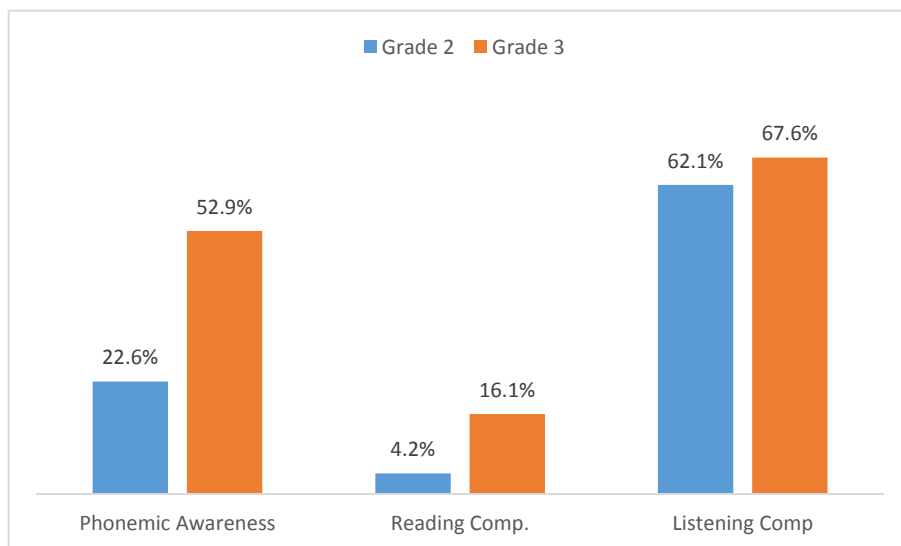
**Figure 12. Mean Fluency Scores in Tigrigna by Gender**

### 6.5 Mean Scores of Untimed Tasks by Language and Grade

This section shows the mean scores of the untimed tasks disaggregated by grade for each of the three languages. The untimed subtasks are: Phonemic Awareness (PA), Reading Comprehension (RC) and Listening Comprehension (LC). In all the three languages, the mean differences between the two grade levels were found statistically significant in favor of Grade 3. The effect sizes were found small in listening comprehension, moderate in reading comprehension and large in phonemic awareness (Annex II shows detail statistics).

#### 6.5.1 Mean Scores of Untimed Tasks in Aff Somali by Grade

Figure 13 shows that in all the three subtasks Grade 3 students performed better than Grade 2 and the widest gap was observed in phonemic awareness (29%). The mean difference in listening comprehension was very small indicating no major grade gain while the grade gain in reading comprehension was relatively high (12%).



**Figure 13. Mean Scores of the Untimed Tasks in Aff Somali by Grade**

### 6.5.2 Mean Scores of Untimed Tasks in Sidamu Affo by Grade

In all the three subtasks Grade 3 students performed slightly better than Grade 2 and the widest gap was observed in reading comprehension (Figure 14).

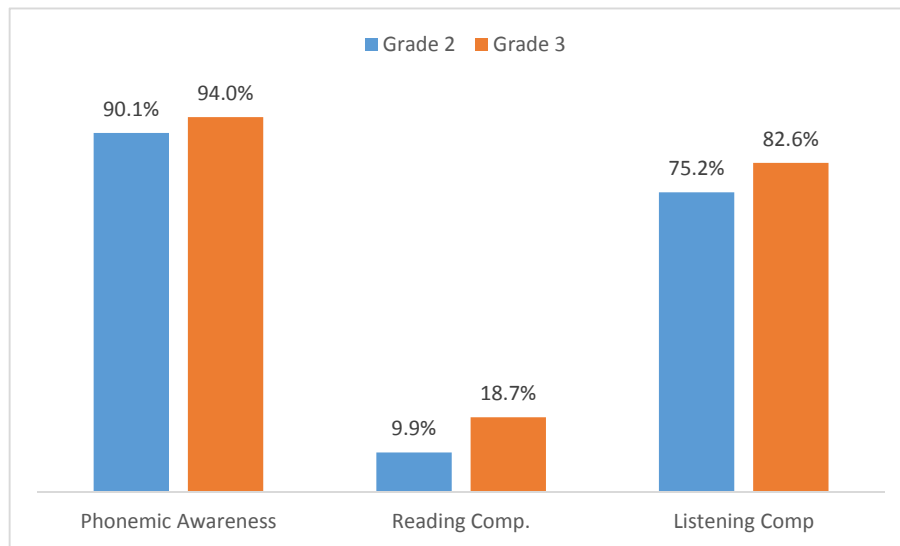


Figure 14. Mean Scores of the Untimed Tasks in Sidamu Affo by Grade

### 6.5.3 Mean Scores of Untimed Tasks in Tigrigna by Grade

Figure 15 shows in all the three subtasks Grade 3 students performed better than Grade 2 and the widest gap was observed in phonemic awareness (16.6%).

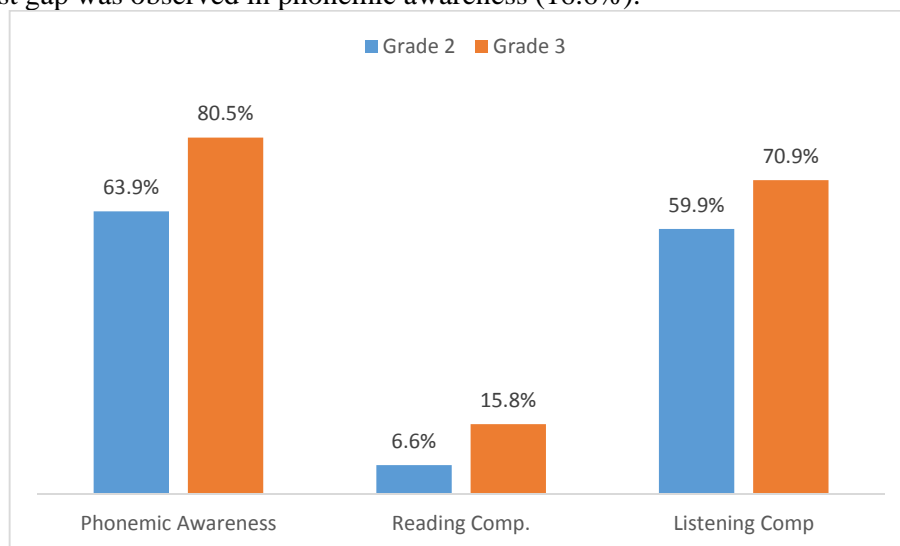


Figure 15. Mean Scores of the Untimed Tasks in Tigrigna by Grade

## 6.6 Mean Scores of Untimed Tasks by Language and Gender

This section shows the mean scores of the untimed tasks disaggregated by gender for each of the three languages. The untimed subtasks are: Phonemic Awareness (PA), Reading Comprehension (RC) and Listening Comprehension (LC). In Aff Somali, boys performed better than girls and the mean differences were statistically significant. However, the effect sizes were small in all cases. In the other two languages, the findings are mixed but the mean differences were not statistically significant and the effect sizes were very small in all cases (Annex IV).

### 6.6.1 Mean Scores of Untimed Tasks in Aff Somali by Gender

Figure 16 shows that in all the three subtasks boys performed better than girls. The widest gap was observed in phonemic awareness (16.2%) followed by reading comprehension (7%).

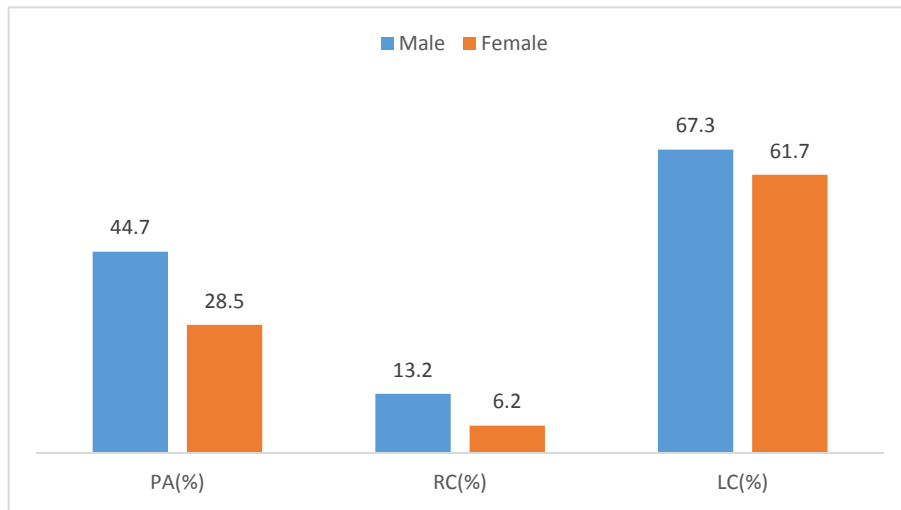
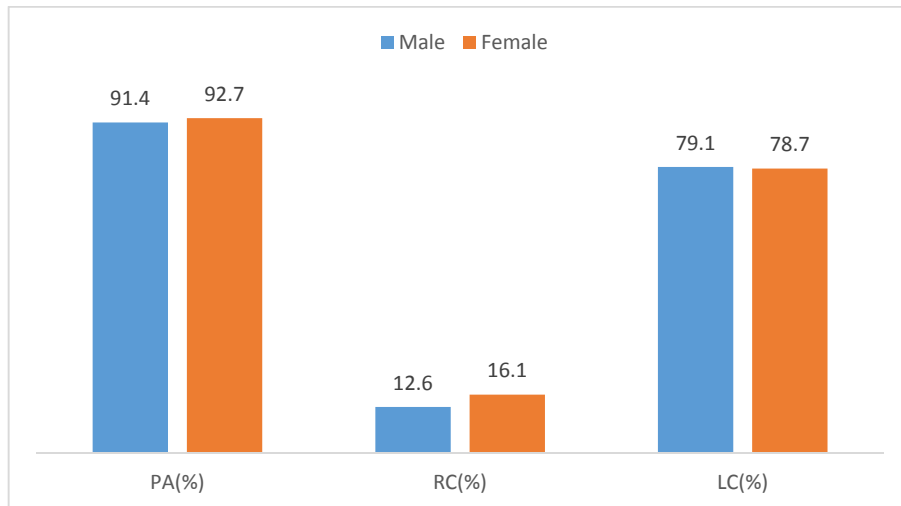


Figure 16. Mean Scores of the Untimed Tasks in Aff Somali by Gender

### 6.6.2 Mean Scores of Untimed Tasks in Sidamu Affo by Gender

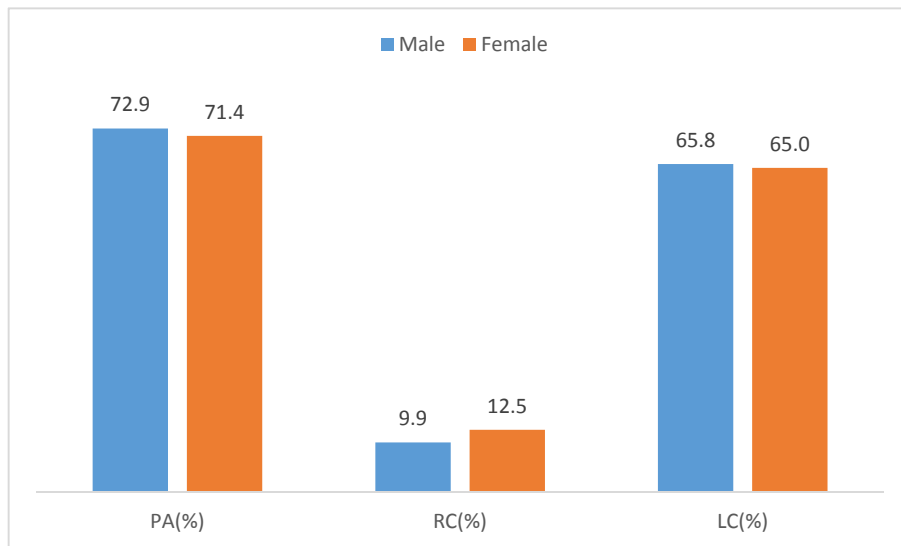
Figure 17 shows that in reading comprehension girls performed slightly better than boys but in the other two subtasks no major differences are observed.



**Figure 17. Mean Scores of the Untimed Tasks in Sidamu Affo by Gender**

### 6.6.3 Mean Scores of Untimed Tasks in Tigrigna by Grade

Figure 18 shows that in reading comprehension girls performed slightly better than boys while in the other two subtasks boys performed slightly better than girls but the mean differences are negligible.



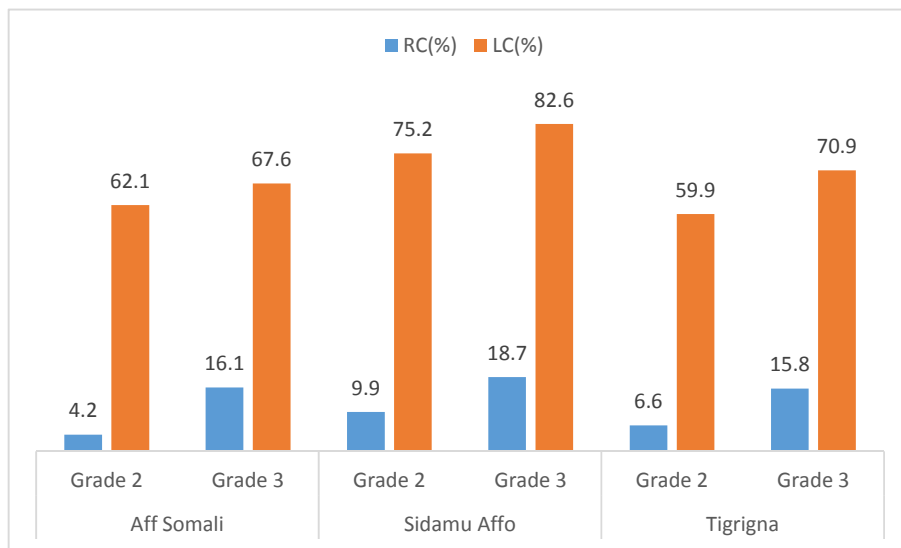
**Figure 18. Mean Scores of the Untimed Tasks in Tigrigna by Gender**

### 6.6.4 Comparisons between RC (%) and LC (%) Scores

Figure 16 shows comparison between the Reading Comprehension and Listening Comprehension scores by language and grade. In all the three languages there exist very wide gap between the two in favor of listening comprehension. The goal of reading is to understand what has been read. The Oral Reading and Comprehension subtasks require a student to read a passage of connected text aloud, and then to answer questions about the portion of the passage that the student had read within a minute. However, about 70% of the students obtained a zero score on the Reading Comprehension subtask, reflecting the fact that



nearly seven out of ten students were not able to answer even one comprehension question correctly. The accepted benchmark for comprehension is 80 percent or more correct on the Reading Comprehension Task. Average student scores fell far below this threshold.



**Figure 19. Reading (RC) and Listening Comprehension (LC) Scores by Language**

This finding suggests that although some students were able to respond correctly to the Reading Comprehension subtask questions, they were very few in number. There is still a high percentage of students who cannot read enough words in 1 minute to develop an understanding of what they read. To obtain even the minimal score on the Reading Comprehension subtask, the students must have read at least one complete sentence on the Oral Reading subtask.

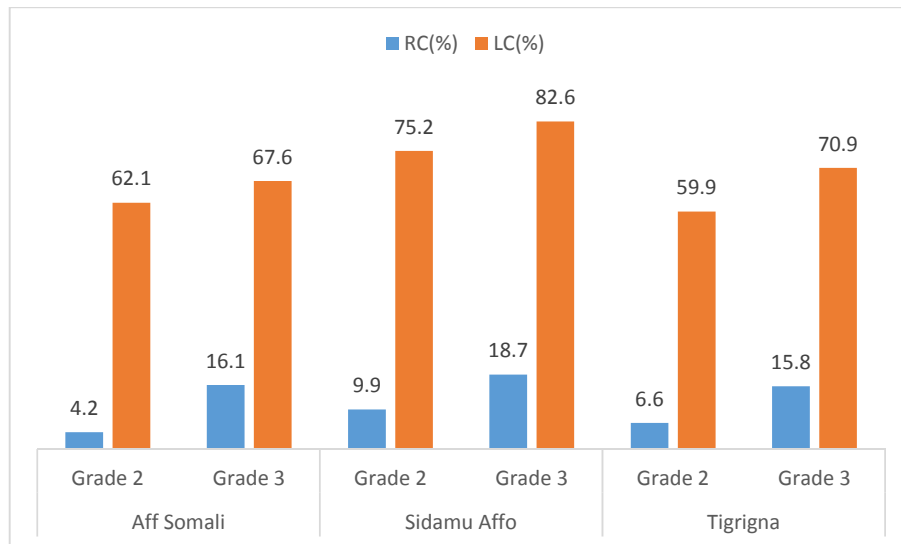
Students appear to understand their mother tongue as reflected by relatively higher scores on the Listening Comprehension subtask. The Listening Comprehension subtask reflects an individual’s listening skills, but it also requires memory skills, including short-term memory and working memory. Therefore, this subtask requires more than listening and language comprehension capacity. Although the Listening Comprehension subtask does not relate strongly to other EGRA tasks, listening skills are considered to be a key element of a well-designed reading instruction program.

## 7 Conclusion

The mini EGRA sampled fifteen schools in each of the three regions and as such cannot be considered representative of all schools in the region. Thus no formal conclusions can be made from this small data set. Nor can there be any cross-language discussion as the EGRA does not allow for conclusions between languages. What can be discussed is the data points that somehow stood apart from the others and as such, raise more questions.

In much of the world, girls tend to perform better than boys in this grade level on reading assessment. In Sidamu Affo and Tigrigna, this was the case. However in Aff Somali, in all the subtasks boys performed better than girls. In passage reading boys were able to read an additional six words per minute. It is not clear why this reversal of trends was seen. In order to be sure that this is indeed the circumstance, READ M&E will need to wait until the 2016 May EGRA results with a much larger sample size. Reasons for the reversal of international norms could be cultural in that the girls may not be accustomed to being in the company of a solo male assessor.

Ethiopia is largely an oral culture, especially away from the cities. Thus it is interesting that children in this mini EGRA did not score higher on the oral listening comprehension assessment. This is not to say that children did not do significantly better on the oral comprehension assessment than on the reading assessment. However, it is interesting to note that in an oral culture, children did not score higher on the listening comprehension assessment. The chart is repeated here for convenience.



**Figure 20. Reading (RC) and Listening Comprehension (LC) Scores by Language**

One possible explanation may be that although the culture is largely oral, children are not asked comprehension questions such as those found in the EGRA. School- type questions, such as inferential questions, may not be asked of children in or out of school. With a larger sample size, READ M&E can research which types of questions children were and were not able to answer correctly.

While overall scores on all reading tasks were rather low, it is useful to remember that there are few demands for literacy in many rural settings and relatively few even in urban areas. Improving national reading score is a generation goal that cannot be accomplished through improved curriculum, new reading materials, and school based learning alone but must be thought of as a process that increases from one generation to the next.

Relatively speaking Grade 3 students performed better than Grade 2 ones indicating grade gain. However even Grade 3 students are far below what is expected at Grade 2.



## Annex I: Independent Samples T-Test and Cohen's D for fluency tasks by grade

Language = Aff Somali

Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Three	251	35.2	29.04	10.4	517	0.000	21.7	0.9
	Two	268	13.5	17.28					
CFWPM	Three	251	12.6	19.43	8.0	517	0.000	10.1	0.7
	Two	268	2.5	6.59					
CIWPM	Three	251	11.8	17.27	7.5	517	0.000	8.7	0.7
	Two	268	3.0	7.63					
CPWPM	Three	251	21.6	41.23	3.1	517	0.002	10.3	0.3
	Two	268	11.4	33.67					

Language = Sidamu Affo

Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Three	300	56.8	27.99	4.95	598.00	0.000	12.8	0.4
	Two	300	44.0	35.11					
CFWPM	Three	300	16.7	15.20	6.08	598.00	0.000	7.0	0.5
	Two	300	9.7	12.79					
CIWPM	Three	300	13.7	13.97	5.64	598.00	0.000	5.9	0.5
	Two	300	7.8	11.35					
CPWPM	Three	300	19.9	23.18	1.70	598.00	0.089	4.1	0.1
	Two	300	15.9	34.43					

Language = Tigrigna

	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Three	300	34.9	27.78	6.182	598	0.000	13.3	0.5
	Two	300	21.6	24.87					
CFWPM	Three	300	24.8	20.87	9.382	598	0.000	13.7	0.8
	Two	300	11.1	14.31					
CIWPM	Three	300	14.6	12.40	5.968	598	0.000	5.5	0.5
	Two	300	9.2	9.98					
CPWPM	Three	300	18.2	16.20	8.704	598	0.000	9.9	0.4
	Two	300	8.3	11.22					

## Annex II: Independent Samples T-Test and Cohen's D for untimed tasks by grade

Language = Aff Somali

Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
PA(%)	Three	251	53.3	40.18	9.599	517	0.000	31.0	0.8
	Two	268	22.4	33.15					
RC(%)	Three	251	16.5	30.63	6.104	517	0.000	12.5	0.5
	Two	268	4.0	13.44					
LC(%)	Three	251	67.7	24.38	2.491	517	0.013	5.7	0.2
	Two	268	62.0	27.65					

Language = Sidamu Affo

Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
PA(%)	Three	300	94.0	15.77	2.392	598	0.017	3.9	0.2
	Two	300	90.1	23.43					
RC(%)	Three	300	18.9	23.99	5.258	598	0.000	9	0.4
	Two	300	9.9	17.43					
LC(%)	Three	300	82.5	20.57	3.651	598	0.000	7.1	0.3
	Two	300	75.3	26.87					

Language = Tigrigna

Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
PA(%)	Three	299	80.4	32.57	5.594	597	0.000	16.5	0.5
	Two	300	64.0	39.17					
RC(%)	Three	300	15.7	20.89	6.351	598	0.000	9.1	0.5
	Two	300	6.7	13.22					
LC(%)	Three	300	70.9	22.32	5.807	598	0.000	11.1	0.5
	Two	300	59.9	24.32					

### Annex III: Independent Samples T-Test and Cohen's D for fluency tasks by gender

Aff Somali									
Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Male	285	28.8	28.36	4.775	517	0.00	10.8	0.4
	Female	234	18.1	21.58					
CFWPM	Male	285	9.9	17.42	4.325	517	0.00	5.7	0.4
	Female	234	4.2	11.13					
CIWPM	Male	285	9.5	15.76	4.117	517	0.00	5.0	0.4
	Female	234	4.5	10.62					
CPWPM	Male	285	19.0	39.23	1.817	517	0.07	6.1	0.2
	Female	234	13.0	35.87					
Sidamu Affo									
Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Male	299	47.2	34.46	-2.409	598	0.016	-6.3	0.2
	Female	301	53.6	29.86					
CFWPM	Male	299	11.3	12.85	-3.218	598	0.001	-3.8	0.3
	Female	301	15.1	15.70					
CIWPM	Male	299	9.2	11.77	-2.953	598	0.003	-3.1	0.4
	Female	301	12.3	14.06					
CPWPM	Male	299	17.2	30.45	-0.599	598	0.549	-1.4	0.0
	Female	301	18.6	28.34					
Tigrigna									
Task	Grade	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean Difference	Cohen's D
CLSPM	Male	300	26.2	25.50	-1.867	598	0.062	-4.1	0.2
	Female	300	30.3	28.64					
CFWPM	Male	300	16.6	18.34	-1.629	598	0.104	-2.5	0.1
	Female	300	19.2	19.87					
CIWPM	Male	300	11.0	10.79	-1.867	598	0.062	-1.8	0.2
	Female	300	12.8	12.27					
CPWPM	Male	300	12.1	13.62	-1.893	598	0.059	-2.3	0.2
	Female	300	14.4	15.79					

**Annex IV: Independent Samples T-Test and Cohen's D for the untimed tasks by gender**

<b>Aff Somali</b>									
<b>Task</b>	<b>Grade</b>	<b>N</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>t</b>	<b>df</b>	<b>Sig. (2-tailed)</b>	<b>Mean Difference</b>	<b>Cohen's D</b>
PA (%)	Male	285	44.7	41.97	4.707	517	0.000	16.2	0.4
	Female	234	28.5	35.11					
RC (%)	Male	285	13.2	27.27	3.331	517	0.001	7.0	0.3
	Female	234	6.2	19.16					
LC (%)	Male	285	67.3	25.57	2.425	517	0.016	5.6	0.2
	Female	234	61.7	26.80					
<b>Sidamu Affo</b>									
<b>Task</b>	<b>Grade</b>	<b>N</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>t</b>	<b>df</b>	<b>Sig. (2-tailed)</b>	<b>Mean Difference</b>	<b>Cohen's D</b>
PA (%)	Male	299	91.4	21.02	-0.745	598	0.457	-1.2	0.1
	Female	301	92.7	19.05					
RC (%)	Male	299	12.6	19.65	-1.97	598	0.049	-3.4	0.2
	Female	301	16.1	22.96					
LC (%)	Male	299	79.1	24.70	0.233	598	0.816	0.5	0.02
	Female	301	78.7	23.68					
<b>Tigrigna</b>									
<b>Task</b>	<b>Grade</b>	<b>N</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>t</b>	<b>df</b>	<b>Sig. (2-tailed)</b>	<b>Mean Difference</b>	<b>Cohen's D</b>
PA (%)	Male	300	72.9	36.86	0.495	597	0.621	1.5	0.041
	Female	299	71.4	37.04					
RC (%)	Male	300	9.9	16.56	-1.722	598	0.086	-2.5	0.145
	Female	300	12.5	19.37					
LC (%)	Male	300	65.8	23.47	0.409	598	0.683	0.8	0.033
	Female	300	65.0	24.49					