የትምህርት ምዘናና ፈተናዎች አገልግሎት

Educational Assessment And Examinations Service

EDUCATIONAL ASSESSMENT AND EXAMINATIONS SERVICE

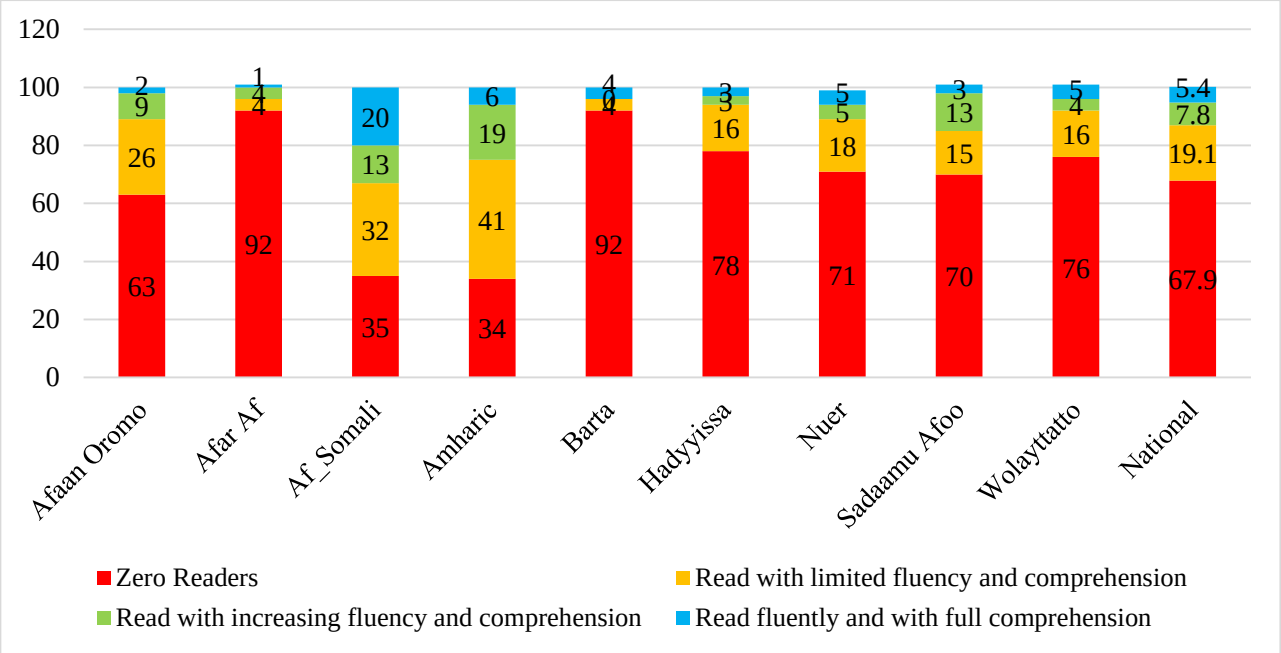
EARLY GRADE READING ASSESSMENT (EGRA 2021) REPORT



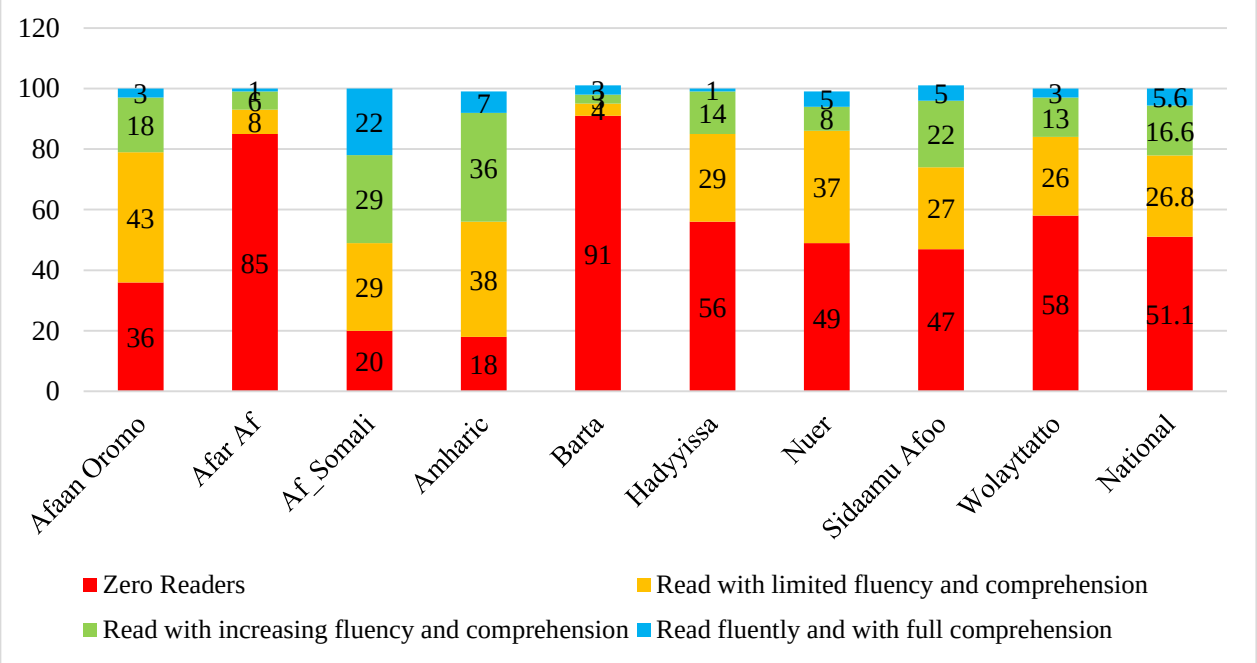


APRIL,2022

Percentage of students at benchmark levels by language



Percentage of grade 2 students at benchmark levels by language



Percentage of grade 3 students at benchmark levels by language

A Study carried out by

Educational Assessment and Examinations Service (EAES)

Technical Working Group:

Effa Gurmu (Dr.): Educational Assessment Directorate Director

Bekele Geleta: Educational Assessment Expert

Solomon Teferi: Educational Assessment Expert

Yeshiwork Simegn: Educational Assessment Expert

Aregawi Gidey: Educational Assessment Expert

Asefa Leta: Educational Assessment Expert

Abiy Kefyalew: Educational Assessment Expert

Educational Assessment and Examinations Service ©

April, 2022

Addis Ababa

Table of Contents

Tables iv

Figures v

Acronyms and Abbreviations vi

Acknowledgements vii

Executive summary viii

1. Introduction 1

1.1 Overview of Early Grade Reading Assessment (EGRA) 1

1.2 EGRA Subtasks 2

1.3 Rationale 4

1.4 Objectives 5

1.5 Basic Research Questions 6

1.6 Limitation 6

2. Method of the Study 7

2.1 Research Design 7

2.2 Particpants and Sampling Techniques 7

2.3 EGRA Instrument Development 9

2.4 Main EGRA Administration Procedures 11

2.5 Data Analysis 12

3.1. Overall Mean Scores of Timed and Untimed Subtasks 13

3.1.1 Students' performance on timed subtasks by language and grade level 13

3.1.2 Students' performance on untimed subtasks by language and grade level 14

3.2 Zero Readers on Each Subtasks 15

3.3 The mean differences of students' performance by subgroups 17

3.3.1 The significant mean differences in timed subtasks by grade level 17

3.3.2 The mean differences of students' performance by gender in timed subtasks 19

3.3.4 The mean differences of students' performance by location in timed subtasks 21

3.3.6 The mean differences of students' performance by gender in untimed subtasks 24

3.3.7 The mean differences of students' performance by location in untimed subtasks 26

3.4 Overall performance of students' by age group 27

3.5 Students' performance by program type 29

3.6 Overall percentage of students at benchmark levels by language 32
3.7 Comparison across years in attaining benchmark levels 36
3.8. The trend of overall performance in each languages by year across subtasks 42
3.9 Contextual Questionnaires Analysis 44
3.9.1 Participants Background Characteristics 44
3.9.2 Principals and Teachers Training 51
3.9.3 School resources / instructional materials 54
3.9.4 Parental Support in Mother Tongue 58
3.9.5 Factors associated with students' oral reading fluency 60
4. Conclusions and Recommendations 67
4.1 Conclusions 67
4.2 Recommendations 71
Appendices 75
Reference 95

Tables

Table 1. The planned sample schools and students by language 9
Table 2. Mean scores of Students in Timed Fluency Subtasks by Language and Grade level 14
Table 3. Mean scores of Students in Untimed Fluency Subtasks by Language and Grade levels 15
Table 4. Percentage of zero readers on timed subtasks by language and grade levels 16
Table 5. Percentage of zero readers on untimed subtasks by language and grade levels 17
Table 6. The mean differences of students' performance by grade level in timed subtasks 18
Table 7. The mean differences of students' performance by gender in timed subtasks 20
Table 8. The mean differences of students' performance by location in timed subtasks 22
Table 9. Mean differences of students' performance by grade in untimed subtasks 23
Table 10. The mean differences in students' performance by gender in untimed subtasks 25
Table 11. The mean differences in students' performance by location in untimed subtasks 26
Table 12. Performance of students' in each subtask by age group 28
Table 13. Overall students' mean score by the types of program in EGRA subtasks 29
Table 14. Performance of students' by types of program across gender in timed tasks 31
Table 15 Performance of students' by types of program across gender in untimed subtasks 32
Table 16. Grade 2 overall percentage of students at benchmark levels by language 34
Table 17. Grade 3 overall percentage of students at benchmark levels by language 35
Table 18. Over all percentage of students at benchmark levels by years and grade levels 38
Table 19. Overall percentage of students at benchmark levels by year 39
Table 20. Percentage of students at benchmark levels by year in GEQIP Phase I schools 41
Table 21. Students' Background Characteristics 47
Table 22. Correlation between Oral Reading Fluency and Principals' characteristics 60
Table 23. Relationship between support given to mother tongue teachers and ORF 61
Table 24. Correlation between Oral Reading Fluency and Teachers' characteristics 61
Table 25. Correlation between oral reading fluency and students' characteristics 62
Table 26. School Resources correlated with Students' ORF as principals' response 63
Table 27. Correlation between oral reading fluency and school resources as teachers' response 64
Table 28. Correlation between oral reading fluency and school resources as students' response 65
Table 29. Correlation between Oral Reading Fluency and some teachers' activities 66

Figures

Figure 1. Mean differences of students' performance by grade level in timed subtasks 18
Figure 2. Mean differences of students' performance by gender in timed subtasks 20
Figure 3 Mean differences of students' performance by location in timed subtasks 22
Figure 4 Mean differences of students' performance by grade in untimed subtasks 24
Figure 5 Mean differences in students' performance by gender in untimed subtasks 25
Figure 6 Mean differences in students' performance by location in untimed subtasks 27
Figure 7 Students' performance in each subtask by age group 28
Figure 8 Students' performance in means scores for each subtask by program types 30
Figure 9 Percentage of grade 2 students at benchmark levels by language 34
Figure 10 Grade 3 overall percentage of students at benchmark levels by language 36
Figure 11 Overall percentage of students at benchmark levels by year and grade 38
Figure 12 Percentage of students at benchmark levels by year 40
Figure 13 Percentage of students at benchmark levels by year in GEQIP Phase I schools 41
Figure 14 Trend of overall performance in each languages by year across timed subtasks 43
Figure 15 Percentage of school principals by sex across language 44
Figure 16 Percentage of school principals by age category 45
Figure 17 Principals' Education Level by Language in Percentage 46
Figure 18 Students' Background Characteristics 47
Figure 19 Percentage of Teachers by Gender in each Language 48
Figure 20 Teachers' Education Level by Language in Percentage (%) 49
Figure 21 Service year overall as a teacher in percentage by language 50
Figure 22 Teachers' service year in percentage as a mother tongue teacher by language 50
Figure 23 Percentage of principals who received trainings on MT to support teachers 51
Figure 24 Percentage of teachers trained as a teacher across language 52
Figure 25 Percentage of teachers' participated in-service or CPD trainings on MT by duration 53
Figure 26 Percentage of teachers trained how to teach with the mother tongue materials 54
Figure 27 Percentage of school principals' response to textbook -student ratio 55
Figure 28 Percentage of students' response in having mother tongue textbook 55
Figure 29 Percentage of teachers' response in having mother tongue textbook, teacher's guide and frequency of using 56
Figure 30 Response of teachers' regarding availability of reading materials and school facilities in percentage 57
Figure 31 Response of teachers' regarding availability of reading materials and school facilities in percentage 57
Figure 32 Response of principals and teachers regarding PTSA support in percentage 58
Figure 33 Response of students regarding family's reading and support in percentage 59

Acronyms and Abbreviations

AIR	American Institute of Research
CPD	Continuous Professional Development
EAES	Educational Assessment and Examinations Service
EGRA	Early Grade Reading Assessment
FWR	Familiar Words Reading
GEQIP	General Education Quality Improvement Program
GEQIP-E	General Education Quality Improvement Program for Equity
IWR	Invented Words Reading
LC	Listening Comprehension
LNR	Letter Name Recognition
MoE	Ministry of Education
MT	Mother Tongue
NEAEA	National Educational Assessment and Examinations Agency
NEGRA	National Early Grade Reading Assessment
ORF	Oral Reading Fluency
PA	Phonemic Awareness
PTSA	Parent Teacher Student Association
RC	Reading Comprehension
RTI	Research Triangle Institution
READ	Reading for Ethiopia's Achievement Developed
SRM	Supplementary Reading Materials
USAID	United States Agency for International Development
WB	World Bank
WPM	Words Per Minute

Acknowledgements

Many people and organizations have contributed in different ways throughout the process of carrying out in EGRA 2021 study, and we are indebted to all of them. Above all, we acknowledge with many thanks Dr. Dilamo Otore, Ex-Director General of National Educational Assessment and Examinations Agency and Ato Tefera Feyissa and Ato Yilikal Wondimeneh Deputy Director Generals of Educational Assessment and Examinations Service for their follow up and support throughout the study. We would like to thank USAID/Ethiopia READ II Project Office for its financial support in data collection and analysis.

We also thank language experts from sampled language groups in adapting and reviewing EGRA data collection instruments or tools. The contributions of regional coordinators, supervisors and enumerators selected from each language groups who travelled to the schools are highly praiseworthy for collecting the required data. Moreover, our many thanks also go to the sample schools principals, teachers and students who enthusiastically participated in this EGRA study.

Moreover, we would like extend our deepest gratitude to W/o Lily Abebe from EAES ICT Directorate, for her dedication in uploading EGRA data collection instruments into Tangerine software and managing the process of data collection via software. In addition, we want to thank Mr. Solomon Worku, Senior English Language Curriculum Developer at the Ministry of Education, for his fantastic editing work of this report. Finally, we would like to thank Mr. Redi Shifa, Director of Communication Directorate at EAES, for his work in designing the EGRA report's cover pages.

Executive summary

Introduction

Early grade reading assessment (EGRA) is a type of assessment designed to provide reliable evidence that help in the acquisition and comprehension of reading skills. EGRA has been undertaken on some languages in Ethiopia since 2010 by various institutions such as RTI, FHI, AIR/READ M&E in partnership with the Ministry of Education and Regional Education Bureaus. In 2020, National Educational Assessment and Examinations Agency (NEAEA) took over the process of conducting EGRA in Ethiopia, and in May 2021, NEAEA administered the first national-wide EGRA to students in grades 2 and 3 in nine languages across seven Ethiopian regions: Afan Oromo, Af Somali, Afar Af, Amharic, Berta, Hadiyyissa, Nuer, Sidamuu Affo, and Wolayttatto, with the exception of Tigrigna due to conflict.

The fundamental goal of EGRA is to promote important policy and pedagogical dialogues about where, when, and how to enhance early grade reading ability in Ethiopia. EGRA results are also intended to help education policymakers, administrators, teacher educators, teachers, parents, and donors in establishing priorities design effective interventions that will in turn improve students' foundational reading skills.

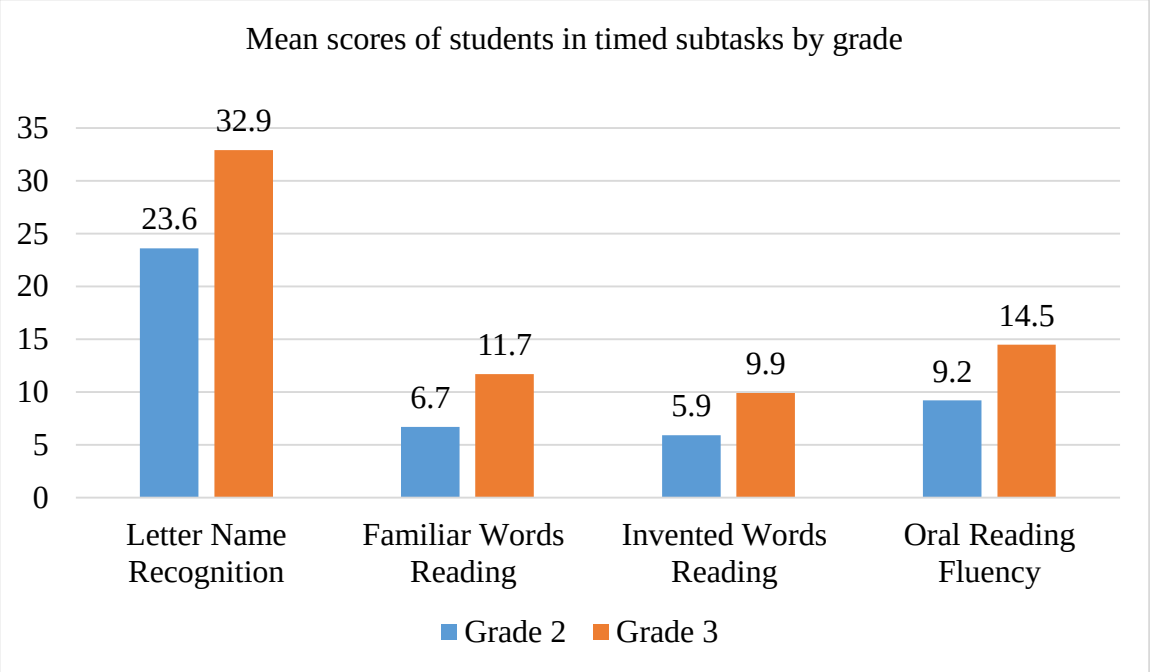
Method

The EGRA tool and subtask selection are based on reading research on the types of skills required for essential reading acquisition. There are four timed subtasks (letter name recognition, familiar words reading, invented words reading, and oral reading fluency/passage reading) and three untimed subtasks (phonemic awareness, reading comprehension, and listening comprehension) in Ethiopia's EGRA tool. READ M&E had earlier adapted and developed these tools and the contextual questionnaires that help to collect information about the home background, classroom and school in 2018 and NEAEA amended to use them in 2019. In EGRA 2021, a two stage cluster random sampling was used. In this sampling technique, while sample schools were randomly selected from each language groups in the first stage, students were selected randomly from each sample schools in the second stage. In this regard, 484 primary schools/school principals, 968 teachers and 19,360 students were selected. At each school, 20 students were selected from each grade 2 and 3. An equal number of girls and boys were randomly selected from each grade.

For the main data collection, assessors and supervisors received training in two rounds of five days on how to administer EGRA orally on a one-on-one basis using tablets. Both theoretical and practical sessions were included in the training. In the practical sessions, all data collectors took turns acting as assessors and students, and then they displayed their demonstrations so that potential faults could be recognized by the large audiences and comments given for each group. The data collectors were divided into 51 groups, each with four participants, one team leader, and three assessors. For the purpose of quality data collection, the national team had visited a sample schools in all regions where EGRA is conducted through field supervision. The data was collected in April 2021.In data analysis, mainly descriptive statistics was used to summarize the findings and inferential statistics for the purpose of generalization.

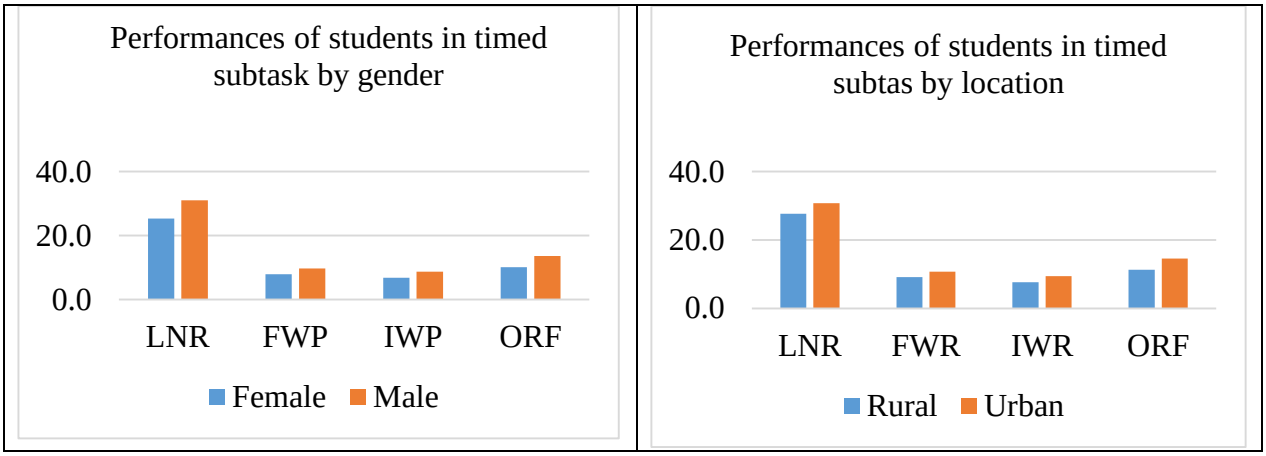
Findings

An overall reading performance on EGRA 2021 timed subtasks across grade levels was found better in letter name recognition, where grade two and three students could read on average 23.6 and 32.9 letters per minute, respectively. The least total mean scores for reading words per minute was found in reading invented words. In this case, the average mean scores for grades 2 and 3 were 5.9 words per minute and 9.9 words per minute, respectively. In both grade levels, students from Amharic performed better in familiar words reading, as well as invented words reading in grade 3. Similarly, grade 2 students from Af Somali performed better in letter name recognition and invented word reading, as well as oral reading fluency in both grades. In Afan Oromo, grade 3 students performed better in letter name recognition. However, those who learned in Nuer language performed lower in letter name recognition, while students who learned in Afar Af performed lower in the remaining timed subtasks.

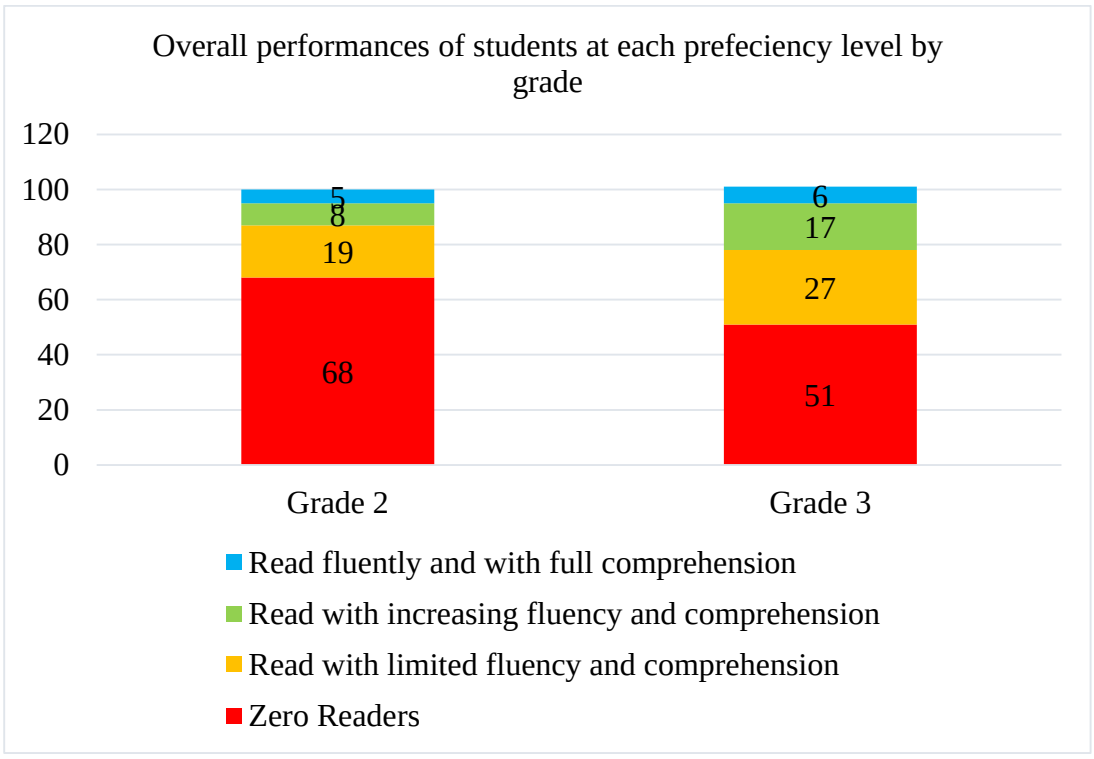


In untimed subtasks, students scored better in phonemic awareness and listening comprehension than in reading comprehension in untimed subtasks. The overall percent-correct average score for reading comprehension is only 11%, whereas it is about 53% and 68% for phonemic awareness and listening comprehension, respectively. That is, students are more likely to recognize the sounds of the letters in a word and comprehend what they listen rather than what they read in the passage.

In grade-level comparisons, the mean differences between the two grade levels in all EGRA subtasks were statistically and practically significant which in favor of grade 3 in each language. When the findings were disaggregated across gender, overall male students performed relatively better than females in each subtasks of EGRA. However, in some subtasks, female students from Af Somali and Berta performed better than males. Similarly, the disparity between urban and rural students is still persisting where urban students performed better than rural students in all subtasks at national level. Nevertheless, rural students from Af Somali, Nuer, and Sidaamu Afoo performed better than urban students in most subtasks.

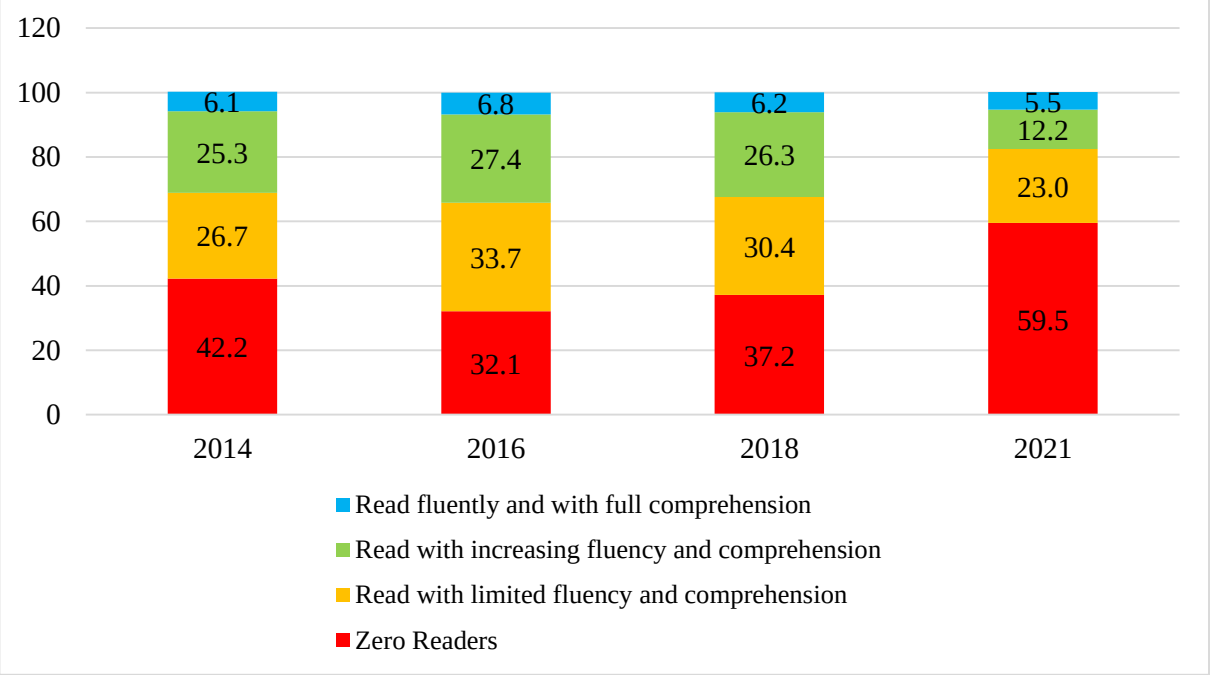


Concerning student performances by reading proficiency levels, the results show that at the national level, the percentage of pupils who attained the required highest benchmark or level 1 (reading fluently with full or almost full comprehension) was about 5% and 6% for grades 2 and 3, respectively. Similarly, the percentage of students who reached the level 2 benchmark, which refers to "reading with increasing fluency and comprehension," was nearly 8% and 17% for grades 2 and 3, respectively. Overall, only about13% of students in grade 2 and 23% in grade 3 achieved the two highest levels, referring to "full or nearly full" and "increasing" reading proficiency known as functional reading level, which almost reduced by half for both grade levels as compared to EGRA 2018. On the other hand, about 68% and 51% of the grades 2 and 3 students were respectively categorized under zero readers in oral reading fluency subtask.



Comparing the students' reading performances across years, the result of EGRA 2021 was the least for each language as compared to all previously conducted EGRA studies. According to this information, despite the fact that many interventions have been in place since the first EGRA findings released, the performances of students in oral reading fluency decreased from time to time.

Performances of students by years at each proficiency levels



According the contextual questionnaires analysis completed by school principals, teachers, and students, majority of school principals 441(92%) are males and 38 (8%) are females. The percentage of female principals in each language ranged from 2% to 19% for Wolayttatto and Berta, respectively. Among the sample teachers, only 3% of them possessed a bachelor's degree, and nearly 97% had a diploma or certificate and 7% of them are facilitators who had no took training. Concerning mother tongue teachers by gender, 46% are females. Moreover, most teachers had 1 to 5 years of experience in teaching mother tongue language. With regard to professional development, about 61% of the school principals did not receive training on how to support mother tongue teachers during instruction.

On the contrary, on average 84% of teachers responded as they took a teacher training. However, still about 16% reported that they do not take any teacher training. The percentage of teachers who did not get any teacher training ranges from 4% to 26% in Sidaamu Afoo and Berta, respectively. Particularly, about 46% of teachers did not obtain training on how to teach mother tongue. In this case, Afar Af (58%), Sidaamu Afoo (55%), and Afan Oromoo (55%) have a large number of teachers who did not get training.

Regarding the availability of school facilities and instructional materials, about half of the principals responded as most of students shared a mother tongue textbook in a 1 to 5 ratio. Similarly, it was found that 34% and 67% of students replied as they did not have any mother tongue textbooks and additional reading books respectively. This result confirms that textbook scarcity is critical in mother tongue instruction for both grades. According to the response of

the teachers, around 7% and 27% of them lacked student textbooks and teacher's guides respectively. In addition, 54%, 83%, and 60% of teachers said that their school do not have a functional library, reading corner, and supplemental reading materials, respectively. Similarly, 58% and 85% of students confirmed that their schools did not have a functioning library and reading corner, correspondingly.

Concerning the involvement and assistance of parents in student learning, 47% and 77 % of school principals said that parent-student-teacher association is not functional and it does not assist student learning, respectively. In addition, 33% and 42% of students stated that they do not have enough time to study/read at home and do not get any support at home, respectively.

Looking the relationships between some contextual variables and students' performance, though, most of the variables showed weak correlations: gender, highest education level, provision of trainings on assisting MT teachers, and the length of trainings were some of the school principal's variables that were significantly associated with the students' ORF. Furthermore, according the response of the participants, variables related to school resources that showed positive correlations with students' reading performances includes: the availability of mother tongue textbooks for students, availability of mother tongue textbooks and teachers' guides for teachers, presence of reading corners and functional libraries, existence of functional PTSA and accessibility of supplementary reading materials for students. Particularly, the finding from teachers' responses showed that, the method of teaching and assessment techniques that teachers applied, positively associated with the students reading performance in addition to class size had also a negative correlation.

As the current findings indicate, the reading performances of the students is very low. Among the additional factors that contribute to the students' low performances are COVID-19 and internal conflicts as well as the drought in certain areas can be mentioned. The impact of the COVID-19 pandemic on schooling was one of the key reasons for students' poor reading performance. Due to the school closure, the formal schooling was disrupted, and students were not actively involved in schoolwork for more than four months. This is supported by the research finding, 44% of students were without schoolwork during the COVID-19 schools closure and 61% of teachers did not engage in any schoolwork during the closures (IEA & NEAEA, 2021). Internal conflict and instability in some parts of the country could also the other reason for the students' low performance, as 20 and 16 sample schools were replaced at the spot of data gathering due to conflict and drought, respectively.

Conclusions

The overall performance of students in EGRA 2021 was the least as compared to all previously conducted EGRA studies. The performance of students across subgroups showed that grade 3 students performed better than grade 2 students. This indicates a positive gains across grade levels. With regard to gender and location, males performed better than females and urban students performed better than rural students in most EGRA subtasks in all language groups. That means, the variability of performance by gender and school location is still a persisting problem.

When comparing student performance across proficiency levels, only about 13% and 23% of grade 2 and 3students achieved the functional reading level or the two upper to levels respectively. On the other hand, about 59 % of students categorized under zero readers at national level.

Regarding the findings from contextual questionnaires, the percentages of female principals are very small in each language groups and ranged from 2% to 19% for Wolayttatto and Berta,

respectively. However, the percentages of mother tongue female teachers are 46% as compared to males. Concerning the qualification of the teachers, 97% are diploma or certificate holders at national level. However, 7% of them are facilitators who did not attend any formal training. Looking the working experiences of the mother tongue teachers, most of them had 1 to 5 years of experiences. Regarding the professional development for principals and teachers, 61% of the school principals did not receive training on how to support mother tongue teachers. On the other hand, 84% of teachers responded that they have obtained a teacher training. However, still 16% reported that they did not take any teacher training.

With regard to school facilities and instructional materials, nearly half of the principals responded as most of the students shared a mother tongue textbook in a 1 to 5 ratio. Similarly, 34% and 67% of the students replied as they did not have any mother tongue textbooks and additional reading books respectively. In addition, as the response of the teachers, 7% and 27% of them lacked student textbooks and teacher's guides respectively. Moreover, 54%, 83%, and 60% of teachers said their school lacked a functional library, reading corner and SRM correspondingly. Likewise, 58% and 85% of the students confirmed that their schools did not have a functional library and reading corners respectively.

According the findings of this study, the main principals contextual variables correlated with the students' oral reading fluency were their gender, highest education level, training on assisting MT teachers, and the length of trainings provided. The findings from the teachers' responses indicated that their teaching methods and assessment techniques were positively associated with the students reading performances. In addition, variables related to school resources that showed positive correlations with students' reading performances include: the availability of mother tongue textbooks for students, availability of mother tongue textbooks and teachers' guides for teachers, presence of reading corners and functional libraries, presence of functional PTSA and accessibility of SRM for students.

Among the additional factors that contribute to the students' low performances COVID-19 pandemic, internal conflicts and drought in certain areas can be mentioned. Since, because of the school closure, the formal schooling was disrupted, and students as well as teachers were not actively involved in schoolwork for more than four months. On the other hand, 20 and 16 sample schools were replaced at the spot of data collection due to the internal conflict and drought respectively.

Recommendations

The overall results of EGRA 2021 indicate the low performances of students as compared to the previous EGRA studies. So, to improve the performance of students' reading skills, all the concerned governmental bodies, non-governmental organizations and stakeholders are required to make a rigorous effort. Moreover, based on the suggestions forwarded by the stakeholders who participated in EGRA 2021 dissemination workshop, there is need to make the issue of EGRA a national agenda and establish system level accountability in general education to promote students' learning.

Specifically, minimize the disparity in reading performance across subgroups such gender, location and language groups, improve the quality of school leadership, strengthen teachers' professional development, encourage teachers to use various methods of teaching and assessment techniques, provide school facilities and instructional materials, aware educational leaders how to cope with educational disruptions, regular monitoring and evaluating the impact of EGRA interventions are some recommendations forwarded for further actions (see the details under 4.2 section).

1. Introduction

1.1 Overview of Early Grade Reading Assessment (EGRA)

Early Grade Reading Assessment (EGRA) is a tool which was initially created to provide a reliable and valid measure of skills that contribute to reading acquisition and comprehension (Dubeck & Gove, 2015). Since 2007, the United States Agency for International Development (USAID) has supported the development and administration of the Early Grade Reading Assessment (EGRA) in more than 70 countries throughout the world (AIR, 2018). In Ethiopia, EGRA has been conducted since 2010 by different institutions like RTI, FHI, AIR/READ M&E in collaboration with the Ministry of Education and Regional Education Bureaus (REBs).

Dubeck and Gove (2015) in their study on the significance of early reading to academic success stated that learning to read is the skill that a child will use for the rest of his or her life, which will ultimately help and affect their future academic success. Reading is strongly linked to academic achievement, since it is incorporated into every subject at higher grade levels. According to Dubeck and Gove, students who are able to read and who have more advanced vocabularies and reading comprehension skills tend to do better in school and score higher on standardized tests. On the contrary, students who struggle to learn how to read, or who give up and do not learn at all, have more difficulties with the rest of their education.

The EGRA tool and the selection of the subtasks is based on reading research regarding the types of skills essential for reading acquisition. The EGRA tool developed in Ethiopia has four timed subtasks (letter name recognition, familiar words reading, invented words reading, and passage reading) and three untimed subtasks (phonemic awareness, reading comprehension, and listening comprehension).

EGRA results are intended to help education policymakers, administrators, teacher educators, teachers, parents, and donors in establishing priorities to improve interventions that will in turn increase foundational reading skills. EGRA does not serve as a high-stakes accountability tool for evaluation of students, teachers or schools. Nor does it serve for direct cross-language comparisons of reading achievements. On the other hand, EGRA is not concerned on an individual student reading skill progress; rather, it is an overall measure of the early reading performances in the education system (AIR, 2018).

EGRA has been conducted in Ethiopia for four cycles so far. The last consecutive cycles (EGRA 2016 and EGRA 2018) assessments were both conducted by American Institute for Research/AIR/ READ M&E. EGRA 2021 is the fifth round of this type which was conducted under the auspices of the National Educational Assessment and Examinations Agency (NEAEA). The assessment was carried out in nine instructional languages; namely, Amharic, Afan Oromo, Af Somali, Sidamu Afoo, Hadiyyissa, Wolayttatto, Afar Af, Berta and Nuer. Be that as it may, it was conducted for the first time in the three languages, that is to say, Afar Af, Berta and Nuer and where their results may serve as a baseline.

Overall, the purpose of EGRA is to measure basic reading skills of Grades 2 and 3 students that they must acquire to read fluently and with comprehension. With this objective in view, EGRA results are used to inform policy makers and stakeholders for subsequent plans and interventions.

1.2 EGRA Subtasks

EGRA has seven subtasks that measure basic skills of students's reading acquisition. Each of the seven EGRA subtask is described below.

Letter Name Identification: This sub task measures knowledge of letter names. Accordingly, most frequently used letters were selected from reading books or any grade-appropriate reading texts. Hundred letters were selected in random order from both upper and lower cases for those using English alphabet and some *Fidels* in case of Amharic. This subtask was timed to 60 seconds and discontinued if none of the letters in the first line was read correctly. Letter identification knowledge is a fairly common assessment approach and is used in several early reading assessments, including the preschools (Dubeck & Gove, 2015 & RTI, 2016). The subtask assesses students' knowledge of the alphabetic principle which is the foundation of learning to read.

Phonemic Awareness: This sub task is concerned with how sounds work with words. It is considered as a pre-reading skill, and can be assessed in a variety of ways. Depending on the language tested, the task might be designed to determine whether students could differentiate the first or the last syllable in a word, or whether they could identify all of the letters in a word. For this subtask a list of 10 randomly selected frequent words were used. The frequency of words in everyday use is determined during development of the subtask by text analysis and calculations of word count frequencies.

Familiar Words Reading: This sub task presents a list of words that students are expected to be able to read at their grade level and will have likely encountered before**.** It measures student's ability to read words that are familiar to the students of grade 2. Fifty most frequently used words are selected from grade 2 textbooks and they are arranged in random order in grids (Dubeck & Gove, 2015).

Invented Words Reading: invented words sometimes called unfamiliar or nonsense words present a reading subtask for students with a written list of pseudo words that follow the phonological and spelling rules of the language but are not actual words in the language (AIR, 2018). Students are asked to read out loud as many of the non-words as they can, as quickly and carefully as they can.

This subtask requires students to apply their decoding skills based on their knowledge of the grapheme-phoneme mappings. Because non-words will not have any whole-word representation previously stored in long-term memory to be accessed directly, students must rely on decoding in order to identify them (RTI, 2016).

The sub task measures the ability of students to decode one-and two-syllable non-words that could plausibly exist in the language. To achieve fluency in reading, students need to acquire both sight-reading and decoding skills.

Like the familiar words reading task, fifty non-words are randomly arranged in a list in the student booklets, and students are asked to read as many of the non-words as they can.

Oral Reading Fluency: This is an oral passage reading subtask which measures the ability of a student to read passage texts with speed, accuracy, and proper expression (AIR, 2018; RTI, 2019;

2016). Each student is provided with a grade-appropriate short story and given one minute to read aloud. The purpose of the timed oral reading fluency (ORF) subtask is to examine whether pupils in Grades 2 and 3 can read a passage with speed and accuracy of grade-appropriate words as presented in the pupil booklets.

Reading Comprehension: This subtask relies on questions about the text read in the ORF subtask. Comprehension as an ultimate goal of reading enables students to make meaning out of reading text they read and use that meaning to learn new things (Anteneh et al., 2016; RTI, 2016). This task determines students' understanding of the text and their ability to answer factual questions and make inferences based on what they read. Reading comprehension is also a highly complex task that requires both extracting meanings from text. After a student completes the ORF subtask, the administrator moves to the reading comprehension task which includes a series of questions about the passage.

Listening Comprehension: Listening comprehension is a pre-reading skill that measures oral fluency and vocabulary, for whether students are able to actively listen and understand the language the teacher uses as he/she teaches (RTI, 2019). Although the length of reading text varies with language type, the subtask includes approximately a paragraph about 40 words. The test administrator reads the passage aloud only once at a pace of one word per second. When the administrator completes reading of the text aloud, he or she then asks students five comprehension questions.

1.3 Rationale

Acquiring literacy becomes more difficult as students grow older; students who do not learn to read in the early grades are more likely to repeat grades and eventually drop out of school. That is, if strong foundational skills are not acquired in early grades, gaps in learning outcomes in the next upper grades become widened (RTI, 2016). The common metaphor of "the rich get richer and the poor get poorer" supports the state of disparities that occur between fluent and non-fluent readers for students who are unable to acquire reading and comprehension skills in the early grades.

However, majority of students enrolled in the early grades in Ethiopia have difficulty to read texts written in the instructional language or their mother tongues. Trends in the reading performances of previous assessments show that students' performances are unsatisfactory in almost all the

reading subtasks. For instance, if we look at the 2018 EGRA results, only few number of Ethiopian students in Grades 2 and 3, and across all languages combined, achieved the targeted reading benchmark or reading fluently with full or almost full comprehension.

Moreover, according to EGRA 2018, the performances of students with zero-scores (nonreaders) was ranging from high percentage in Hadiyyissa (60.2%) to relatively low percentage of non-readers in Amharic language (11.3%).

1.4 Objectives

The general objective of EGRA 2021 is to measure reading skills of early grade (Grades 2 and 3) students in order to inform governments and donors regarding the students' performances system needs for improving instruction. Accordingly, EGRA 2021 includes the following specific objectives:

- to find out the overall performance of student's reading skills by each language groups;
- to analyze the performance differences of students in reading skills across subgroups such as grade level, gender, and location;
- to identify the proportion of students at each benchmark of the proficiency levels;
- to indicate what changes have been made over time across each years of EGRA administration;
- to identify factors that are associated with the students' performance in reading; and
- to suggest action points for policy makers and stakeholders.

1.5 Basic Research Questions

The basic research questions of this EGRA study are stated as follows:

- What is the overall students' reading performance across language groups in timed and untimed subtasks?
- What are the differences in performances among the various groups in EGRA 2021?
- What are the performances of students at each proficiency level or benchmark?
- What is the current performance of students as compared to the previous EGRA studies?
- What are the factors associated with the students oral reading performances?

1.6 Limitation

In this study, due to the internal conflict, schools from Tigray region and the Metekel zone of Benishangul Gumuz region were not part of this study. Furthermore, because of conflict and drought in the local areas, some sampled schools from various regions were replaced at the moment of data collection.

2. Method of the Study

Section II overviews the design and the method of study. It is presented in five parts. The first part details the research design applied in this study. The second part presents the sampling techniques and the participants of the study. The third part delves deep into the process of developing the data collection instruments. The fourth part describes the procedures of data collection and data management. The fifth part culminates section II. It outlines the data analysis techniques used in the study.

2.1 Research Design

In EGRA 2021, a survey research design was applied in order to assess the reading performance of Ethiopian students at Grades 2 and 3. Particularly, a cross-sectional survey study was used to capture information based on data gathered for a specific point in time. Such survey is used to gather information to make inferences about a population of interest at a point in time from the samples. This research design is especially useful when one needs to make direct statistical generalizations from a sample to a population (Christensen, Johnson, & Turner, 2013). According to McMillan and Schumacher (2006), the reason for the researchers turn face to the survey research was, because of the nature of its flexiblity and more appropriate for collecting a large data from diversefied type of populations. That means, it is the concept for working with large data on geographically scattered populations, when the other methods of research are not always feasible (Brewer, 2009).

2.2 Particpants and Sampling Techniques

The target population of this study consisted of Grades 2 and 3 students in all Ethiopian primary schools. Therefore, the participants of the EGRA 2021 study were students enrolled in Grades 2 and 3 in 2021, their mother tongue (MT) teachers and the principals of the sample schools.

Regarding the sampling and the sampling techniques, nine instructional language groups were selected for EGRA 2021 study based on the interest of the policy makers. The selected languages were: Afan Oromo, Afar Af, Af Somali, Amharic, Berta, Nuer, Hadiyissa, Sidaamu Afoo and Wolyattatto. In this case, due to security concerns in Tigray region, the Tigrigna language, which had been part of the EGRA project for a decade was not included in this assessment.

Using a two-stage cluster sampling method, the sample of schools and students were selected randomly from each instructional language groups. To determine the sample size, the formula of a two-stage cluster sample design a= $\frac{400}{b}$ × [1+ (b-1)ρ] which would provide an equivalent sampling accuracy for a simple random sample of 400 was applied (Ross, 2007). In this formula, 'a' represents the number of clusters or schools, 'b' represents the number of cluster size and ρ (rho) stands for the coefficient of intraclass correlation or rate of homogeneity. In this case, the cluster size 'b' was 40 students, where each 20 of them from Grades 2 and 3. In addition, taking the data of oral reading performance as the dependent variable and the schools as a random factor from the EGRA 2018, the variance component analysis of an interclass correlation rho was found 0.12. The value of rho indicates a measure of the tendency of student characteristics to be more homogeneous within schools than would be the case if the students assigned to the schools at random.

Thus, according to the formula indicated above, theoretically about 57 sample schools are enough to conduct EGRA 2021 at the national level. However, if these sample schools are shared by each nine instructional language groups using probability proportional to their school sizes, the ratio of some language groups with smaller number of school population could be under representative or even could be zero. So, in order to have a statistically powerful data across each language group, it is necessary to optimize their sample size. Hence, an adjustment was made by adding a constant number of 30 schools to each of them. As Gay, Mills and Airasian (2012) indicated, if the population is larger than 5,000, the sample of 400 would be considered adequate. In this regard, 484 sample schools were selected at national level with a range of 30 to 85 schools from each language group. This results in a total number of 19,360 students' and 1,200 to 3,400 range of number of students per language group. Among 484 schools, 320 schools were selected for National EGRA (NEGRA). The remaining 86 and 78 schools were selected for GEQIP phase I and READ II, respectively. In case of GEQIP phase I, sample schools were selected from all language groups expect for Nuer, due to the fact that Nuer language was not part of phase I schools (see Table 1). Finally, under this two-stage cluster sampling design, in the first stage, the sample schools were selected randomly from each of the sample language groups, and then 20 students (10 boys and 10 girls) from Grade 2 and 20 students (10 boys and 10 girls) from Grade 3 were selected randomly from each sample schools during the second stage. At this point, mother tongue

teachers of Grade 2 and 3 selected students and the school principals were automatically selected from the randomly sampled schools.

Table 1. The planned sample schools and students by language

Language	NEGRA		GEQIP Phase I		READ II		Total	
	Schools	Students	Schools	Students	Schools	Students	Schools	Students
Afan Oromo	51	2040	17	680	17	680	85	3400
Af Somali	34	1360	11	440	10	400	55	2200
Afar Af	31	1240	4	160	-	-	35	1400
Amharic	50	2000	16	640	16	640	82	3280
Berta	30	1200	2	80	-	-	32	1280
Hadiyyissa	31	1240	12	480	10	400	53	2120
Nuer	30	1200	-	-	-	-	30	1200
Sidaamu Afoo	32	1280	13	520	13	520	58	2320
Wolayttato	31	1240	11	440	12	480	54	2160
	320	**12800**	**86**	**3440**	**78**	**3120**	**484**	**19360**

2.3 EGRA Instrument Development

EGRA data collection instrument in seven languages: Afan Oromo, Amharic, Af somali, Hadiyyissa, Sidama Afoo, Tigrigna and Walayittato were previously adapted, piloted and became operational in previous EGRA, 2018. The same tool was used in EGRA, 2021. However, there was a reviewing and further validation process by language experts which was held in February, 2020. During validation, the language groups took into an account the characteristics of each language and made a minor revision.

In addition to the previous EGRA seven instruction languages, three additional languages namely: Afar Af, Nuer, and Berta were included in EGRA, 2021 in order to check the reading status of students in emerging regions and based on the recommendation from policy makers and stakeholders. For these three languages, EGRA instrument adaptation and translation had been conducted in the last week of October, 2020. For adaptation purposes: Firstly, language experts and native speakers in each language were recruited. Secondly, the team of language experts took trainings on how to adapt the instrument to the local context. Besides, manual was developed and given for each language group regarding instrument adaptation and translation.

During adaptation and translation of EGRA instruments, due attention was given to ensure that the instruments were culturally and socially appropriate, relevant to the peculiar nature of the language, free from religious and gender bias. Therefore, adapting the instrument means that student tasks were developed from scratch based on the rules and structures of the language of assessment as used locally and based on the vocabulary in the relevant local reading materials or syllabi. Two forms (versions) were developed (A and B). Thus, for both forms stimulus materials were designed for the subtasks: letter identification, familiar word reading, invent word reading and oral passage reading.

The instruments were piloted on 30 sample schools that were randomly selected from the list of all the schools in the sampling frame. That means, from each language groups 10 schools were selected. Then, training was given for enumerators and supervisors/team leaders/ in February 2021 for 5 days. Each language group consisted of three enumerators and one team leader. The training was accompanied by practical demonstrations on how to administer EGRA orally with the help of electronic data collection. Debriefs and discussions were made using model demonstrations to facilitate peer learning. In this case, Samsung Tablets A loaded with Tangerine Software were used for the data collection. Data collection took place in mid-February 2021 for ten consecutive working days. In order to ensure the quality of data process and reliability of the tools, pilot schools were visited through field supervision by the NEAEA experts. Furthermore, supervisors from NEAEA were assigned to monitor the process of data collection using the observation checklist.

Item analysis of the pilot study was presented to the language groups through workshop. Item difficulty index of both forms (A and B) in all tasks has been presented to the language groups and items in both forms were critically investigated task by task in each language group. Form A had relatively better difficulty level than form B for most of the tasks for the three instructional languages. However, some items from form B had good difficulty level. Items with good difficulty index were retained but some improvements were made for those items with poor quality. For some of the tasks, items were selected from both forms by examining the difficulty level of the items. Lastly, one final tool was developed by selecting and assembling items from both forms for each language. The final tool was used for EGRA operational purpose which was conducted in April 2021.

The National Educational Assessment and Examinations Agency (NEAEA) used Tablets loaded with Tangerine Software to complete both the pilot for the three languages and operational data collection for the nine languages. In addition to student subtasks, there was other data collection instruments (students', teachers' and principals' background questionnaires) uploaded to the Tangerine Software to evaluate the association between student reading performance on the subtasks and background factors.

2.4 Main EGRA Administration Procedures

The NEAEA team conducted a training workshop for assessors and supervisors in two rounds each for five days in the last week of March and first week of April 2021. The purpose of the training was to equip the assessors and supervisors with the necessary skills on how to administer EGRA 2021 orally on a one-on-one basis using tablets. The training accommodated both theoretical and practical sessions. In practical sessions, all data collectors acted as assessors and students turn by turn and lastly they made demonstrations so that possible errors could be identified by the large audiences and comments were given for each group. The same procedure was repeated until each assessor was able to manage the oral assessment of student using the tablet. Since a large proportion of assessors had experience in previous years EGRA data collection, the vast majority of the assessors did not face problems in the utilization of the Tablet.

EGRA data collection was conducted from April 5-24, 2021. Data collectors were organized into 51 teams and each team consisted of four members (one supervisor/team leader and three assessors). Supervisors/team leaders and assessors sampled students from both Grade 2 and 3 as per the EGRA administration protocol. Administration manual was given to each data collector. The three assessors in each group took the responsibility for administering instruments directly with students of both grade levels. Accordingly, assessing one student with the EGRA tool lasted for 15-20 minutes. From this, one can conjecture that an assessor can assess about 10–13 students in a typical school day.

In the meantime, while the assessors were conducting oral assessment of students, the team leader or supervisor was involved in assessing Grades 2 and 3 MT teachers as well as the school principals using the Tablet. Besides, the team leader was also responsible for managing students who were sampled from both grades until the assessors finish their duties.

In order to ensure the quality and reliability of data, the national team had visited schools in all regions where EGRA is operational through field supervision. Thus, members of the national team observed and checked for the following points.

- How the sampling of students has been accomplished in schools.
- The convenience of assessment room or place (whether or not the assessment room/place is free from destruction and noise).
- How assessors interact with students.
- How assessors administer each task of the EGRA tool (whether or not it is as per EGRA Administration Protocol).
- Pace of assessors while administering the tasks.
- The proper utilization of the Tablets while assessing the student.
- Completion of student assessment forms.

2.5 Data Analysis

The collected data was analized by using both descriptive and inferential statistics. As a precursor to the analysis, the scores for EGRA timed tasks were calculated as the number of letters or words correctly read per minute, whereas the scores for untimed tasks were calculated as the percentage of correct responses out of the total number of questions in the subtask. Then, the descriptive summary statistics such as frequencies, percentages and means scores were computed to summarize the overall finding of the study. Independent samples t-test used to compare the mean differences of two groups such as gender and grade levels and Analysis of Variance (ANOVA) was used to compare mean differences among the program types. Correlation was also used to associate the student backround variables with their performces on oral reading fluency.

To clean the data and analyize the findings, Microsoft Excel and the Statistical Package for Social Science (SPSS) Version 25 were applied. Here in all cases, the results were interpreted based on significance of statistical analysis, 95% confidence level with an accepted level of probability significance value $P = 0.05$

3. EGRA Findings

This section provides the data analysis and findings. The main purpose of this EGRA 2021 assessment is to measure the reading skills of early grade (Grades 2 and 3) students. This study used a cross-sectional survey design to capture information based on data gathered for a specific point in time. As this section is the backbone of the study, it subsumes nine parts and each part is further broken down into chunks as the need arises. This section is organized to address the five research questions of this study. They include, among other things, the overall students' reading performance across language groups in timed and untimed subtasks; the differences in performances among the various groups in EGRA 2021; the performances of students at each proficiency level or benchmark; the current performance of students as compared to the previous EGRA studies and the factors that are attributed to the students oral reading performances.

3.1. Overall Mean Scores of Timed and Untimed Subtasks

3.1.1 Students' performance on timed subtasks by language and grade level

Table 2 presents the overall performance of students by language groups and grade levels in each timed subtasks (letter name recognition, familiar words reading, invented words reading and oral reading fluency). These subtasks were measured by calculating the number of letters or words learners read per minute. In this regard, the better performance was found in letter name recognition, where Grades 2 and 3 students could read on average 23.6 and 32.9 letters per minute respectively. The least total mean scores for reading words per minute was found in reading invented words. In this case, while the average mean scores for Grade 2 was 5.9 words per minute and that of Grade 3 was 9.9 words per minute. In both grades, students from Amharic language performed better in reading familiar words, as well as reading invented words in Grade 3. Similarly, Grade 2 students from Af Somali performed better in letter name recognition and invented word reading, as well as oral reading fluency in both Grades 2 and 3.As far as Afan Oromo is concerned, Grade 3 students performed better in letter name recognition. However, those who learned in Nuer language performed lower in letter name identification, while students who learned in Afar Af performed lower in the remaining timed subtasks (familiar words reading, invented words reading, and oral reading fluency).

Table 2. Mean scores of Students in Timed Fluency Subtasks by Language and Grade level

Language	Letter Name Recognition		Familiar Words Reading		Invented Words Reading		Oral Reading Fluency	
	Gr. 2	Gr. 3	Gr. 2	Gr. 3	Gr. 2	Gr. 3	Gr. 2	Gr. 3
Af Somali	44.5	46.6	14.3	18.0	14.3	17.9	22.5	29.8
Afan Oromo	31.8	53.3	5.7	13.6	4.5	10.2	5.9	15.5
Afar Af	14.1	18	2.1	3.9	1.8	3.6	2.2	4.2
Amharic	22.4	32.8	15	24.7	12.5	18.6	17.2	29.4
Berta	19	18.8	4.5	7.4	4.2	8	9.9	6.4
Hadiyyissa	23.5	37.8	4.6	9.7	4.1	8.7	4.6	9.7
Nuer	9.1	16.3	3.4	6.3	2.6	5.5	8	11.1
Sidaamu Afoo	29.3	44.4	6.1	12.3	4.7	9.4	7	14.4
Wolayttatto	19.1	28.5	5	9.1	4	7.2	5.8	10.3
National	**23.6**	**32.9**	**6.7**	**11.7**	**5.9**	**9.9**	**9.2**	**14.5**

3.1.2 Students' performance on untimed subtasks by language and grade level

The overall performance of students by language group in untimed subtasks (phonemic awareness, reading and listening comprehension) is shown in Table 3. These subtasks were measured by calculating the percentage mean scores that students respond correctly to each provided items. As a result, the total percent mean scores for Grade 2 learners in phonemic awareness, reading comprehension and listening comprehension were 53.5%, 7.6% and 64.5%, respectively. Likewise the total percent mean scores for Grade 3 learners in phonemic awareness, reading comprehension and listening comprehension were 63.2%, 15% and 70.8% respectively. The table also discloses that reading comprehension received the lowest mean score for all languages and both grade levels, which call for due attention. On the contrary, the performance of students in both grade levels across all languages was better in listening comprehension. Similarly, the performance of students was better in phonemic awareness except for students from Afar Af in both grade levels and Afan Oromo, Berta and Nuer languages for Grade 2 students.

Table 3. Mean scores of Students in Untimed Fluency Subtasks by Language and Grade levels

Language	Phonemic Awareness		Reading Comprehension		Listening Comprehension	
	Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3
Af Somali	77.0	73.8	19.2	29.9	66.2	74.7
Afan Oromo	39.1	63.0	6.7	19.3	64.7	75.5
Afar Af	15.8	23.6	1.1	3.5	50.9	56.5
Amharic	72.6	80.1	16.4	31.1	69.3	73.8
Berta	31.2	40.0	2.1	2.6	55.8	62.2
Hadiyyissa	73.7	78.6	4.8	8.9	74.2	81.5
Nuer	30.3	54.6	5.5	14.6	76.8	80.5
Sidaamu Afoo	82.0	85.6	6.2	13.8	75.8	82.5
Wolayttatto	60.0	69.4	6.3	11.2	47.1	50.0
National	**53.5**	**63.2**	**7.6**	**15.0**	**64.5**	**70.8**

3.2 Zero Readers on Each Subtasks

Table 4 shows the percentage of zero readers on timed subtasks by language groups and grade levels. In this regard, the total average percentages of zero readers in Grade 2 are about 33%, 65%, 70% and 68% for letter name recognition, familiar words reading, invented words reading and oral reading fluency subtasks respectively. Similarly, in Grade 3, the total percentages of zero readers are 23%, 49%, 54% and 51% for letter name recognition, familiar words reading, invented words reading and oral reading fluency subtasks respectively. In this case, a large percentage of zero readers are from grade two in each timed subtasks.

Looking across language groups, in all timed subtasks in both grade levels, the largest zero readers are from Afar Af and Berta languages. On the other hand, in familiar words reading, invented words reading, and oral reading fluency, the smallest zero readers are from Amharic language for both grade levels. However, in letter name recognition, the least percentages of zero readers are from Afan Oromo in both grade levels.

Table 4. Percentage of zero readers on timed subtasks by language and grade levels

Language	Letter Name Recognition		Familiar Words Reading		Invented Words Reading		Oral Words Reading	
	Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3
Af Somali	16	17	36	30	40	32	35	20
Afan Oromo	15	4	54	26	66	42	63	36
Afar Af	63	55	88	82	91	84	92	85
Amharic	27	16	35	19	33	18	34	18
Berta	53	46	91	85	94	89	92	91
Hadiyyissa	26	15	69	48	74	54	78	56
Nuer	49	25	78	55	81	61	71	49
Sidaamu Afoo	16	6	64	39	70	48	70	47
Wolayttatto	34	21	71	54	77	62	76	58
National	**33**	**23**	**65**	**49**	**70**	**54**	**68**	**51**

Table 5 indicates the percentages of zero readers in untimed subtasks by language groups and grade levels. As a result shows, in grade two, the total percentages of zero readers are 26%, 83%, and 10% for phonemic awareness, reading comprehension and listening comprehension subtask respectively. Similarly, in Grade 3, the total percentages of zero readers are 18%, 69%, and 7% for phonemic awareness, reading comprehension and listening comprehension subtasks respectively. In each untimed subtask except for phonemic awareness for Grade 3 in Af Somali, a large percentages of zero readers from Grade 2 rather than Grade 3.

In reading comprehension, there is a higher percentage of zero readers as compared to listening comprehension in all language groups for both grade levels. Afar Af had the highest percentage of zero readers in all untimed subtasks for Grade 2 and also it had the highest percentage of zero readers in Grade 3 for phonemic awareness and listening comprehension subtasks. The table also indicates that Berta had the highest percentage of zero readers for reading comprehension. On the contrary, in phonemic awareness and listening comprehension, the lowest percentages of zero readers are Grade 2 students from Sidaamu Afoo. Likewise, in reading comprehension, the lowest percentages of zero readers are from Amharic language. In Grade 3, the lowest percentages of zero readers are from Hadiyyissa in phonemic awareness and listening comprehension, while those are from Amharic in reading comprehension.

Table 5. Percentage of zero readers on untimed subtasks by language and grade levels

Language	Phonemic Awareness		Reading Comprehension		Listening Comprehension	
	Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3
Af Somali	7	9	69	53	12	5
Afan Oromo	41	19	85	60	9	3
Afar Af	65	56	96	90	32	24
Amharic	14	9	60	37	3	2
Berta	44	35	96	95	13	12
Hadiyyissa	4	1	85	72	3	1
Nuer	42	21	85	69	4	4
Sidaamu Afoo	3	2	81	64	2	1
Wolayttatto	17	10	88	77	8	8
Total Av.	**26**	**18**	**83**	**69**	**10**	**7**

3.3 The mean differences of students' performance by subgroups

In this part, student performance in the 2021 EGRA administration is presented in terms of mean difference scores for each subtasks, which are disaggregated by certain subgroups such as grade levels, gender, location, age category, and language groups.

3.3.1 The significant mean differences in timed subtasks by grade level

Table 6 and Figure 1 indicate the significant mean differences of students' performance by grade level for each language in timed subtasks. In this case, the results of testing for statistically significant differences between grades (by independent sample t-tests) and practical differences between grades (by Cohen's d) were applied as shown in Appendix A in details.

In each language group, the mean differences between the two grade levels in all EGRA timed subtasks are statistically significant and in favor of Grade 3. This means that in all languages, Grade 3 students were able to read better than Grade 2 students, which indicates positive grade gain. Particularly, the effect sizes of Grade differences for oral reading fluency in Afan Oromo and Amharic, as measured by Cohen's d, were 0.61 and 0.59, respectively, and can be classified as strong educational impact that was showed as something substantially changed. Since the Cohen's d was 0.25 and higher in the Af Somali, Hadiyyissa, Sidaamu Afoo and Wolayttatto languages, it implies that the amount of the change is educationally significant and that something was taught. Notwithstanding to this, there was statistically significant difference, there were a small effect sizes between grade levels for Afar Af, Berta, and Nuer languages.

Table 6. The mean differences of students' performance by grade level in timed subtasks

Language	Grade	Letter Name Recognition	Familiar Words Reading	Invented Words Reading	Oral Reading Fluency	Cohen's d For ORF
Af Somali	Grade 2	44.5	12.1	14.3	22.5	-0.29
	Grade 3	46.6	14.3	17.9	29.8	
Afan Oromo	Grade 2	31.8	5.7	4.5	5.9	-0.61
	Grade 3	53.3	13.6	10.2	15.5	
Afar Af	Grade 2	14.1	2.1	1.8	2.2	-0.16
	Grade 3	18.0	3.9	3.6	4.2	
Amharic	Grade 2	22.4	15.0	12.5	17.2	-0.59
	Grade 3	32.8	24.7	18.6	29.4	
Berta	Grade 2	19.0	4.5	4.2	9.9	0.08
	Grade 3	18.8	7.4	8.0	6.4	
Hadiyyissa	Grade 2	23.5	4.6	4.1	4.6	-0.33
	Grade 3	37.8	9.7	8.7	9.7	
Nuer	Grade 2	9.1	3.4	2.6	8.0	-0.13
	Grade 3	16.3	6.3	5.5	11.1	
Sidaamu Afoo	Grade 2	29.3	6.1	4.7	7.0	-0.46
	Grade 3	44.4	12.3	9.4	14.4	
Wolayttatto	Grade 2	19.1	5.0	4.0	5.8	-0.25
	Grade 3	28.5	9.1	7.2	10.3	

Note. Cohen's d of 0.25 and above indicates the size of difference is educationally significant; something was learned and Cohen's d of 0.50 and above indicates a strong educational effect; something substantially changed (Wolf, 1986)

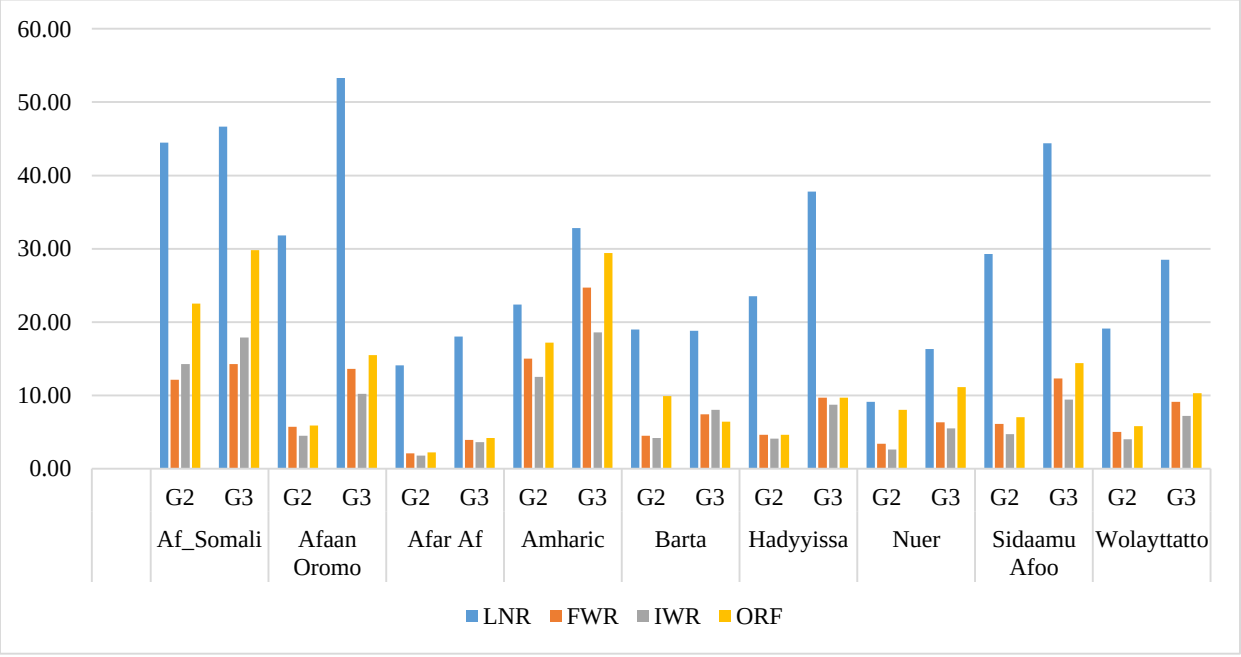


Figure 1. Mean differences of students' performance by grade level in timed subtasks

3.3.2 The mean differences of students' performance by gender in timed subtasks

Table 7 and Figure 2 show the significant mean differences in students' performance by gender in timed subtasks for each language group. In this regard, to analyze the findings of testing for statistically significant differences across gender, the independent sample t-test was used and for the practical differences between males and females Cohen's d was applied, as detailed in Appendix B.

The mean differences between males and females are statistically significant in all EGRA timed subtasks. Males slightly performed better than females in most language categories and timed subtasks. However, females performed better than males in familiar words reading and, invented words reading subtasks in Af Somali. Likewise in Berta, female students performed somewhat better than males in familiar words reading.

In this regard, in Afar Af, the effect sizes disparities by gender for oral reading fluency task, as measured by Cohen's d, the value of 0.71 can be classified as a strong educational impact that was manifested as something was significantly improved. The Cohen's d values 0.29, 0.35, and 0.27 in Afan Oromo, Sidaamu Afoo, and Wolayttatto, respectively, indicating that the degree of difference between males and females is educationally significant and showed that something was taught. On the other hand, even though there was a statistically significant difference between males and females for the remaining five language groups, the effect sizes between males and females were negligible, according to the values of Cohen's d.

Table 7. The mean differences of students' performance by gender in timed subtasks

Language	Gender	Letter Name Recognition	Familiar Words Reading	Invented Words Reading	Oral Reading Fluency	Cohen's d For ORF
Af Somali	Female	44.1	13.6	16.5	24.6	-0.12
	Male	47.1	12.8	15.6	27.6	
Afan Oromo	Female	39.1	7.8	6.0	8.3	-0.29
	Male	46.0	11.4	8.6	13.0	
Afar Af	Female	12.1	1.6	1.6	1.9	-0.71
	Male	18.8	4.0	3.4	4.1	
Amharic	Female	27.1	19.7	15.3	23.2	-0.01
	Male	28.0	19.9	15.8	23.4	
Berta	Female	18.5	7.8	5.0	5.0	0.15
	Male	19.2	4.4	7.1	11.8	
Hadiyyissa	Female	27.2	6.1	5.6	6.3	-0.11
	Male	34.1	8.2	7.2	8.0	
Nuer	Female	9.0	3.6	2.9	8.1	-0.11
	Male	15.6	5.9	4.9	10.6	
Sidaamu Afoo	Female	31.0	6.7	5.3	7.9	-0.35
	Male	43.0	11.8	8.9	13.7	
Wolayttatto	Female	20.0	4.6	3.8	5.5	-0.27
	Male	27.3	9.3	7.3	10.4	

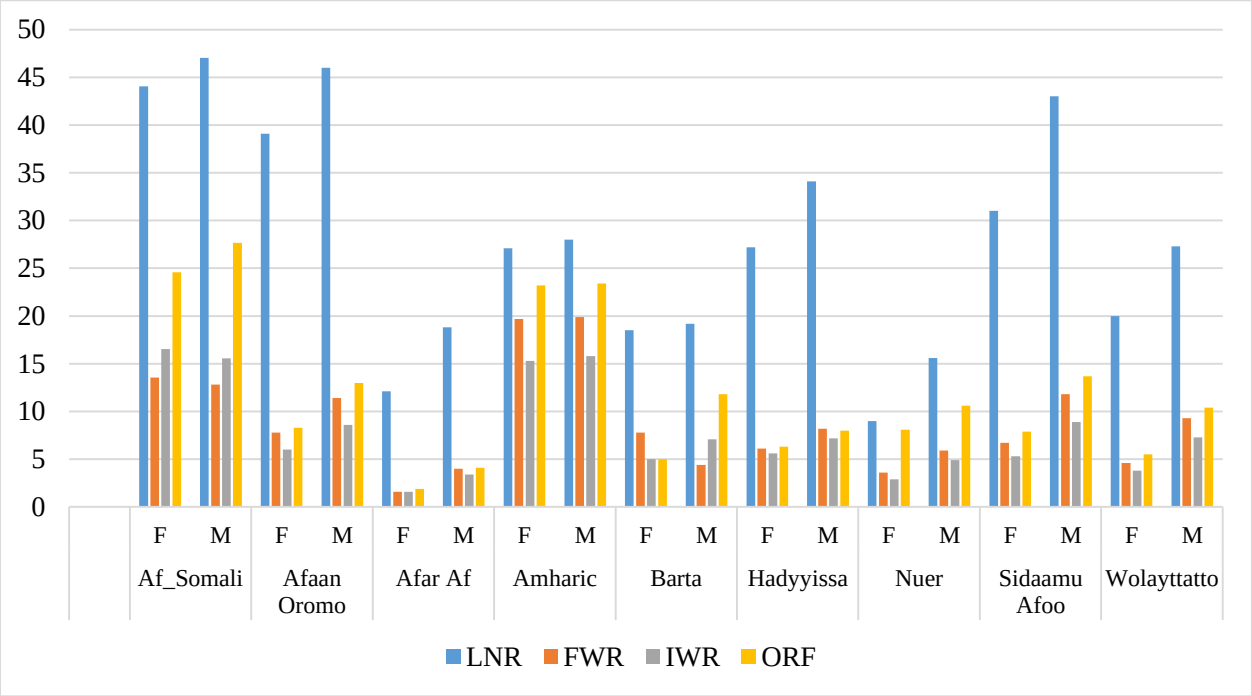


Figure 2. Mean differences of students' performance by gender in timed subtasks

3.3.4 The mean differences of students' performance by location in timed subtasks

For each language group, Table 8 and Figure 3 show the significant mean differences in students' performances by location in timed subtasks. In this regard, the independent sample t-test was used to examine the findings of testing for statistically significant differences between urban and rural students, and Cohen's d was used to analyze the practical differences between the locations, as indicated in Appendix C.

In all timed subtasks for each language group, the mean differences between urban and rural students are statistically significant. In particular, urban students performed better than rural students in all timed subtasks for Afan Oromo, Afar Af, Amharic, Berta, Hadiyyissa, and Wolayttatto languages, with the exception of letter recognition in both Hadiyyissa and Wolayttatto languages as well as invented words reading in Afan Oromo language. Nevertheless, rural students outperformed urban students in all subtasks in Af Somali, Nuer, and Sidaamu Afoo, with the exception of letter name recognition and oral reading fluency for Af Somali and Nuer respectively, where urban students outperformed rural students.

Furthermore, in Amharic language, the effect sizes differences by location for oral reading fluency task, as measured by Cohen's d, the value of 0.65 can be justified as a strong educational impact that was revealed as something was substantially improved. However, despite the fact that there is a statistically significant difference between urban and rural students in all other language groups for each subtask and the effect sizes are very small, according to Cohen's d values.

Table 8. The mean differences of students' performance by location in timed subtasks

Language		Letter Name Recognition	Familiar Words Reading	Invented Words Reading	Oral Reading Fluency	Cohen's d For ORF
Af Somali	Rural	44.6	18.5	17.2	27.2	0.03
	Urban	46.5	13.8	14.9	25.0	
Afan Oromo	Rural	42.4	9.6	7.4	10.7	-0.10
	Urban	46.6	11.2	7.3	12.3	
Afar Af	Rural	15.5	2.8	2.5	2.9	-0.22
	Urban	19.8	4.3	3.8	5.6	
Amharic	Rural	24.8	18.0	14.3	21.1	-0.65
	Urban	41.5	29.0	22.0	34.8	
Berta	Rural	14.8	4.8	4.2	5.8	-0.18
	Urban	29.0	9.0	11.0	13.9	
Hadiyyissa	Rural	31.4	6.9	5.9	6.6	-0.18
	Urban	28.6	8.3	8.3	9.4	
Nuer	Rural	13.6	5.1	4.5	8.7	-0.10
	Urban	11.0	4.4	3.2	10.8	
Sidaamu Afoo	Rural	37.6	9.3	7.1	10.9	0.06
	Urban	31.6	8.6	6.9	9.8	
Wolayttatto	Rural	24.0	6.9	5.3	7.7	-0.11
	Urban	22.3	7.6	6.9	9.7	

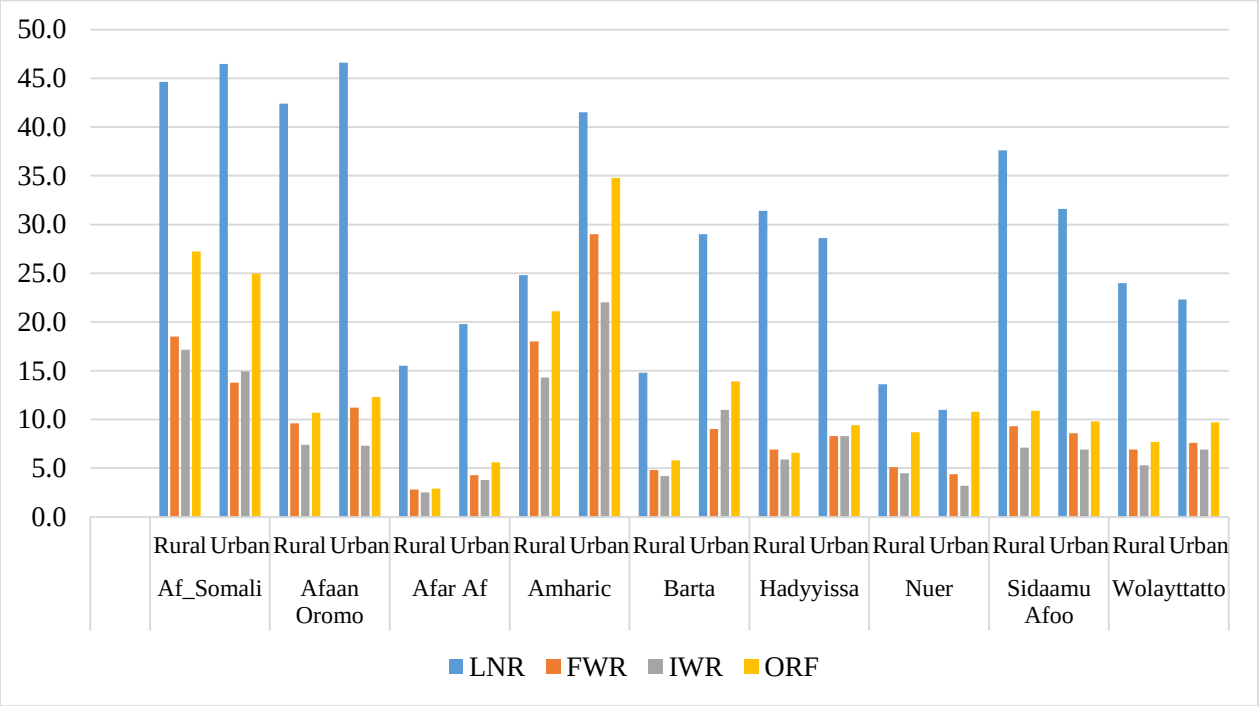


Figure 3 Mean differences of students' performance by location in timed subtasks

3.3.5 The mean differences of students' performance by grade in untimed subtasks

Table 9 and Figure 4 indicate the significant mean differences of students' performance by grade level for each language groups in untimed subtasks. In this case, the results of testing for statistically significant differences between grades (by independent sample t-tests) and practical differences between grades (by Cohen's d) were applied as shown in Appendix D in details.

In each language group, the mean differences between the two grade levels in all untimed subtasks are statistically significant and in favor of Grade 3. This means that in all languages, grade 3 students were able to perform better than Grade 2 students, which indicates positive grade gain. Particularly, the effect sizes of grade differences for reading comprehension subtask as measured by Cohen's d, are higher than 0.25 for most languages which indicate that the level of the change by grade is educationally significant. Even though, there is a statistically significant difference between grade levels for Berta and Wolayttatto, there is a small effect size.

Table 9. Mean differences of students' performance by grade in untimed subtasks

Language	Grade	Phonemic Awareness	Listening Comprehension	Reading Comprehension	Cohen's d For RC
Af Somali	Grade 2	77.0	66.2	19.2	-0.32
	Grade 3	73.8	74.7	29.9	
Afan Oromo	Grade 2	39.1	64.7	6.7	-0.52
	Grade 3	63.0	75.5	19.3	
Afar Af	Grade 2	15.8	50.9	1.1	-0.26
	Grade 3	23.6	56.5	3.5	
Amharic	Grade 2	72.6	69.3	16.4	-0.54
	Grade 3	80.1	73.8	31.1	
Berta	Grade 2	31.2	55.8	2.1	-0.04
	Grade 3	40.0	62.2	2.6	
Hadiyyissa	Grade 2	73.7	74.2	4.8	-0.28
	Grade 3	78.6	81.5	8.9	
Nuer	Grade 2	30.3	76.8	5.5	-0.44
	Grade 3	54.6	80.5	14.6	
Sidaamu Afoo	Grade 2	82.0	75.8	6.2	-0.40
	Grade 3	85.6	82.5	13.8	
Wolayttatto	Grade 2	60.0	47.1	6.3	-0.22
	Grade 3	69.4	50.0	11.2	

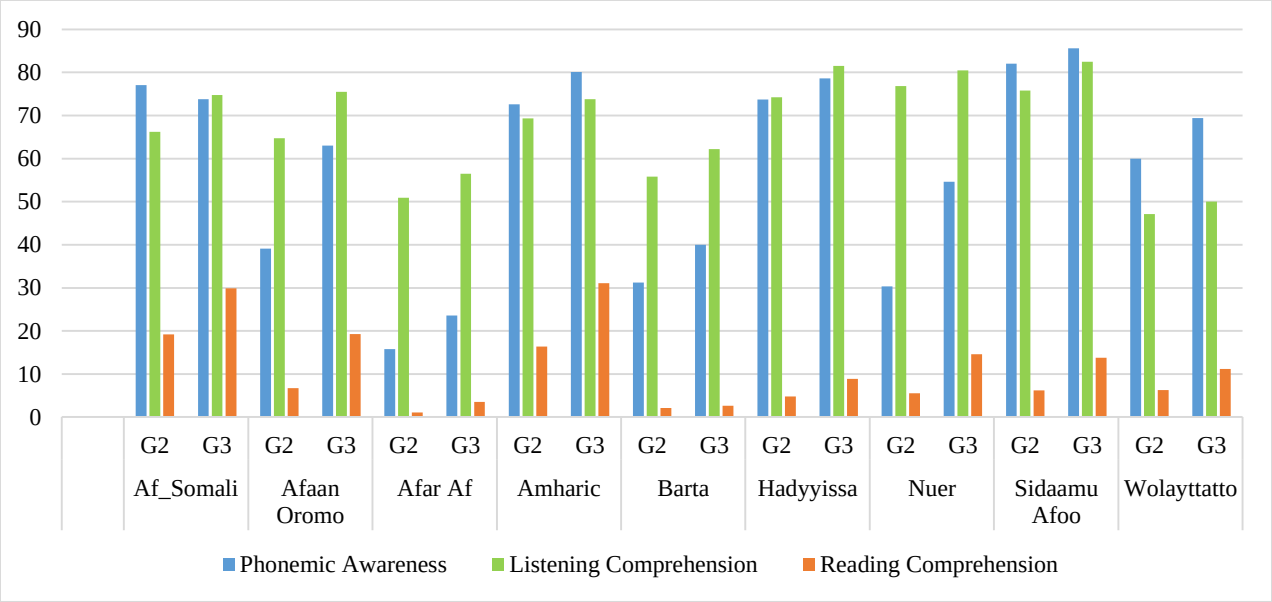


Figure 4 Mean differences of students' performance by grade in untimed subtasks

3.3.6 The mean differences of students' performance by gender in untimed subtasks

Table 10 and Figure 5 show the significant mean differences in students' performance by gender in untimed subtasks for each language group. The independent sample t-test was used for the analysis of statistically significant differences across gender and for the practical differences Cohen's d was applied (Appendix E).

The mean differences between males and females are statistically significant in all untimed subtasks except for Berta in listening comprehension. In all language groups, males performed better than females, though, females performed slightly better than males in reading comprehension for Amharic. Looking at the effect sizes disparities by gender for reading comprehension, as measured by Cohen's d, the values of 0.29, 0.28, and 0.28 in Afan Oromo, Sidaamu Afoo, and Wolayttatto, respectively, indicating that the degree of differences between males and females is educationally significant and showed that something was taught. Even though, there was a statistically significant difference between males and females for the remaining five languages, the effect sizes were found very negligible.

Table 10. The mean differences in students' performance by gender in untimed subtasks

Language	Gender	Phonemic Awareness	Listening Comprehension	Reading Comprehension	Cohen's d For RC
Af Somali	Female	74.3	66.6	20.5	-0.20
	Male	76.6	74.3	28.5	
Afan Oromo	Female	44.5	68.0	9.3	-0.29
	Male	57.2	72.1	16.4	
Afar Af	Female	15.7	50.9	1.6	-0.13
	Male	22.2	55.6	2.7	
Amharic	Female	75.3	70.4	24.2	0.03
	Male	77.4	72.6	23.3	
Berta	Female	37.2	59.0	2.3	-0.01
	Male	34.4	59.2	2.4	
Hadiyyissa	Female	74.1	76.6	6.0	-0.11
	Male	78.2	79.1	7.7	
Nuer	Female	35.1	76.8	6.0	-0.12
	Male	47.8	80.0	13.1	
Sidaamu Afoo	Female	82.7	77.6	7.4	-0.28
	Male	84.9	80.7	12.7	
Wolayttatto	Female	63.0	47.6	5.4	-0.28
	Male	66.1	49.4	11.7	

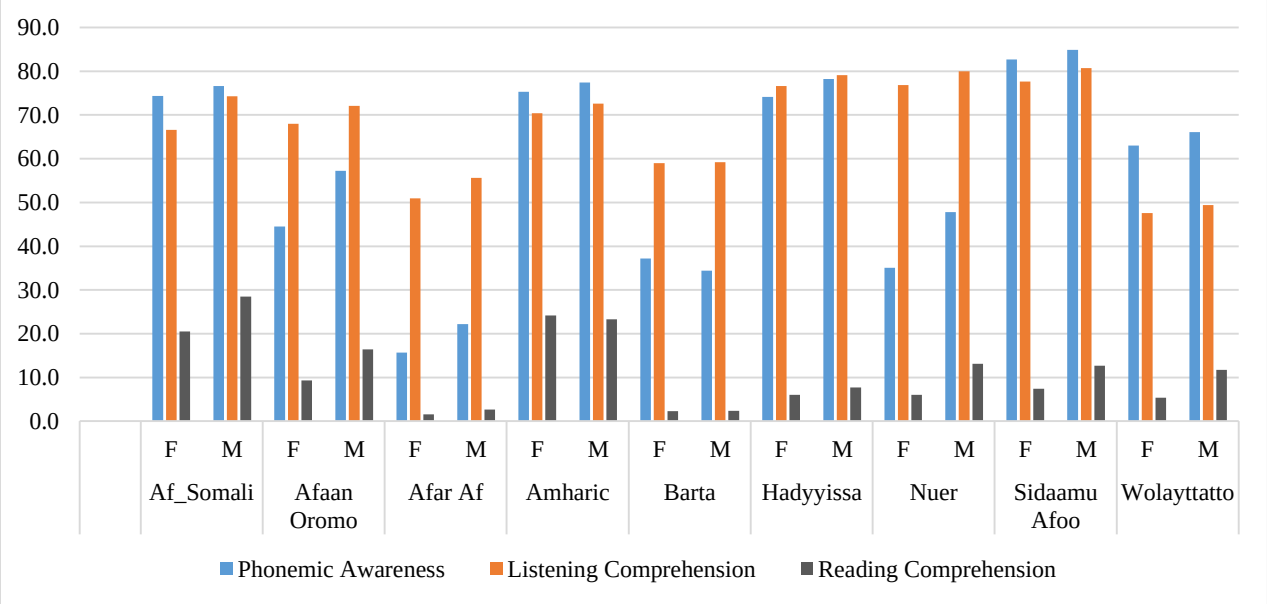


Figure 5 Mean differences in students' performance by gender in untimed subtasks

3.3.7 The mean differences of students' performance by location in untimed subtasks

For each language group, Table 11 and Figure 6 show the significant mean differences in students' performance by location in untimed subtasks. In this regard, the independent sample t-test was used to examine the statistically significant differences between urban and rural students, and Cohen's d was used to analyze the practical differences between the locations, as indicated in Appendix F.

In all untimed subtasks for each language group, the mean differences between urban and rural students are statistically significant except for Berta in listening comprehension. For most of language groups, urban students slightly performed better than rural students across the three untimed subtasks. Nevertheless, rural students outperformed urban students in phonemic awareness for Afan Oromo, Nuer, and Hadiyyissa; in listening comprehension for Af Somali, Berta and Hadiyyissa as well as in reading comprehension for Af Somali. Furthermore, despite the fact that there was a statistically significant difference between urban and rural students in all other language groups, the effect sizes were very small, according to Cohen's d values except for Amharic 0.63 which indicates as educationally something was taught significantly.

Table 11. The mean differences in students' performance by location in untimed subtasks

Language		Phonemic Awareness	Listening Comprehension	Reading Comprehension	Cohen's d For RC
Af Somali	Rural	70.3	72.1	26.6	0.14
	Urban	80.6	68.7	22.3	
Afan Oromo	Rural	51.6	69.9	12.9	-0.09
	Urban	49.6	72.9	15.2	
Afar Af	Rural	19.3	53.6	2.1	-0.18
	Urban	21.1	53.7	3.7	
Amharic	Rural	74.3	69.4	20.9	-0.63
	Urban	86.8	82.4	38.3	
Berta	Rural	32.5	59.1	2.1	-0.09
	Urban	43.7	59.0	3.2	
Hadiyyissa	Rural	77.9	78.9	6.6	-0.09
	Urban	70.2	74.1	8.0	
Nuer	Rural	46.8	74.1	9.5	-0.05
	Urban	34.2	86.0	10.5	
Sidaamu Afoo	Rural	83.6	78.9	10.0	-0.02
	Urban	85.2	81.5	10.5	
Wolayttatto	Rural	63.6	47.8	8.1	-0.14
	Urban	69.6	51.9	11.4	

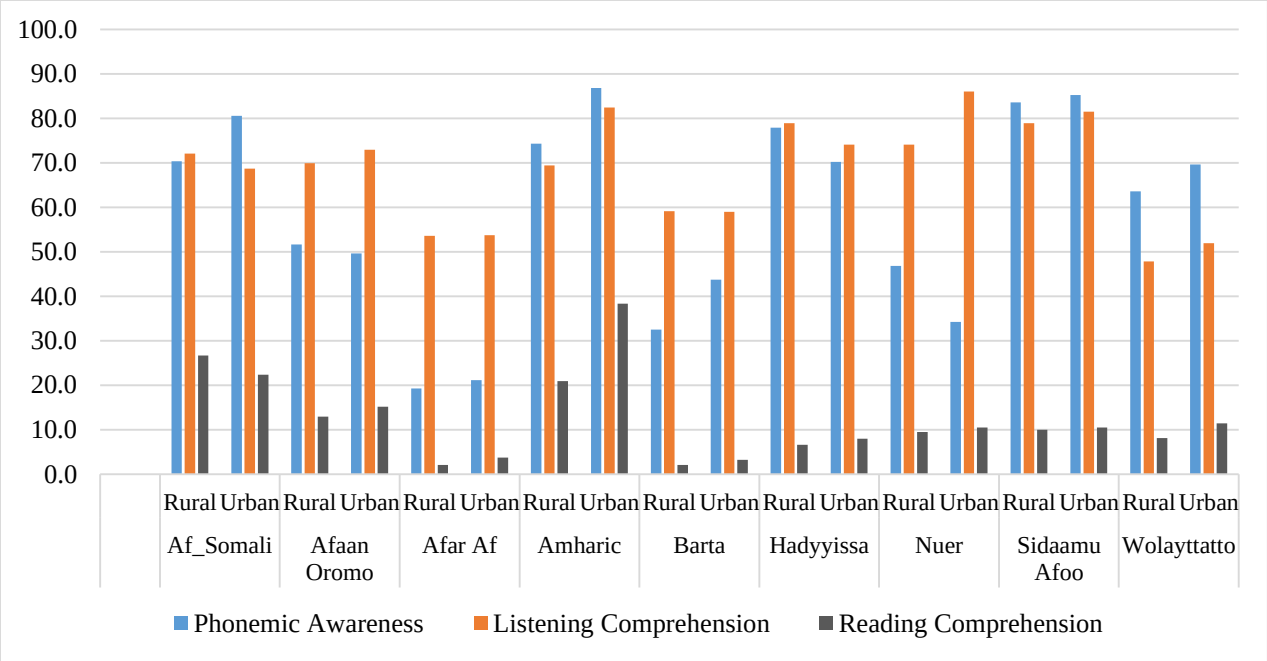


Figure 6 Mean differences in students' performance by location in untimed subtasks

3.4 Overall performance of students' by age group

Table 12 and figure 7 indicate the overall performance of students' in each subtask by age group. In this aspect, the students of ages more than years old performed better than students of ages less than nine years old in all subtasks. That means, as age increased the performance of students in reading skills also increased. However, the eta squared values varied from 0.009 to 0.011 for timed subtasks and 0.012 to 0.018 for untimed subtasks, indicating a weak difference in effect size among age groups.

Table 12. Performance of students' in each subtask by age group

Subtask		Age of the Students				Eta Squared
		< 8 years	8 years	9 years	> 9 years	
Timed	Letter Name Recognition	17.95	28.55	32.38	34.55	0.009
	Familiar Words Reading	5.48	8.48	10.60	12.51	0.010
	Invented Words Reading	3.71	6.54	8.73	9.96	0.008
	Oral Reading Fluency	8.23	9.31	12.30	15.00	0.011
Untimed	Phonemic Awareness	37.51	54.16	59.41	65.20	0.018
	Reading Comprehension	6.57	9.30	12.45	16.24	0.013
	Listening Comprehension	56.02	64.03	68.17	71.37	0.012

Eta squared ranges between 0 and 1 and it is interpreted as 0-0.1 is a weak effect, 0.1-0.3 is a modest effect, 0.3-0.5 is a moderate effect and >0.5 is a strong effect (Daniel, 2004).

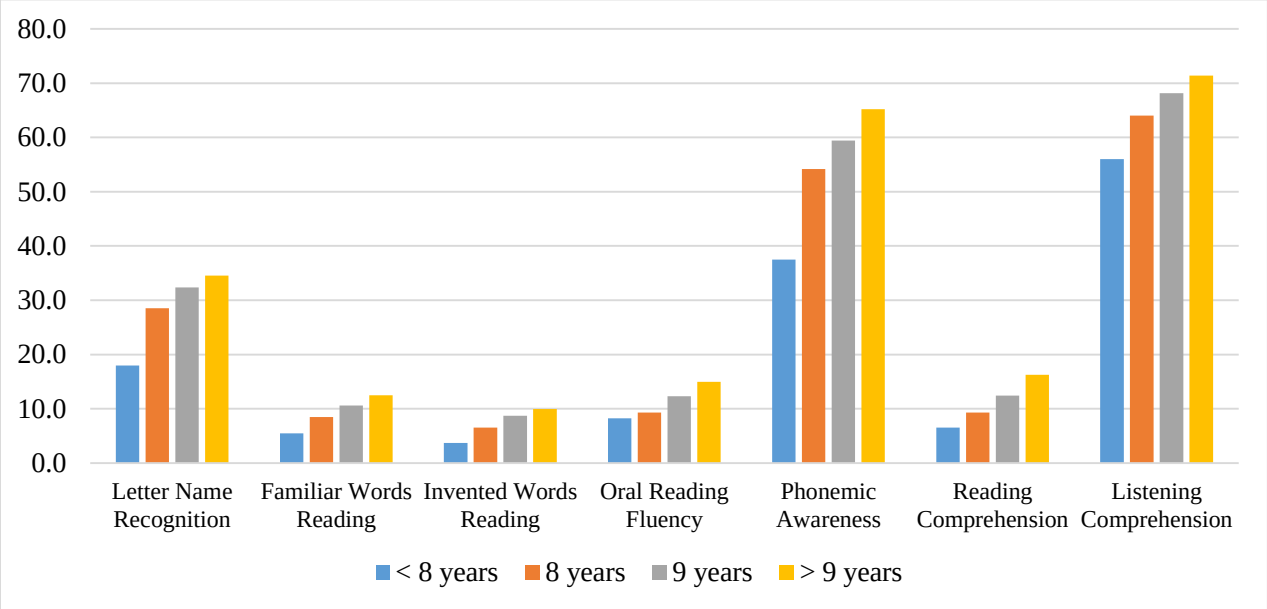


Figure 7 Students' performance in each subtask by age group

3.5 Students' performance by program type

The overall performances of students' in EGRA timed and untimed subtasks by the types of program is shown in Table 13 and Figure 8. In this regard, in all EGRA subtasks except in listening comprehension, students from GEQIP phase I sample schools performed slightly better than students from both NEGRA and READ II. On the other hand, students from READ II schools performed slightly better than students from NEGRA in all timed subtasks except in oral reading fluency. However, in untimed subtasks, students from NEGRA schools slightly performed better than students from READ II in reading comprehension and listening comprehension except in phonemic awareness. Furthermore, the eta squared values for both timed and untimed subtasks ranged from 0.000 to 0.005, revealing a small difference in effect size of students' performance among the program types.

Table 13. Overall students' mean score by the types of program in EGRA subtasks

Subtasks		Type of Program			Total	Eta Squared
		GEQIP Phase I	NEGRA	READ II		
Timed	Letter Name Recognition	37.4	31.5	31.8	32.7	0.005
	Familiar Words Reading	12.6	10.7	11.5	11.2	0.002
	Invented Words Reading	10.1	8.7	8.8	9.0	0.001
	Oral Reading Fluency	14.0	13.1	13.0	13.3	0.000
Untimed	Phonemic Awareness	65.8	59.3	64.6	61.4	0.005
	Reading Comprehension	15.3	13.7	13.5	14.0	0.001
	Listening Comprehension	68.3	69.3	69.2	69.1	0.000

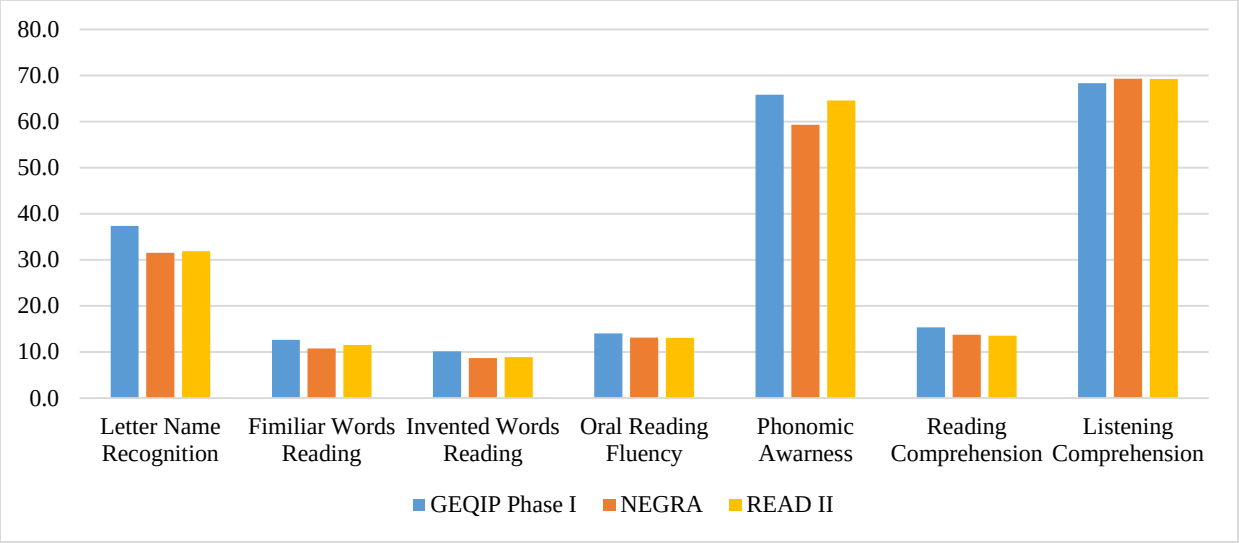


Figure 8 Students' performance in means scores for each subtask by program types

Table 14 shows the performances of students in timed subtasks by program types and gender. In Grade 2 and 3, male students performed better than female students in each program types (GEQIP Phase I, NEGRA, and READ II) and in each EGRA timed subtasks. From the findings, it can be suggested that all types of programs should be given special attention to improve the reading abilities of girls.

Table 14. Performance of students' by types of program across gender in timed tasks

Grade	Program	Sex	Letter name recognition	Familiar words reading	Invented words reading	Oral reading fluency
Grade 2	GEQIP Phase I	Female	24.0	6.9	5.4	7.4
		Male	31.2	9.6	7.7	10.6
		Total	27.7	8.3	6.6	9.1
	NEGRA	Female	22.8	7.3	6.0	9.5
		Male	27.3	8.1	7.0	9.8
		Total	25.2	7.7	6.5	9.7
	READ II	Female	22.2	6.6	5.4	7.1
		Male	25.6	8.1	6.4	8.8
		Total	24.0	7.4	5.9	8.0
Grade 3	GEQIP Phase I	Female	44.4	15.6	12.4	16.8
		Male	49.6	18.4	14.7	21.1
		Total	47.1	17.0	13.6	19.0
	NEGRA	Female	34.5	12.1	9.4	14.8
		Male	40.5	15.0	11.9	17.9
		Total	37.7	13.7	10.8	16.5
	READ II	Female	35.9	13.8	10.4	15.5
		Male	43.1	17.3	12.9	20.3
		Total	39.7	15.6	11.7	18.0

Table 15 indicates students' performance in untimed tasks by program type across gender of the students. In this case, in both grade levels, male students performed better than female in each EGRA untimed subtask and program types (GEQIP Phase I, NEGRA, and READ II). This shows that each of the programs should be given due attention in order to boost females' reading performances.

Table 15 Performance of students' by types of program across gender in untimed subtasks

Grade Levels	Program	Sex	Phonemic awareness	Reading comprehension	Listening comprehension
Grade 2	GEQIP Phase I	Female	55.2	6.4	63.4
		Male	64.7	11.6	65.5
		Total	60.0	9.1	64.5
	NEGRA	Female	50.1	8.5	63.8
		Male	53.6	9.6	67.0
		Total	52.0	9.1	65.5
	READ II	Female	53.6	6.1	65.0
		Male	58.7	8.6	67.3
		Total	56.3	7.4	66.2
Grade 3	GEQIP Phase I	Female	66.8	18.1	71.2
		Male	76.0	24.8	73.1
		Total	71.5	21.5	72.2
	NEGRA	Female	62.7	15.2	70.8
		Male	69.4	20.8	74.8
		Total	66.3	18.2	72.9
	READ II	Female	69.0	16.8	71.1
		Male	76.3	22.3	73.3
		Total	72.8	19.7	72.2

3.6 Overall percentage of students at benchmark levels by language

The percentages of students who reached each proficiency level analyzed for each language and grade level. The analysis is based on the benchmarks set by READ M&E in 2015 for seven languages and three additional languages (Afar Af, Berta, and Nuer) on October 2021 by NEAEA as indicated in Appendix H. The benchmark levels were classified as Level 1: Full proficiency, Level 2: Increasing proficiency Level 3: Limited proficiency and Level 4: Non -readers. In other words, students are categorized as readers who read fluently and with full comprehension, readers with increasing fluency and comprehension, readers with limited fluency and comprehension and zero readers under these four benchmarks of oral reading fluency subtask. In this case, policy makers and stakeholders can use the information from these benchmark levels to support schools as well as monitor and evaluate the progress of students' performance in reading.

Table 16 and Figure 9 present the percentage of Grade 2 students at each benchmark levels by language groups. In this regard, at the national level, 68% of the students are put under the category of zero readers in oral reading fluency subtask. In this case, Afar Af and Berta languages had the highest percentage of zero readers (about 92%), while Amharic (34%) and Af Somali (35%) had the lowest percentage. The percentage of zero readers in other languages ranged from 63% in Afan Oromo to 78% in Hadiyyissa. On the other hand, insignificant percentage of student in each language group achieved the highest benchmark (Level 1), which reflects "reading fluently with full or almost full comprehension." In this regard, relatively the highest percentage (20%) is from Af Somali and the least percentage (1%) is from Afar Af in "read fluently and with full comprehension" benchmark. At national level, the percentage of students who reached level 1 benchmark is only 5 %.

With regard to level 2, which refers to "reading with increasing fluency and comprehension", the percentages of students attaining this level vary across languages from zero in Berta to 19% in Af Somali. Based on the findings, an average of only 8% of Grade 2 students in Ethiopia are attaining reading proficiency at the "increasing" level. For the monitoring purposes, the two highest levels referring to "full or almost full" and "increasing" reading proficiency may be combined to render a category of reading proficiency that could be tentatively qualified as acceptable or described as readers who can attain either full or partial reading comprehension leveled as *functional reading level*. Overall, at the national level, the percentage of students at the functional reading level or level 1 and level 2 is only about 13%. This indicates that, the percentage of students who reached the functional reading level is dropped almost by half (12 %) as compared to 2018 EGRA (25%).

Table 16. Grade 2 overall percentage of students at benchmark levels by language

Language	Zero Readers	Read with limited fluency and comprehension	Read with increasing fluency and comprehension	Read fluently and with full comprehension
Afan Oromo	63	26	9	2
Afar Af	92	4	4	1
Af Somali	35	32	13	20
Amharic	34	41	19	6
Berta	92	4	0	4
Hadiyyissa	78	16	3	3
Nuer	71	18	5	5
Sadaamu Afoo	70	15	13	3
Wolayttatto	76	16	4	5
National	**68**	**19**	**8**	**5**

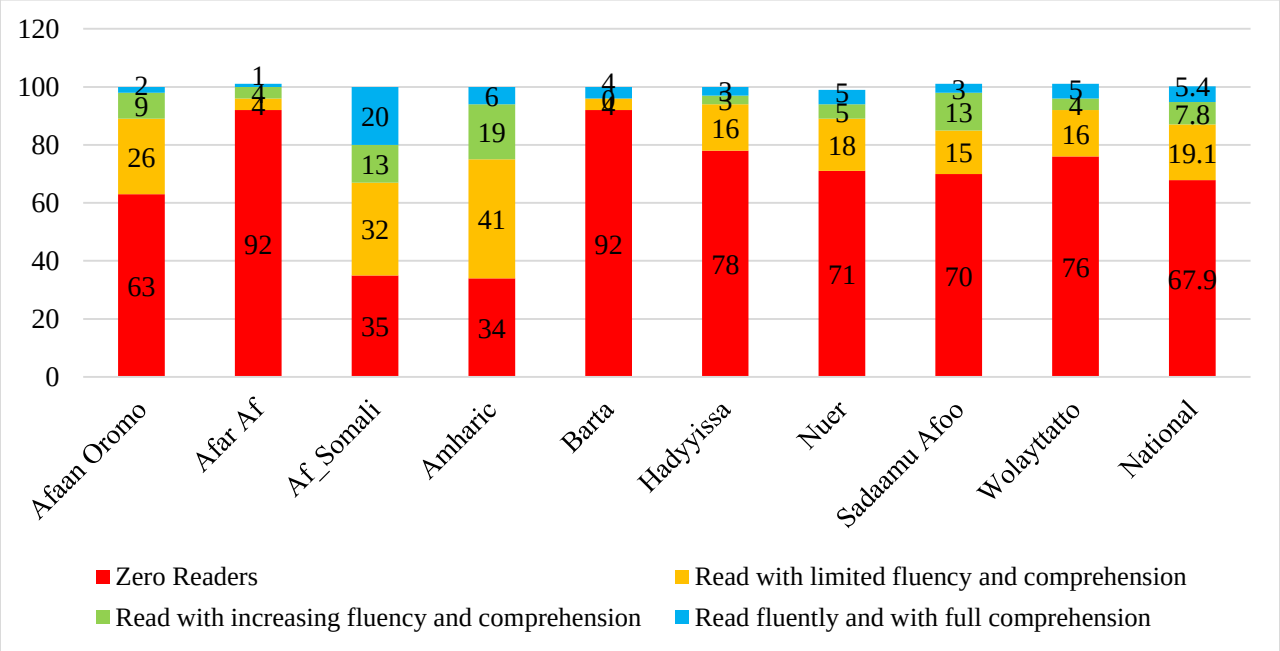


Figure 9 Percentage of grade 2 students at benchmark levels by language

Table 17 and Figure 10 show the percentage of Grade 3 students in each language group at each benchmark level. As the result shows, at the national level, 51% of the students were categorized under zero readers in oral reading fluency task. Like that of Grade 2, Afar Af and Berta had the highest percentage of zero reader students with 85% and 91% respectively, while Amharic (18%) and Af Somali (20%) had the lowest percentage. The percentage of zero readers in other languages ranged from 36% in Afan Oromo to 58% in Wolayttatto. On the other hand, a few percent of students in each language group achieved the highest benchmark (Level 1). In this regard, relatively the highest percentage (22%) is from Af Somali and the least percentage (1%) is from Afar Af and Berta. Regarding reading with increasing fluency and comprehension level, the percentages of students reaching this level vary across languages from 3% in Berta to 36% in Amharic. As depicted in the findings, only 13% of Grade 3 students in Ethiopia are attaining reading proficiency at the "increasing" level.

For the monitoring purposes, the two highest levels referring to "full or almost full" and "increasing" reading proficiency may be combined to render a category of reading proficiency that could be tentatively qualified as acceptable or described as readers who can attain *functional reading level*. Overall, at the national level, the percentage of students at the functional reading level is only 23%. This indicates that, the percentage of students reaching this level is dropped almost by 17 % as compared to 2018 EGRA (40%).

Table 17. Grade 3 overall percentage of students at benchmark levels by language

Language	Zero Readers	Read with limited fluency and comprehension	Read with increasing fluency and comprehension	Read fluently and with full comprehension
Afan Oromo	36	43	18	3
Afar Af	85	8	6	1
Af Somali	20	29	29	22
Amharic	18	38	36	7
Berta	91	4	3	3
Hadiyyissa	56	29	14	1
Nuer	49	37	8	5
Sidaamu Afoo	47	27	22	5
Wolayttatto	58	26	13	3
National	**51**	**27**	**17**	**6**

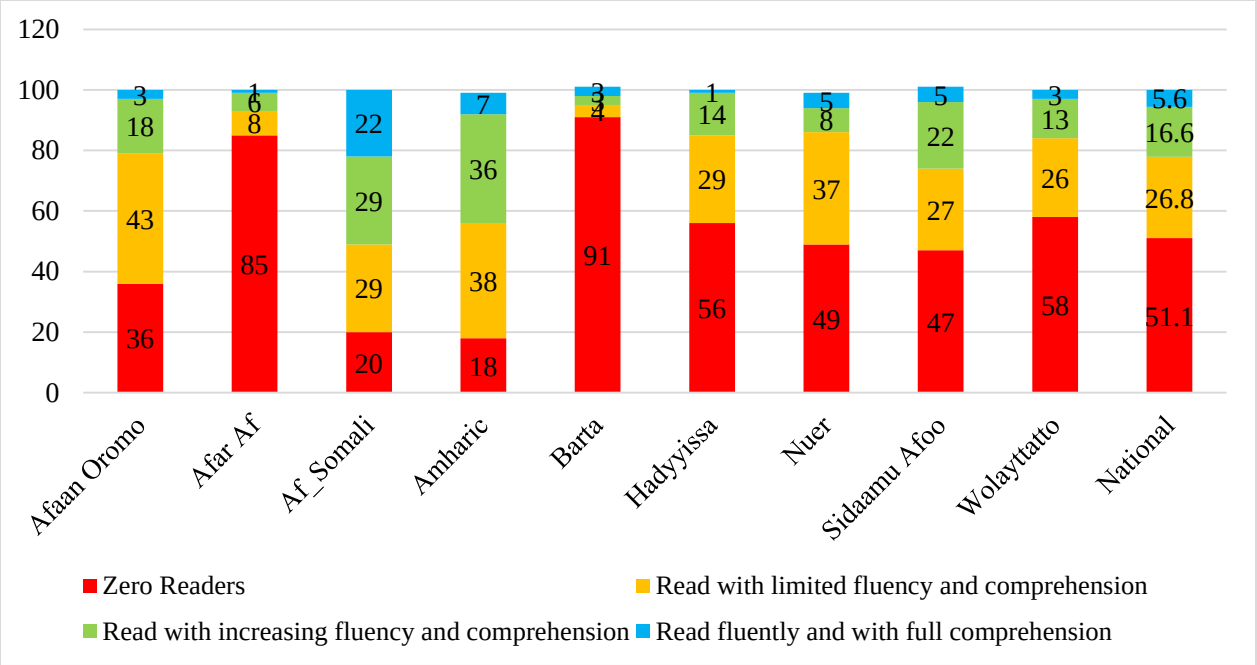


Figure 10 Grade 3 overall percentage of students at benchmark levels by language

3.7 Comparison across years in attaining benchmark levels

This section attempts to compare performances of students in EGRA across four administration years: 2014 (baseline), 2016 (midline), and 2018 (end-line) conducted by the READ M & E project, and EGRA 2021 conducted by the NEAEA. However, due to some reasons, the comparability of the results obtained in the last three EGRA assessments and the EGRA administered by NEAEA require certain cautions. The first reason is that the Tigrigna language was excluded from this study because of regional unrest. As a result, students from Tigray, the second highest scorers in oral reading fluency in 2018 after those from Amharic language, were excluded from EGRA 2021. Secondly, this EGRA study included additional students from the 'emerging' regions who had been performing lower in every national assessment for the last two decades. Thus, these issues may have an impact while comparing the results of EGRA 2021 study to those of previous EGRA studies, albeit, the instruments for data collection were the same. Therefore, readers should take into account our reservations about the comparability of Ethiopian students' oral reading performances throughout the four administration years.

Table 18 and Figure 11 show the percentages of students at each benchmark level by years and grade levels to represent the overall students' performances. When comparing grades within each EGRA administration years, the percentages of students reaching the functional reading levels is higher for Grade 3 than Grade 2, even though the cut score of Grade 3 is also higher due to the expectation of differences between the two grade levels.

When comparing EGRA 2021's students' oral reading performances to that of earlier EGRA administrations, there is a significant decrease in the percentages of students achieving full or increasing reading proficiency. For example, in 2021, the percentage of students' at the functional reading level are 13% and 23 % for Grades 2 and 3 respectively, that were found lower than earlier EGRA studies. On the contrary, the percentages of zero readers in 2021 is higher than previous EGRA findings for both grade levels. In this regard, the performance of students in each language was aggregated for the purpose of comparisons across years at the national level. The cut scores were specifically designed to reflect unique characters of each language, despite the fact, the EGRA benchmarks levels have the same meaning in each language. As a result, the aggregated percentages at each benchmark level can be used as measures of overall students' performances at national levels.

Table 18. Over all percentage of students at benchmark levels by years and grade levels

Benchmarks	2014		2016		2018		2021	
	Gr.2	Gr.3	Gr.2	Gr.3	Gr.2	Gr.3	Gr.2	Gr.3
Zero Readers	50.3	34.0	39.9	24.2	46.7	27.7	67.9	51.1
Read with limited fluency and comprehension	24.8	28.5	32.1	35.3	28.3	32.4	19.1	26.8
Read with increasing fluency and comprehension	19	31.5	22.7	32.2	20.8	31.8	7.8	16.6
Read fluently and with full comprehension	5.3	6.8	5.3	8.3	4.2	8.1	5.4	5.6

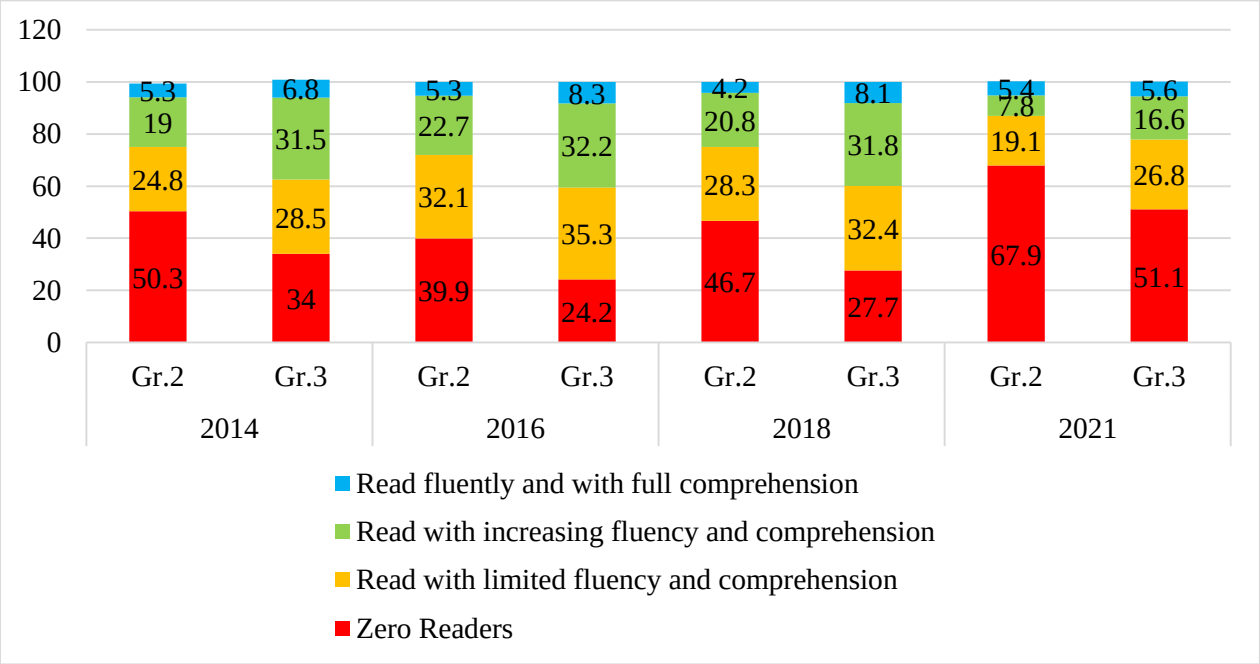


Figure 11 Overall percentage of students at benchmark levels by year and grade

The percentage of students meeting each benchmark level was aggregated across grades to provide direct insights into the overall trend in reading performances of students in Ethiopia over time (Table 19 and Figure 12). Over the last three EGRA administrations, there has been a small variance in students' reading performances. In 2014, about 31% of students met the two upper reading benchmark levels; whereas in 2016, it increased to 34.2%; and in 2018, it decreased to 32.5%. As a result, fluency and comprehension levels remained relatively stable over these time periods, as the deviations are negligible and have little practical effect.

On the other hand, when compared to previous EGRA findings, the percentage of students at the functional reading level in EGRA 2021 declined to 17.7%, about half the percentage recorded in 2016 and the percentage of zero readers raised to about 60%. The most likely explanation for the low performances of students in current EGRA might be the impact of the COVID-19 pandemic on schooling, internal conflict and instability in some parts of the country, the exclusion of students from Tigray region, and the inclusion of additional three languages (Afar Af, Berta and Nuer) from emerging regions. However, reconsidering all these problems, the findings this study revealed that Ethiopian students' reading performances is still in jeopardy. This presupposes that a profound foundation should be laid for an optimal and effective reading interventions.

Table 19. Overall percentage of students at benchmark levels by year

Benchmarks	Administration Years			
	2014	2016	2018	2021
Zero Readers	42.2	32.1	37.2	59.5
Read with limited fluency and comprehension	26.7	33.7	30.4	23.0
Read with increasing fluency and comprehension	25.3	27.4	26.3	12.2
Read fluently and with full comprehension	6.1	6.8	6.2	5.5

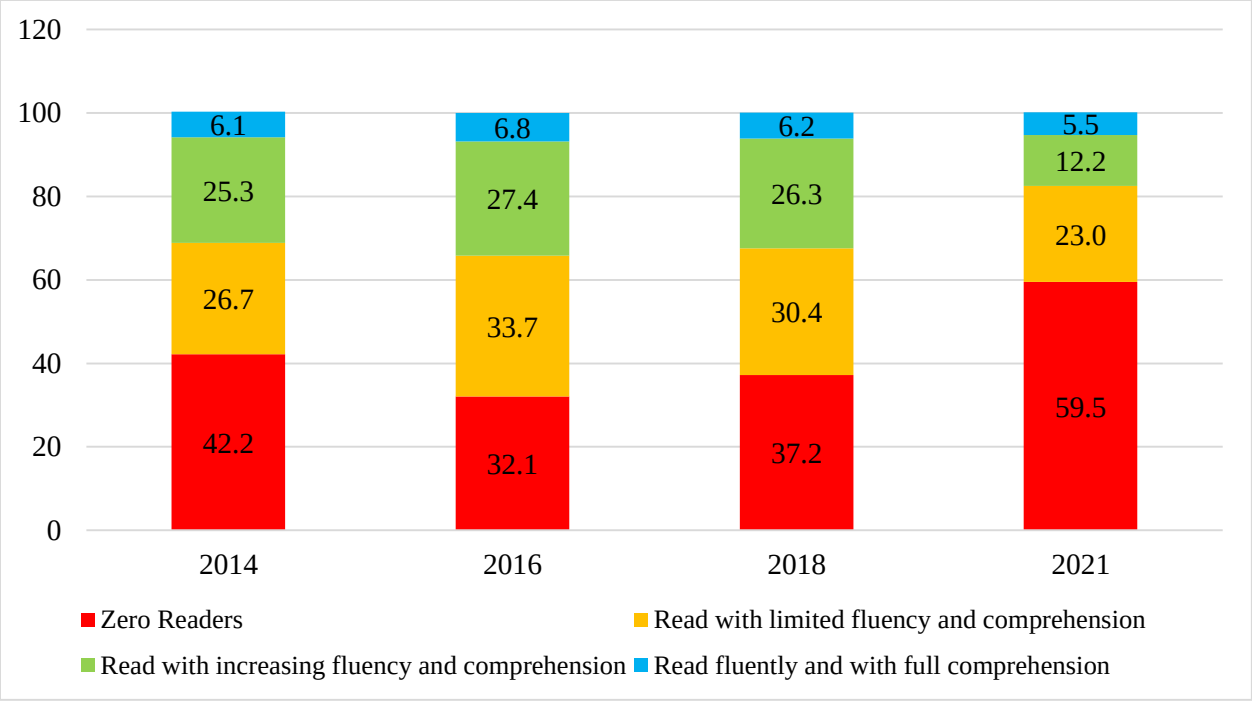


Figure 12 Percentage of students at benchmark levels by year

Table 20 and Figure 13 display, the comparison of GEQIP Phase I schools at benchmark levels across years. In this case, students' from Tigray and the three languages (Afar Af, Berta, and Nuer) were not included as students from Tigray did not participate in the EGRA 2021 and those from additional three languages were not part of the previous EGRA studies.

In this regard, in 2018, overall 30.8% of students met the two upper reading benchmark levels (functional reading level), however, in 2021, the percentage of students dropped to 23%. On the contrary, the percentage of zero readers increased by about 10% from 39% in 2018 to 49.4% in 2021. When comparing the oral reading fluency performances of each Grade 2 and 3 students', they performed lower in EGRA 2021 than 2018. In this case, percentages of Grade 2 students attained at functional proficiency levels is 16.9%, which was declined by 7.4% as compared to EGRA 2018 (24.3%). Similarly, in Grade 3, the percentage of students attained at functional proficiency level in EGRA 2021 is 29.2% that was dropped by 8.1% when compared to EGRA 2018 (37.3%).

Table 20. Percentage of students at benchmark levels by year in GEQIP Phase I schools

Benchmark Levels	Administration Year					
	2018			2021		
	Grade 2	Grade 3	Total	Grade 2	Grade 3	Total
Zero Readers	47.6	30.5	39.0	59.6	39.1	49.4
Read with limited fluency and comprehension	28.1	32.2	30.2	23.6	31.7	27.7
Read with increasing fluency and comprehension	18.1	30.5	24.3	10.9	22.5	16.7
Read fluently and with full comprehension	6.2	6.8	6.5	6.0	6.7	6.3

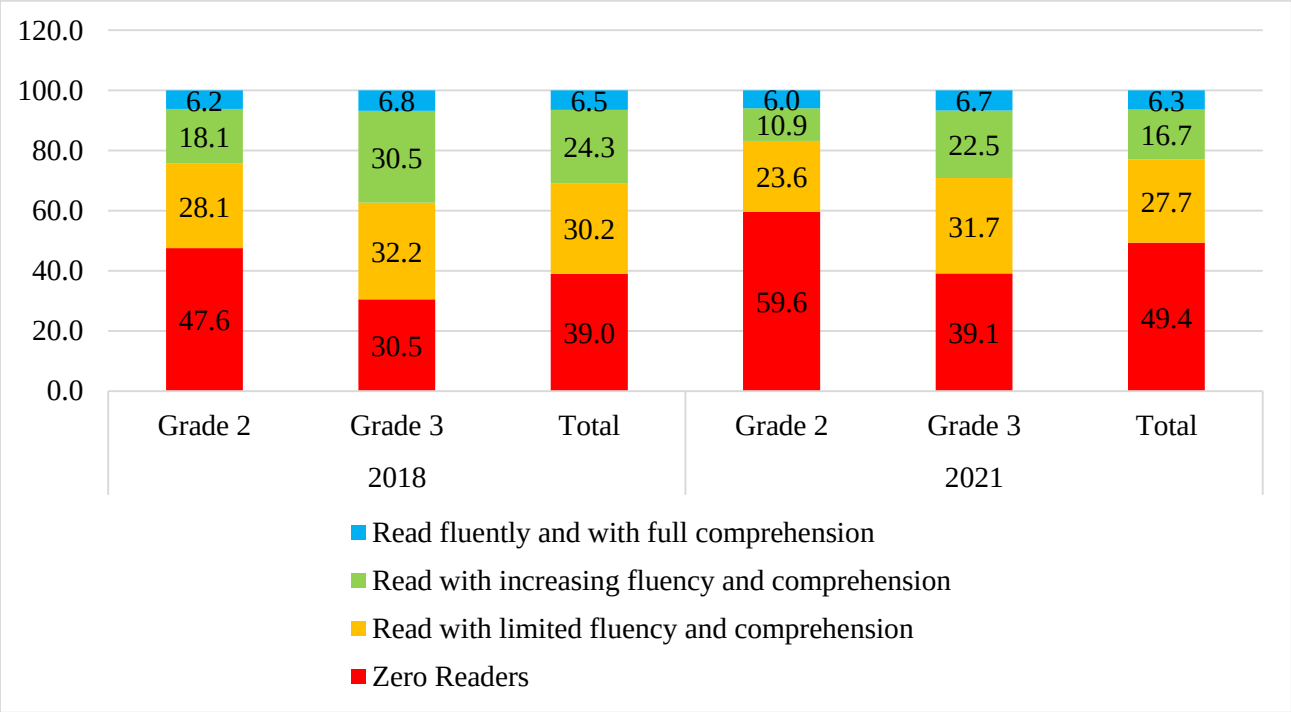


Figure 13 Percentage of students at benchmark levels by year in GEQIP Phase I schools

3.8. The trend of overall performance in each languages by year across subtasks

Figure 14 shows the overall performance of students across timed subtasks in six languages (Afan Oromo, Af Somali, Amharic, Hadiyyissa, Sidamuu Afoo, and Wolayttato) by assessment year. Since they recently joined, the remaining three languages (Afar Af, Berta, and Nuer) do not have a trend in previous EGRA studies. Moreover, it is important to aware that EGRA 2014 was conducted as Mini-EGRA only for Hadiyyissa and Wolayttatto and the other languages were not take part of it.

In EGRA 2016 and 2021, the performances of students were highest in Amharic, but lowest for Sidaamu Afoo than EGRA 2010 and 2018. On the contrary, students' performance was the lowest for Afan Oromo in 2021 than 2010, 2016 and 2018 and it was the lowest for Af Somali in 2016 than in 2010, 2018 and 2021. The highest performance for Afan Oromo and Af Somali was displayed in 2010. In Hadiyyissa and Wolayttatto students' better performances were recorded in 2014 and 2016, respectively, and their least performances were observed in 2021. In general, although there were differences among the languages, the overall trend revealed that students' performances were declining over time.

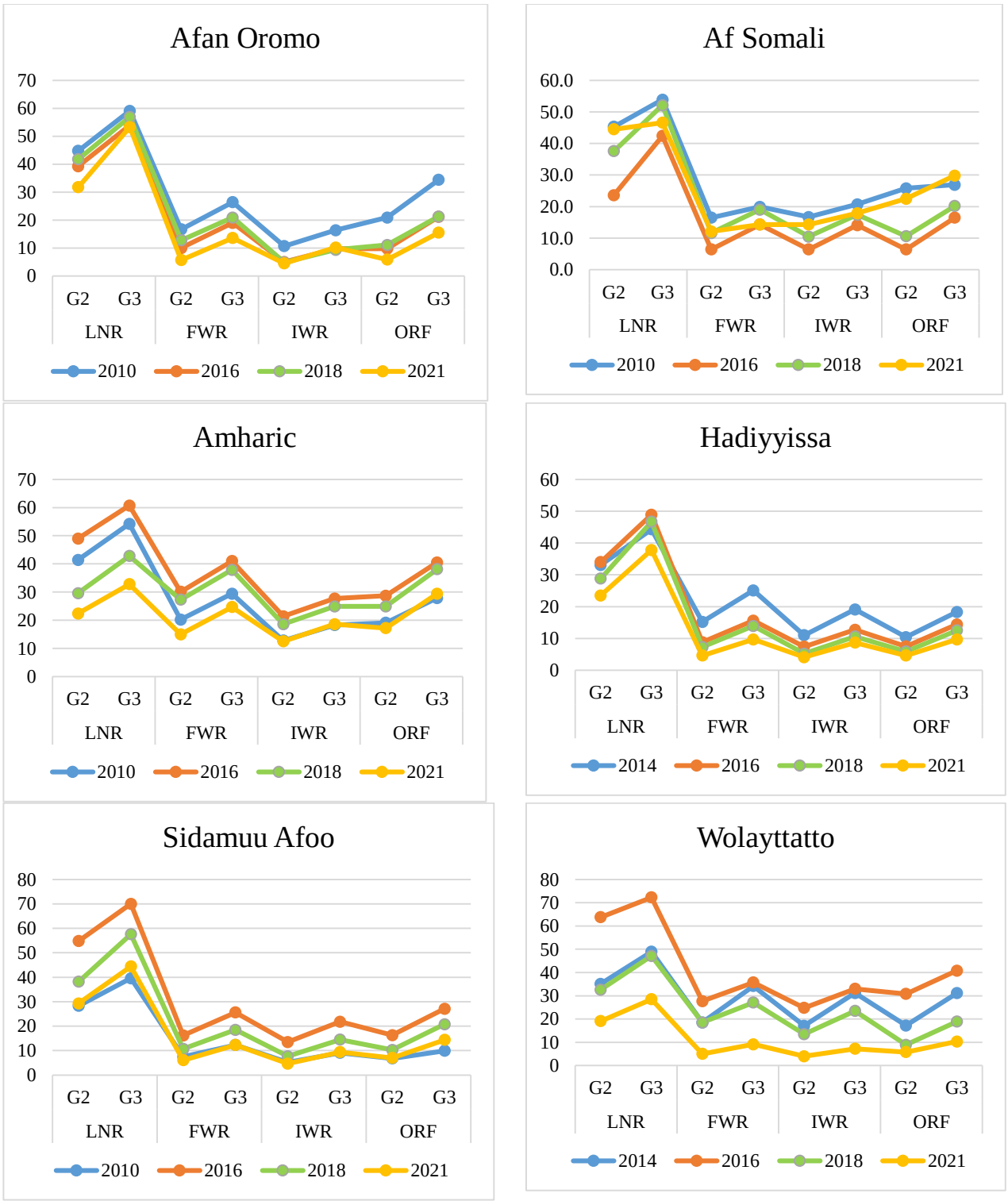


Figure 14 Trend of overall performance in each languages by year across timed subtasks

3.9 Contextual Questionnaires Analysis

The results of the contextual questionnaires completed by school principals, teachers, and students are reported in this part. The background characteristics of the participants are discussed in the first subsection. In the second part, the issue of professional development or trainings provided to school principals and teachers are presented. The availability of school resources and instructional materials are described in the third subsection. Finally, major factors associated with students' oral reading fluency performance are identified.

3.9.1 Participants Background Characteristics

Principals' background characteristics

Figure 15 shows the percentage of male and female principals. Of the total 484 EGRA sampled schools, 479 (99%) of school directors responded to the items included in the principals questionnaires. Among the participant school principals, 441 (92%) of them were males and only 38 (8%) of them were females. Similarly, in each language the number of female principals is very insignificant and ranged from the lowest 2% in Wolayttatto to the highest 19% in Berta.

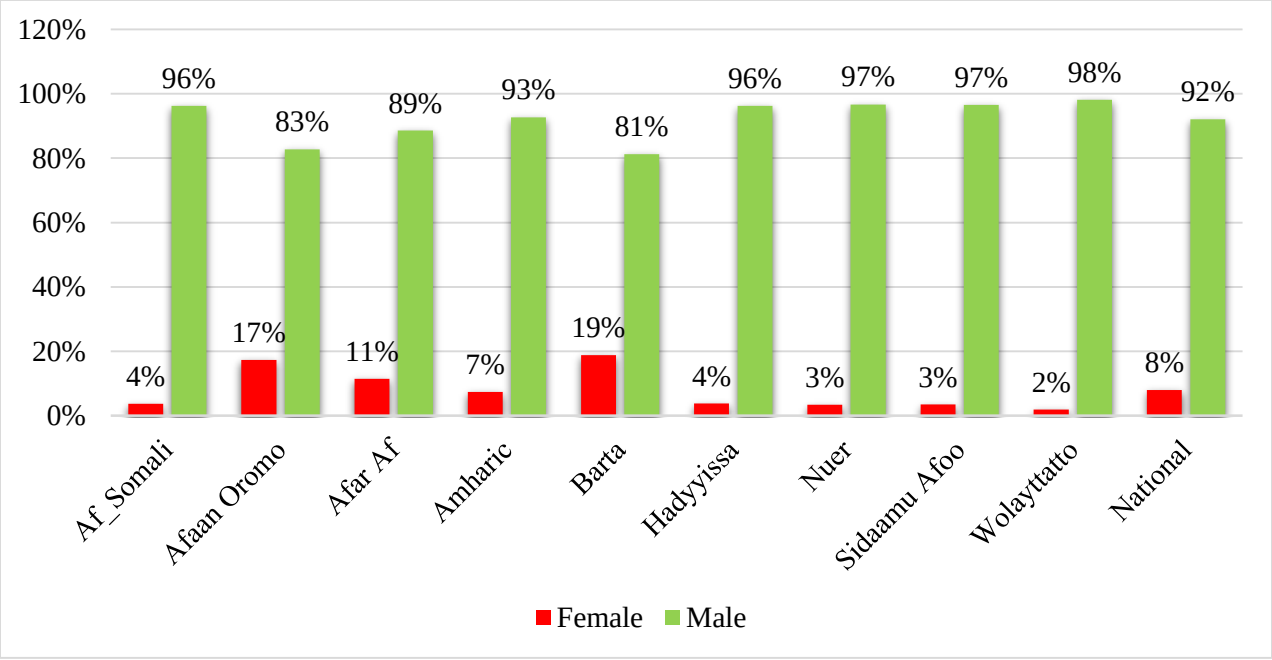


Figure 15 Percentage of school principals by sex across language

Figure 16 shows the age category of principals in the sampled schools. The majority of principals (62%) are in the age category 26-35 years. On the contrary, there is very small percentage (6%) of principals in each age category of 18 to 25 years and 46 and above years. This shows that the majority (88%) of the teachers are in the middle age.

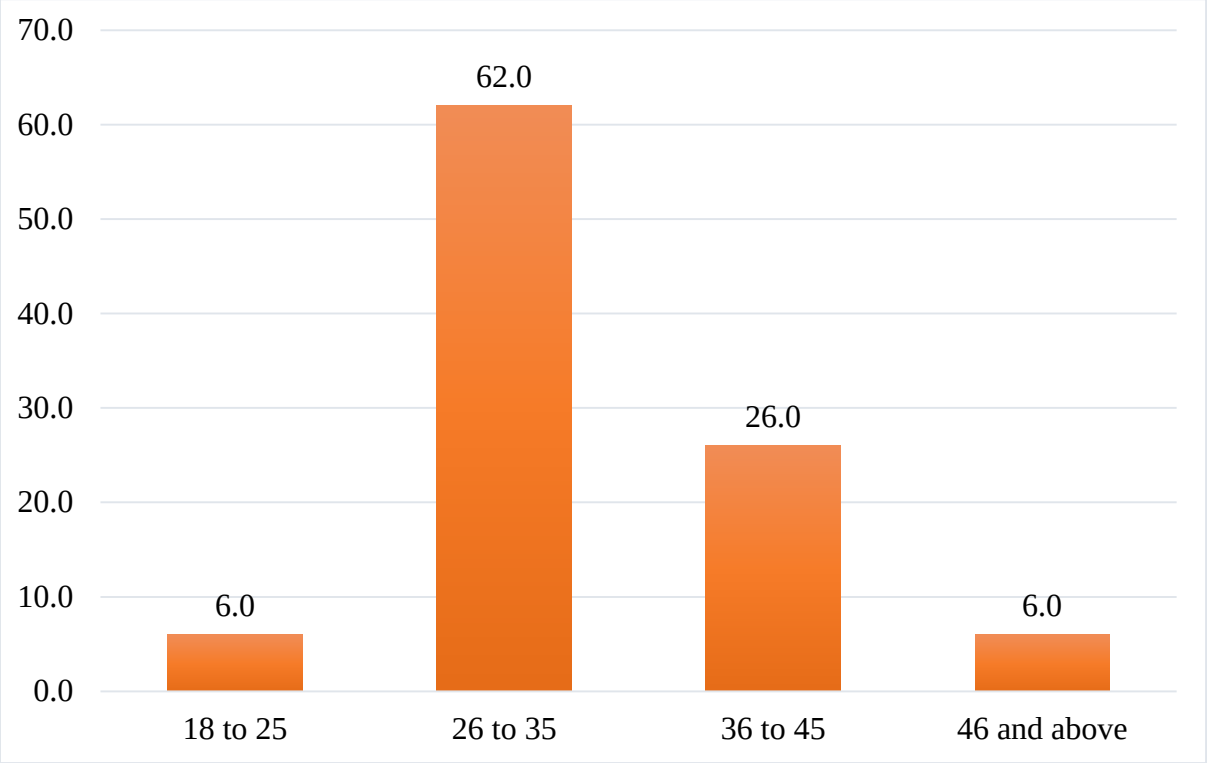


Figure 16 Percentage of school principals by age category

Figure 17 depicts the educational levels of the school principals by language. At the certificate level, there are still a substantial number of principals, with the highest percentage coming from Af Somali (56 percent), followed by Afar Af (40 percent), Berta (38 percent), and Sidaamu Afoo (31 percent). The percentage of diploma holders is high in Amharic (82%) as well as Hadiyyissa and Walayttatto with 81% each. Even though, the percentage of principals with a bachelor's degree or above is low, there is a variation across language groups in primary schools. In overall, the percentage of principals with a certificate, diploma and degree and above was 29%, 68%, and 3%, respectively.

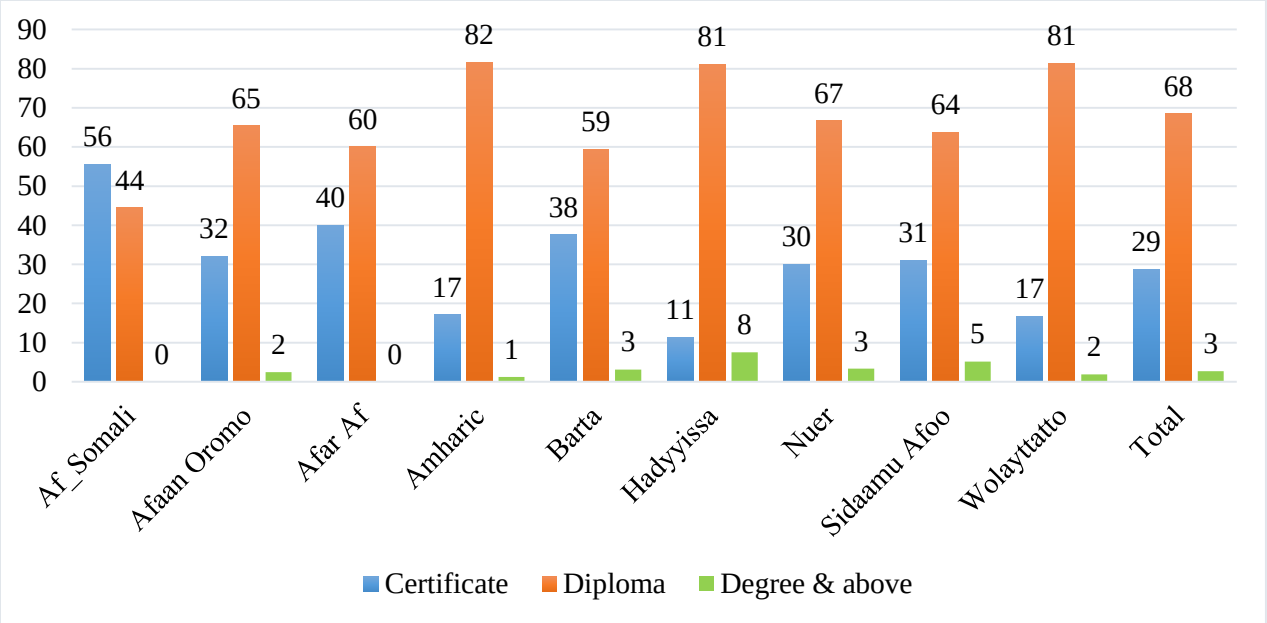


Figure 17 Principals' Education Level by Language in Percentage

Students' background characteristics

Table 21 and Figure 18 present demographic characteristics of the students who participated in this study. Among the sample students participated in the study, half of them are from Grade 2 and the remaining are from Grade 3. With regard to gender, 52% of them were males and 48% of them were females. When considering the sampled students across school location, 85% and 15% of them were from rural and urban schools respectively. The figure perfectly matches with the fact that more than 80% of the population lives in rural areas. Concerning the age of the students, 56%, 25%, 16% and 3% of them are above nine years, nine years, eight years and below eight years old respectively.

Table 21. Students' Background Characteristics

Students Characteristic		Frequency	Percent
Grade Levels	Grade 2	8845	50
	Grade 3	8902	50
	Total	17747	100
Gender	Female	8509	48
	Male	9238	52
	Total	17747	100
Location	Rural	15013	85
	Urban	2734	15
	Total	17747	100
Age Category	Less than 8 years	452	3
	Equal to 8 years	2891	16
	Equal to 9 years	4484	25
	Above 9 years	9920	56
	Total	17747	100

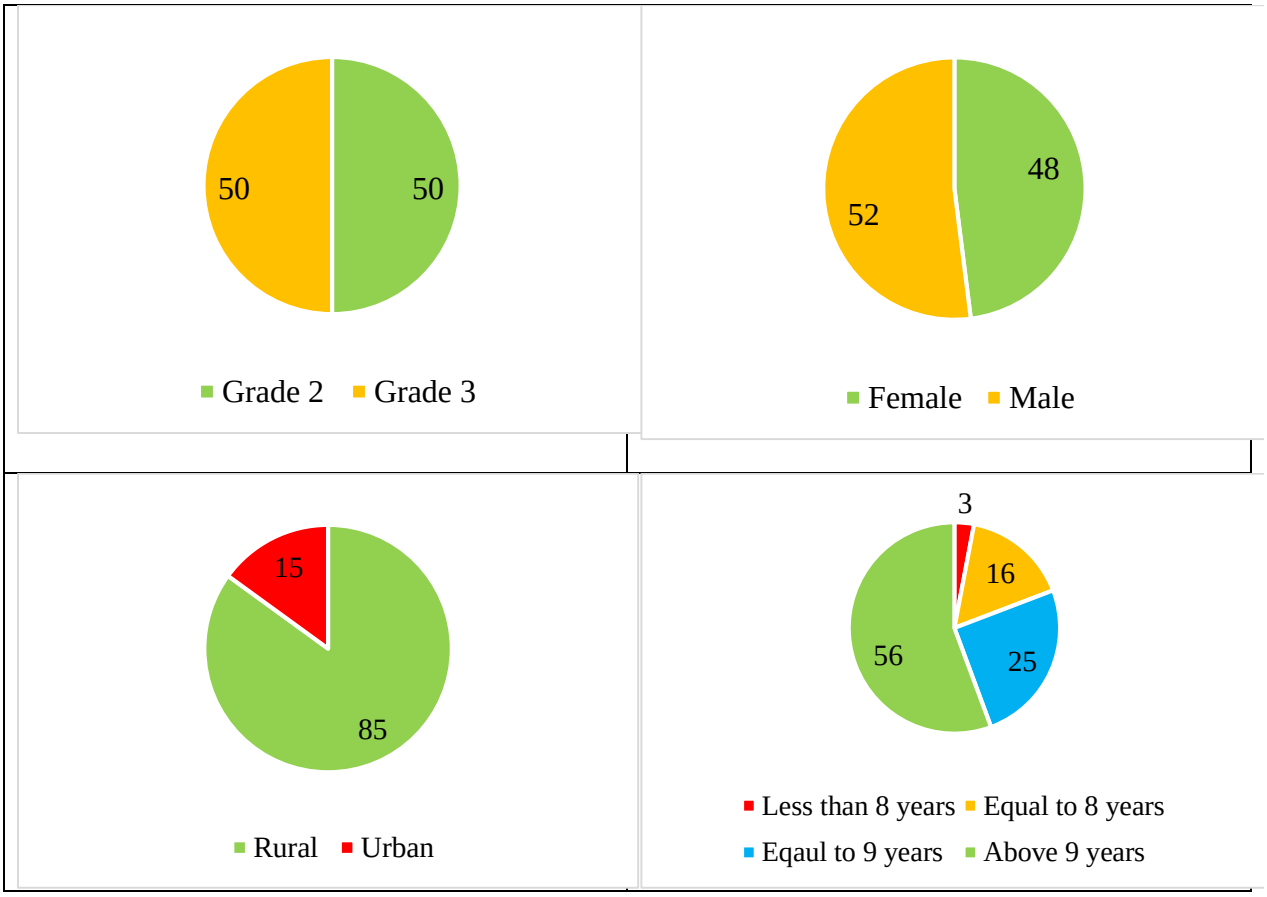


Figure 18 Students' Background Characteristics

Teachers' background characteristics

The percentages of male and female mother tongue language teachers in the sampled schools are shown in Figure 19. The findings show that the percentages of male teachers is higher than female teachers in all language groups except Afan Oromo, Amharic, and Hadiyyissa, where female teachers account for 64%, 69%, and 60%, respectively. On the other hand, in Af Somali, Afar Af, Nuer, and Sidaamu Afoo, more than 60% of teachers were male, followed by Berta (57.4%) and Wolayittatto (53.4%). Particularly, in Afar Af, female teachers constitute for only about 3% which requires more attention to increase female teachers' participation in schools.

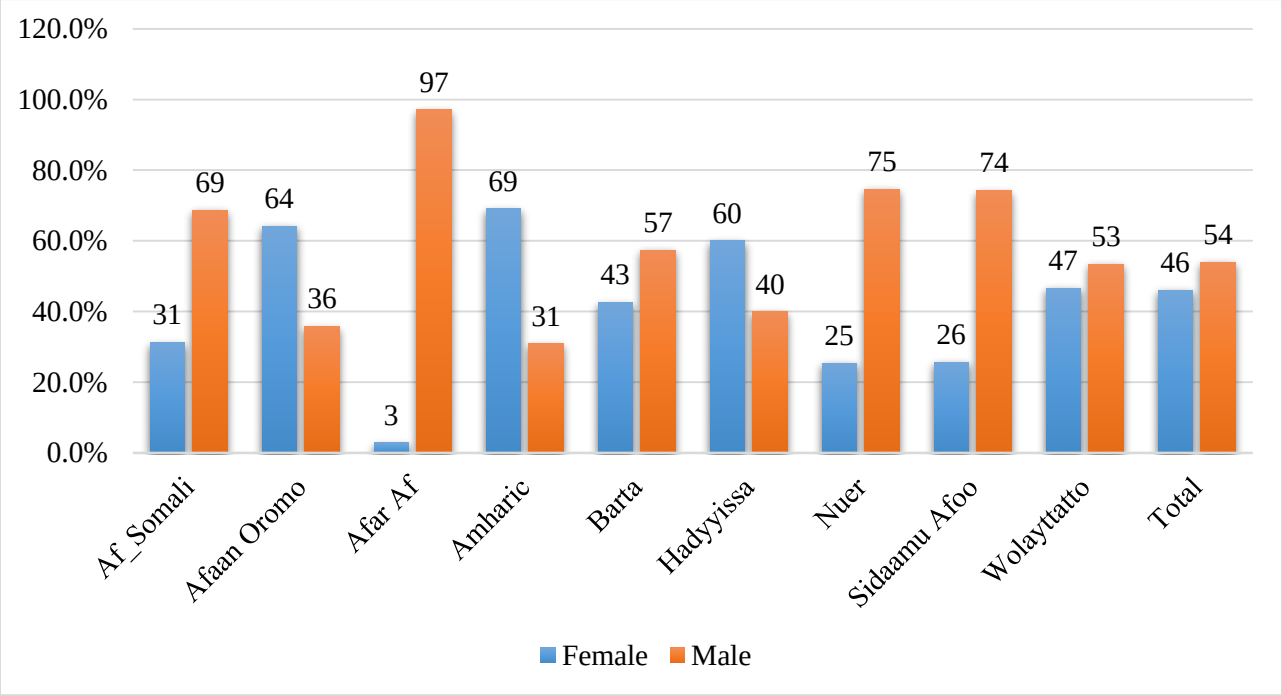


Figure 19 Percentage of Teachers by Gender in each Language

Figure 20 depicts the percentage of mother tongue teachers with their educational qualification per language. In this case, teachers in sample schools were requested to provide information about their qualifications. The results indicate that unqualified teachers are largest in Afar Af, which accounts for 17% of the total. Teachers who teach mother tongue with diploma are large in number in each language groups. For example, teachers in Afan Oromo (79.7%), Nuer (84.7%), Amharic (92.6%), Hadiyyissa (95.2%), Wolayttatto (97.1%), and Sidaamu Afoo (99%) reported that they earned diploma as their highest academic qualification. The result vividly shows that nearly all teachers

in Sidaamu Afoo and Wolayttatto are in possession of diploma. On the other hand, in Af Somali (54.2%), Afar Af (55.1%), and Berta (44.4%) teachers are at the certificate level. In general, 73% of mother tongue teachers in Grades 2 and 3 are diploma holders at the national level. The remaining 24% and 6% are certificate and degree holders respectively, while 7% are uncertified.

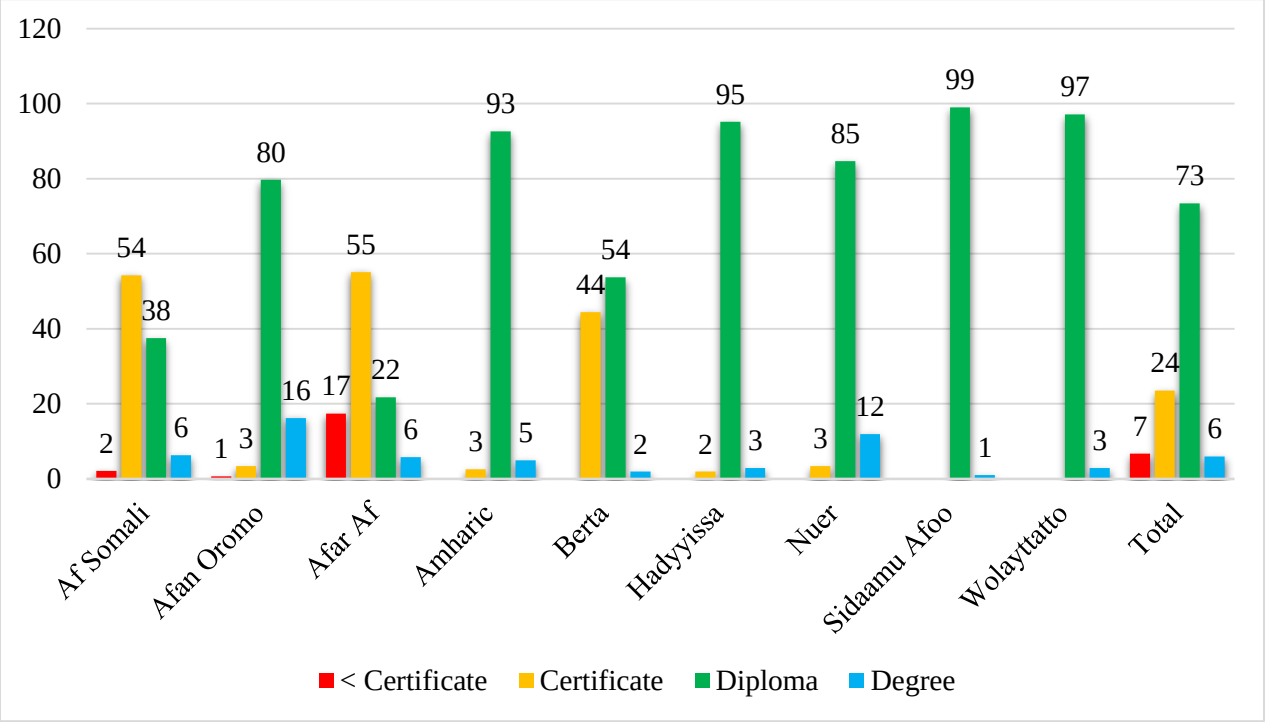
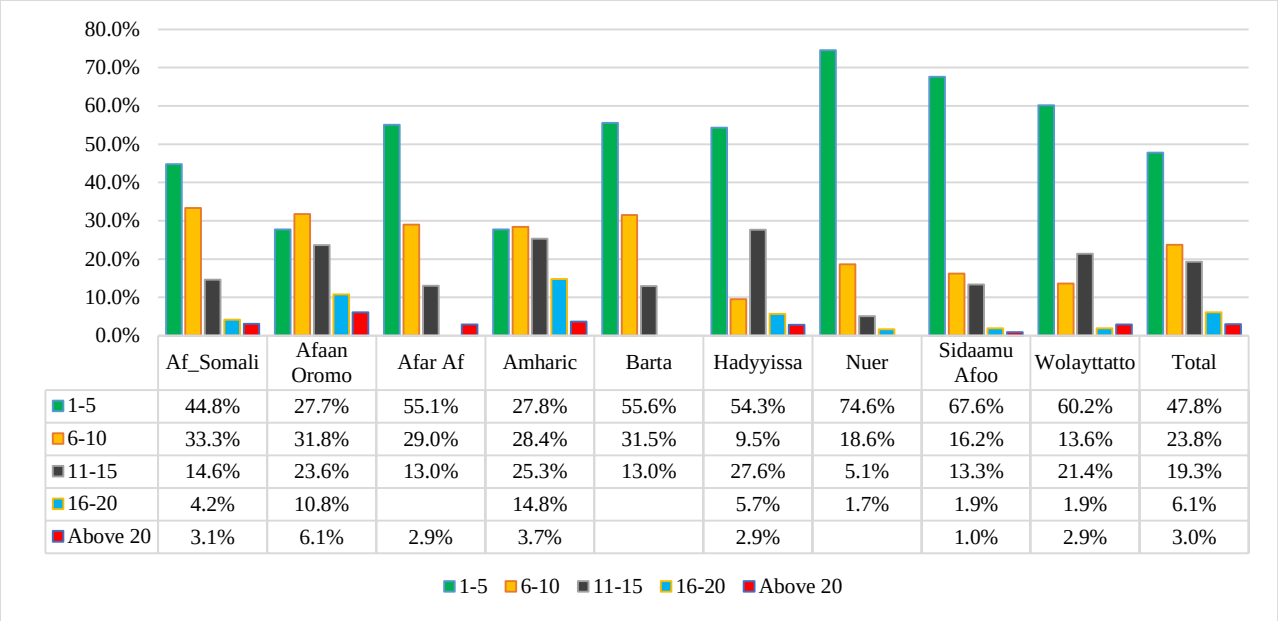


Figure 20 Teachers' Education Level by Language in Percentage (%)

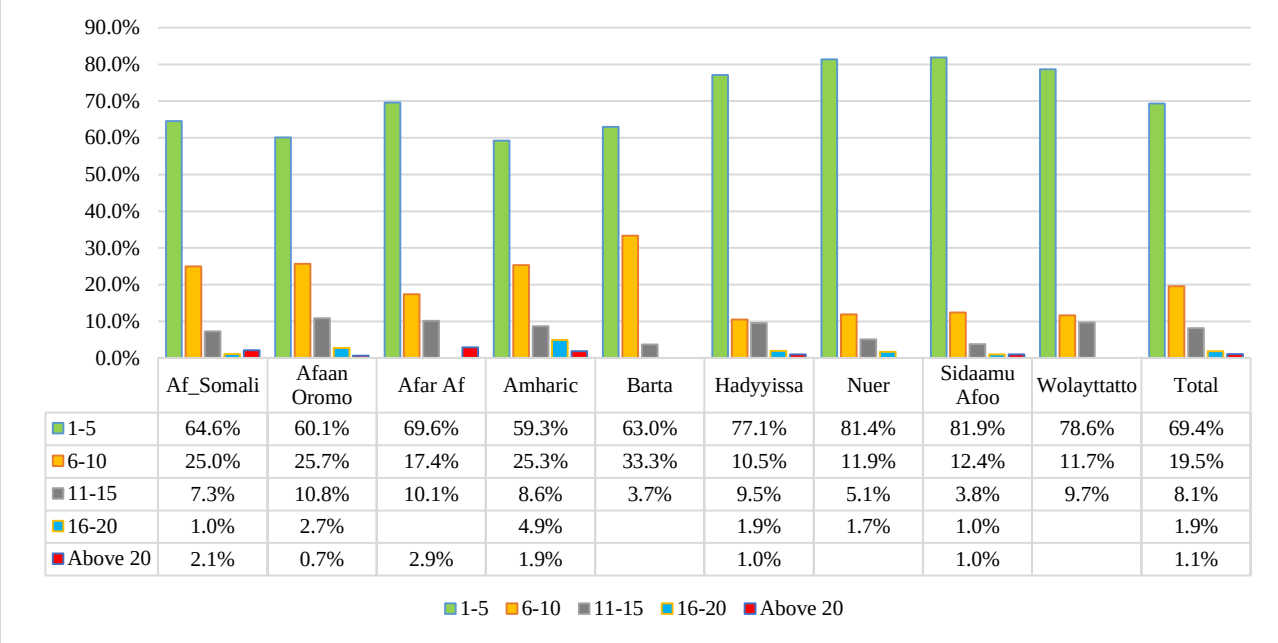
Figures 21 and 22 present, the overall teaching experiences of teachers and their experiences as a mother tongue teacher in percentages by language. At the national level, 43% of teachers have 1 to 5 years of teaching experience. On the contrary, among the participant teachers, there are only 3% of more experienced teachers with more than 20 years of teaching experiences. Nearly half of the teachers have 6 to 20 years of teaching experiences.

Similarly, when it comes to experience of MT teachers, about 69% at the national level, have 1 to 5 years of working experiences. In this case, Sidaamu Afoo (82%) and Amharic (59%) have the highest and lowest percentages of teachers with 1 to 5 years of teaching experience in MT, respectively. On the contrary, there are only about 9% of experienced teachers with more than 15 years of teaching experience in MT (Figure 22).



	Af_Somali	Afaan Oromo	Afar Af	Amharic	Barta	Hadyyissa	Nuer	Sidaamu Afoo	Wolayttatto	Total
1-5	44.8%	27.7%	55.1%	27.8%	55.6%	54.3%	74.6%	67.6%	60.2%	47.8%
6-10	33.3%	31.8%	29.0%	28.4%	31.5%	9.5%	18.6%	16.2%	13.6%	23.8%
11-15	14.6%	23.6%	13.0%	25.3%	13.0%	27.6%	5.1%	13.3%	21.4%	19.3%
16-20	4.2%	10.8%		14.8%		5.7%	1.7%	1.9%	1.9%	6.1%
Above 20	3.1%	6.1%	2.9%	3.7%		2.9%		1.0%	2.9%	3.0%

Figure 21 Service year overall as a teacher in percentage by language



	Af_Somali	Afaan Oromo	Afar Af	Amharic	Barta	Hadyyissa	Nuer	Sidaamu Afoo	Wolayttatto	Total
1-5	64.6%	60.1%	69.6%	59.3%	63.0%	77.1%	81.4%	81.9%	78.6%	69.4%
6-10	25.0%	25.7%	17.4%	25.3%	33.3%	10.5%	11.9%	12.4%	11.7%	19.5%
11-15	7.3%	10.8%	10.1%	8.6%	3.7%	9.5%	5.1%	3.8%	9.7%	8.1%
16-20	1.0%	2.7%		4.9%		1.9%	1.7%	1.0%		1.9%
Above 20	2.1%	0.7%	2.9%	1.9%		1.0%		1.0%		1.1%

Figure 22 Teachers' service year in percentage as a mother tongue teacher by language

3.9.2 Principals and Teachers Training

Figure 23 below shows the percentage of principals who have got trainings to monitor and support MT teachers. The results obtained from the principals revealed that, the majority of them did not receive trainings to monitor and support activities of MT teachers across all languages. It is the highest (89%) for Afar Af and the least for Nuer with 40%. Principals who received training for monitoring and supporting MT teachers, on the other hand, account for the highest percentage of Nuer (60%) and the lowest percentage of Afar Af (11%) followed by Berta (16 %). Overall, except for Aff Somali, Hadiyyissa and Nuer, more than half of the principals from the other languages, did not get any training related to MT language to support and supervise teachers during instruction.

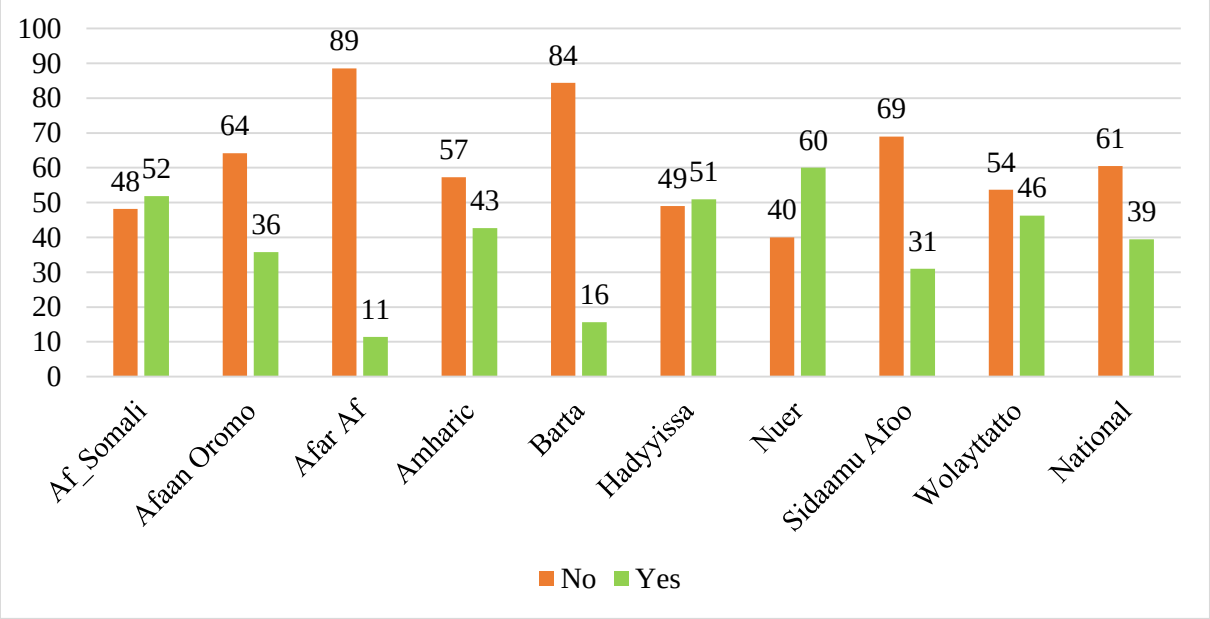


Figure 23 Percentage of principals who received trainings on MT to support teachers

Figure 24 shows the percentages of teachers who have received teachers' training in each of the languages. In this regard, the result obtained from the responses of the teachers unveil that practically all of them are trained in all languages. Particularly, majority of the teachers from Sidaamu Afo (96%), Af Somali (93%), Nuer (90%), and Haddyissa (88%) reported that as they attended formal teachers' training or as they trained as teachers. On the other side, Amharic (25%) and Berta (26%) had the largest percentage of untrained teachers, whereas Aff Somali (7%) and Sidaamu Afoo (4%) had the lowest percentages of untrained teachers. Overall, the percentage of

teachers who have got teaches' training is 84%, whereas the percentage of those who did not receive any teachers' training is 16%.

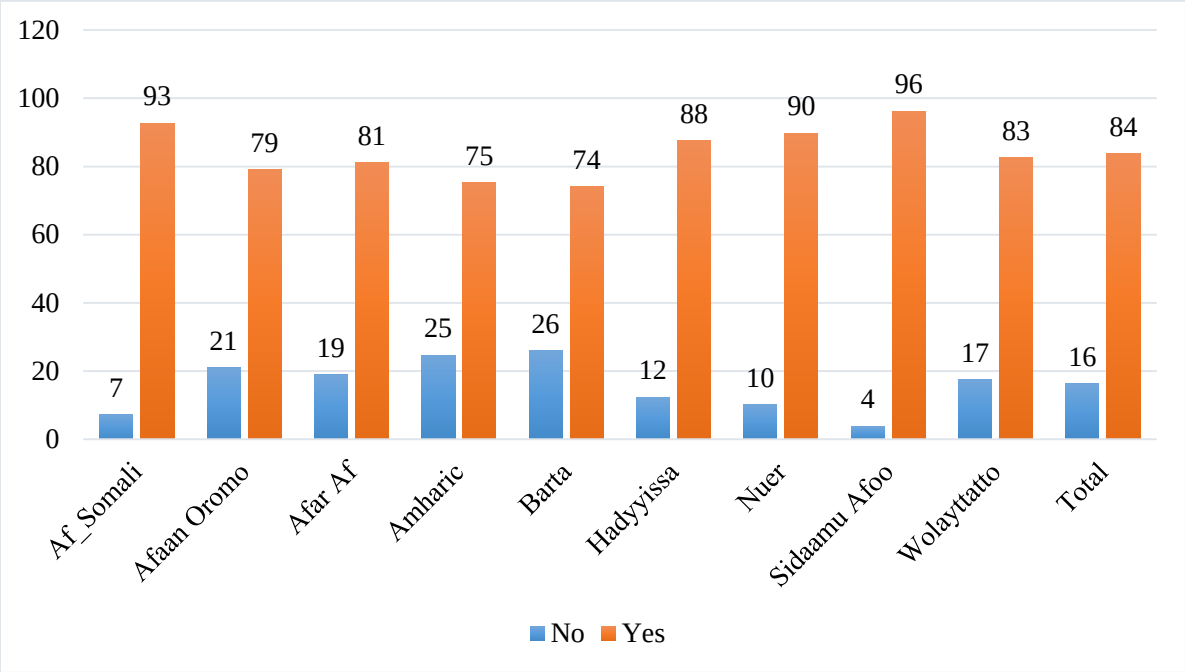


Figure 24 Percentage of teachers trained as a teacher across language

Figure 25 shows the percentage of teachers who took part in In-service or CPD trainings on MT instructional materials across the length of the trainings. As the result shows, at the national level, 46% of teachers did not obtain any training related to MT language instructional materials. However, the percentages of teachers who received training for more than 10 days, 5 to 10 days, and below 5 days are 10%, 17%, and 27%, respectively. Afar Af (58%), Sidaamu Afoo (55%), and Afan Oromo (55%) have the largest percentage of teachers who have not received any training. Hadiyyissa and Amharic have the least untrained teachers, with 34 percent and 38 percent, respectively.

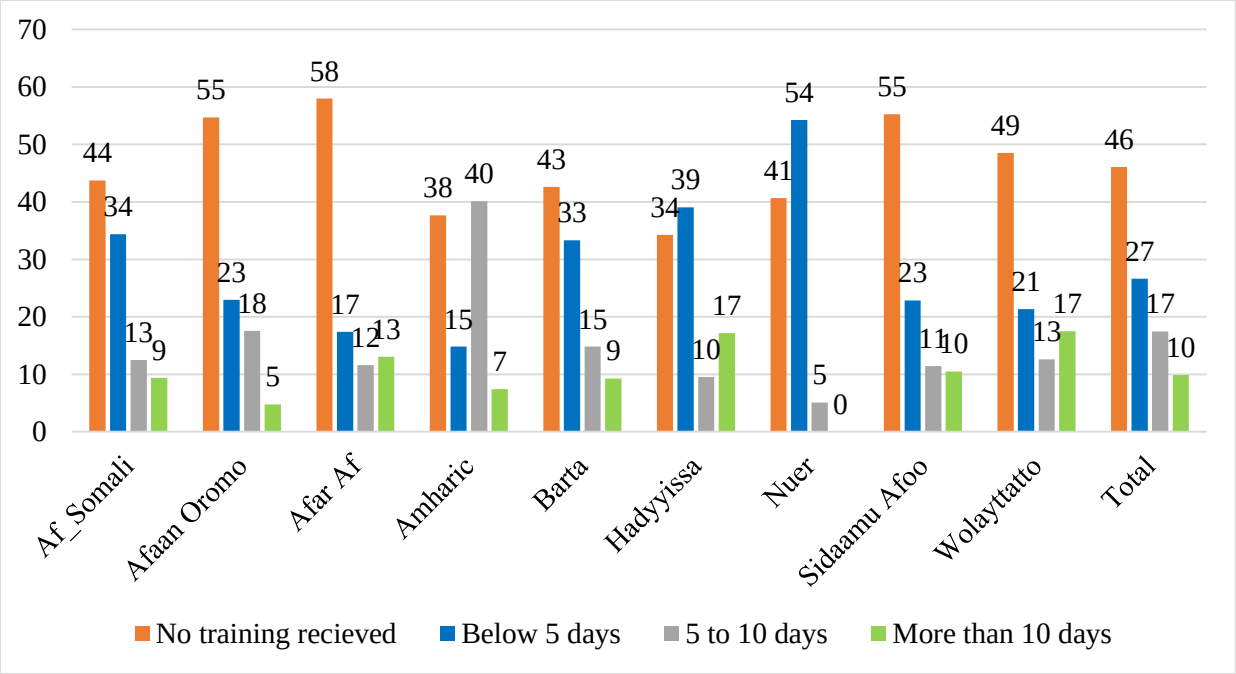


Figure 25 Percentage of teachers' participated in-service or CPD trainings on MT by duration

Figure 26 shows the percentage of teachers who received in-service training or CPD on how to teach the implemented MT instructional materials. As a result, more than 90% of teachers from all languages, except Afar Af, has responded as they received the training on how to teach the newly implemented MT instructional materials, either through in-service training or through a CPD program. The highest and lowest percentages of teachers who received the training were from Nuer (100%) and Afar Af (77%) respectively. The number of teachers who did not obtain the training, on the other hand, ranged from 23 percent in Afar Af to 3 percent in Amharic. Nearly 90% of teachers across the country participated in the training.

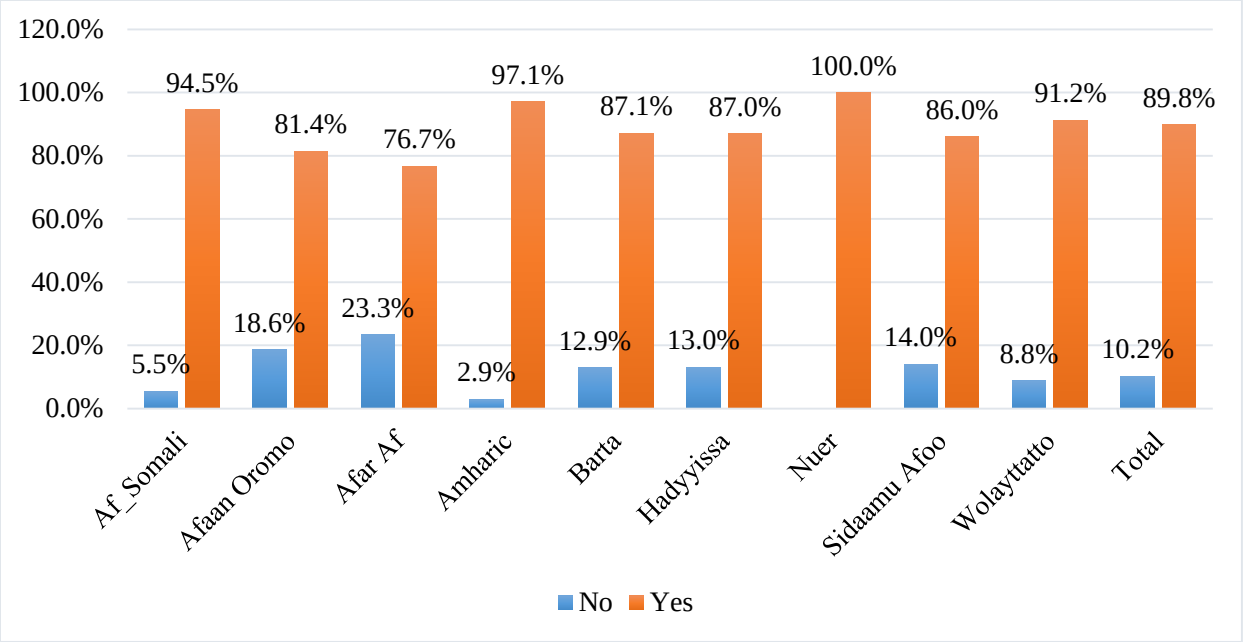


Figure 26 Percentage of teachers trained how to teach with the mother tongue materials

3.9.3 School resources / instructional materials

Figure 27 shows the percentage of school principals' response to textbook -student ratio for Grades 2 and 3. There is ample evidence from the data, that the majority of students in both grade levels share text books in the ratio 1:5, 1:4, 1:3 and 1:2. The percentage of students who are sharing textbooks in the ratio 1:5 is 54% and 48.7% for Grades 2 and 3, respectively. On the contrary, the percentages of students who have their own textbooks without sharing is very low (2.5% for Grade 2 and 5.7% for Grade 3).

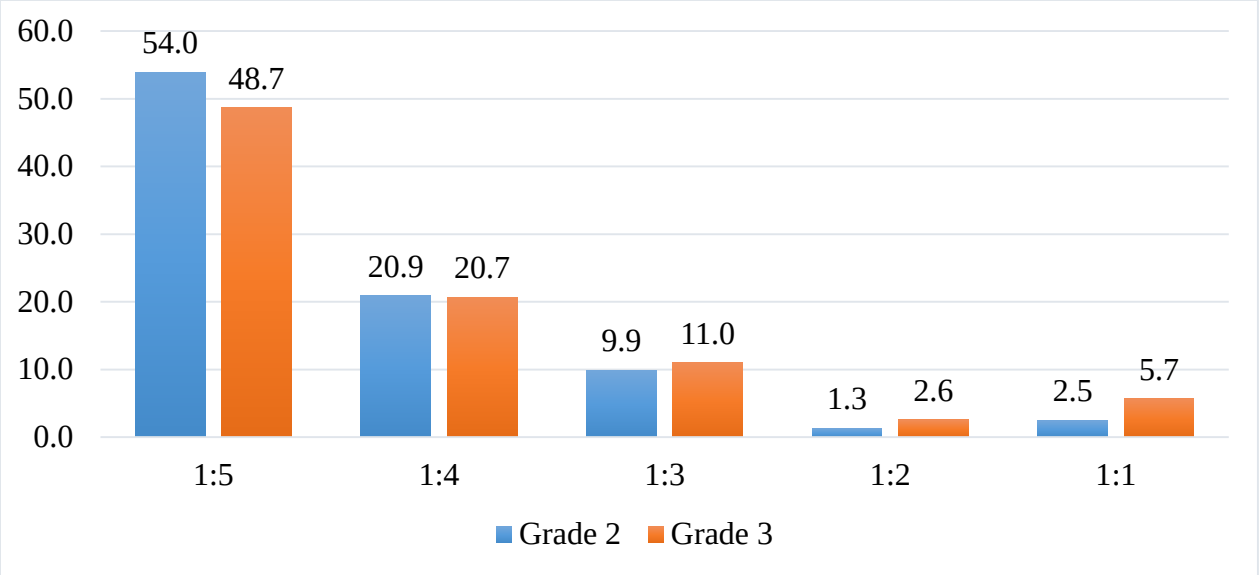


Figure 27 Percentage of school principals' response to textbook -student ratio

Students were asked whether they had MT textbook, as illustrated in figure 28. As the result indicates, majority of them (66%) said they owned an MT textbook. However, 34% of them said they did not have any MT textbooks that they could share with others. This coincides with the principal's response, which indicated that the majority of the students shared MT textbooks.

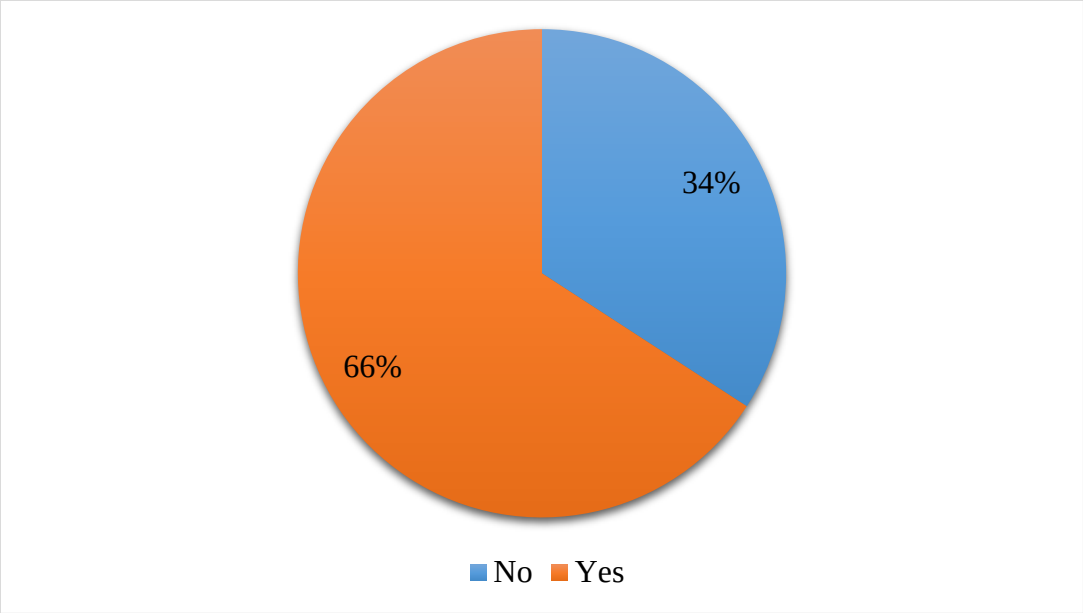


Figure 28 Percentage of students' response in having mother tongue textbook

As indicated in Figure 29, about 93% and 73% of the teachers, said that they had an MT textbook and an MT teacher's guide respectively. The remaining 7% and 27% of them responded as they did not have mother tongue textbooks and teachers guide, respectively. Furthermore, the majority of teachers who possessed these instructional materials responded that they used them every day at school. In this case, almost 90% and 83% of them, respectively, used a mother tongue textbook and teacher's guide every time.

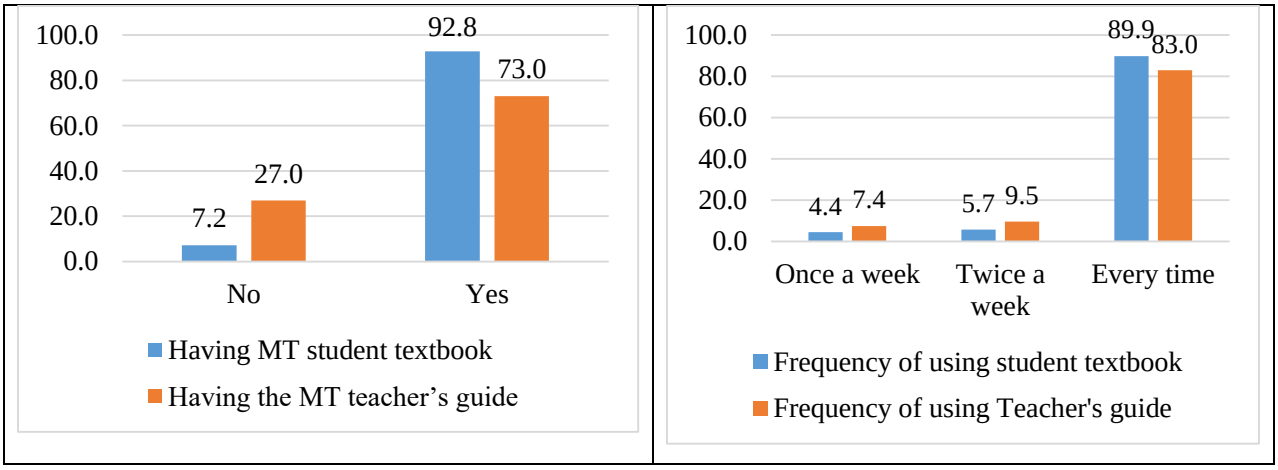


Figure 29 Percentage of teachers' response in having mother tongue textbook, teacher's guide and frequency of using

As illustrated in figures 30 and 31, teachers and students were asked about the availability of some reading materials and facilities at schools. As can be seen from the figures, around 54%, 83 %, and 60% of mother tongue teachers said that their schools lacked a functional library, reading corners, and sufficient supplementary reading materials, respectively. On the other hand, around 59% of them said that their students borrowed supplementary reading materials and read them at home as well as 73% of them responded that supplementary reading materials are accessible to the students.

Similar to the response of the teachers, majority of the students indicated that their schools lacked reading resources and facilities. In this case, 83%, 58%, 89%, and 67% of the students responded, the of absence of reading corners, functional libraries, supplementary reading materials at school, and additional reading materials at home respectively. Furthermore, 69% of them indicated that, they did not read any other reading materials other than their MT textbook.

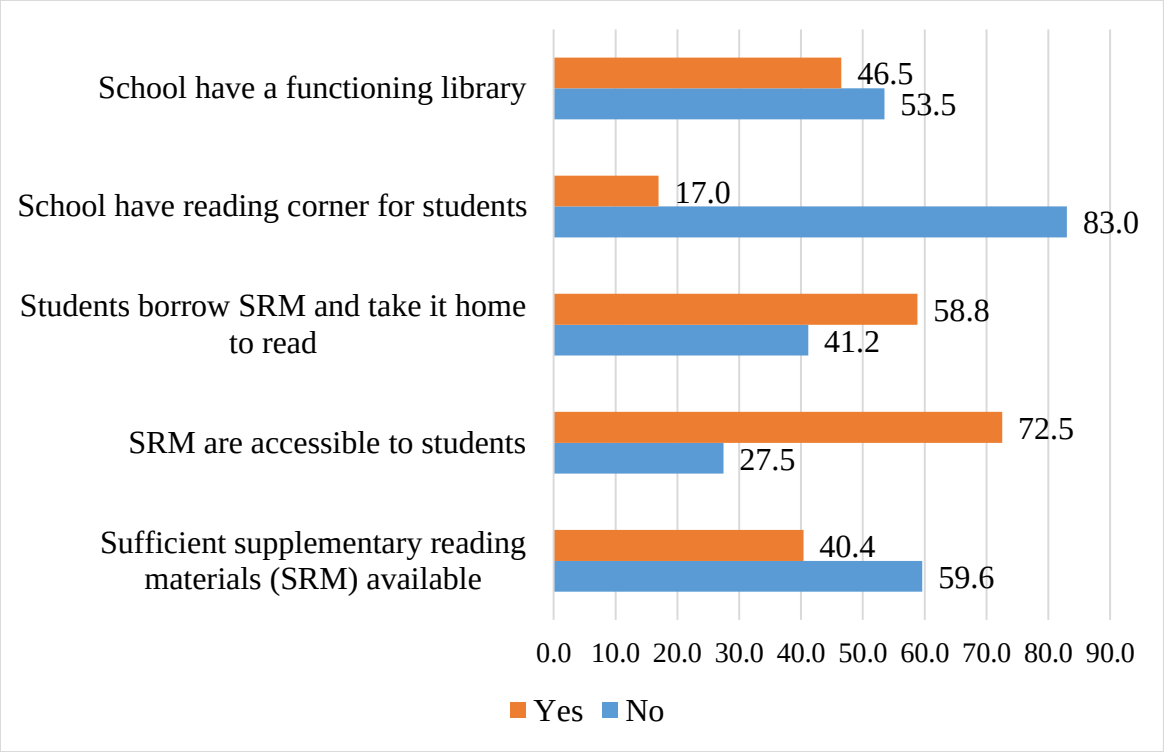


Figure 30 Response of teachers' regarding availability of reading materials and school facilities in percentage

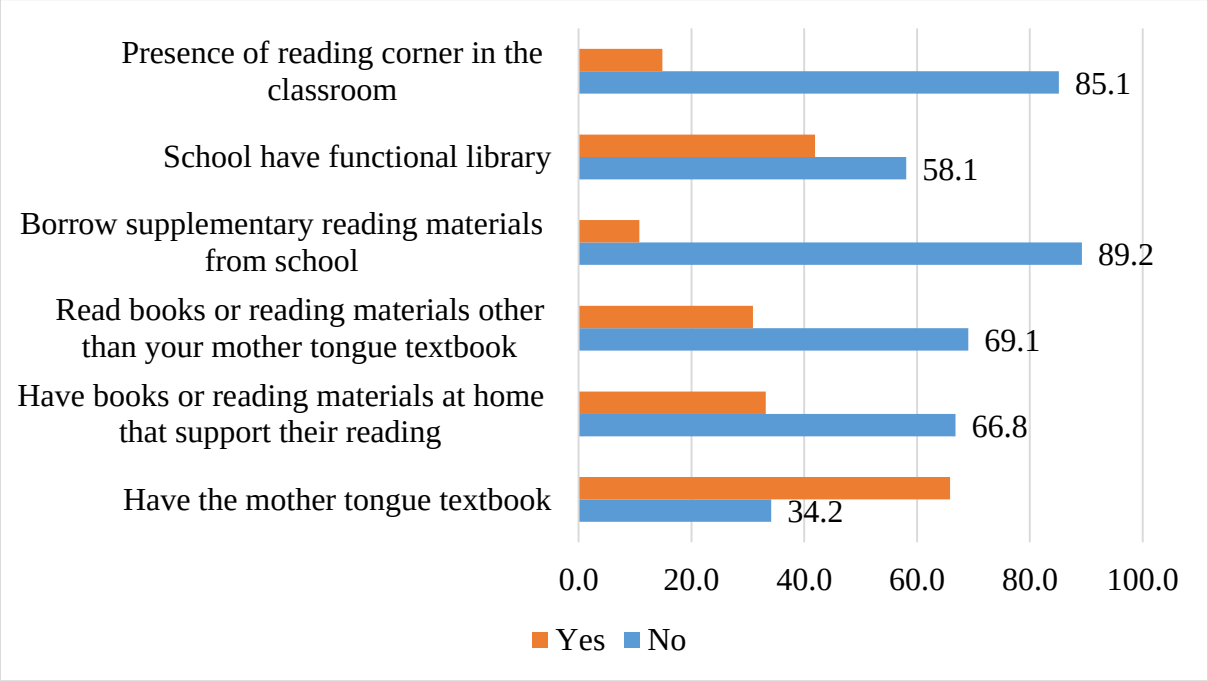


Figure 31 Response of teachers' regarding availability of reading materials and school facilities in percentage

3.9.4 Parental Support in Mother Tongue

School principals and teachers were asked whether the Parent, Teacher and Student Association (PTSA) is functional and provides any help for students learning or reading, as shown in Figure 32. In this regard, while about 53% of school principals stated that the school parent, teacher, and student association (PTSA) was functional, on the contrary, nearly 89 % of the teachers responded that PTSA was not functional. This shows that, the responses of the participants contradict to each other. With regard to the PTSA support, 77% of school principals and 62% of teachers said that, PTSA provided support for the students' learning in mother tongue. On the other hand, 38% of teachers and 23% of principals reported that PTSA did not offer any support for students learning in mother tongue.

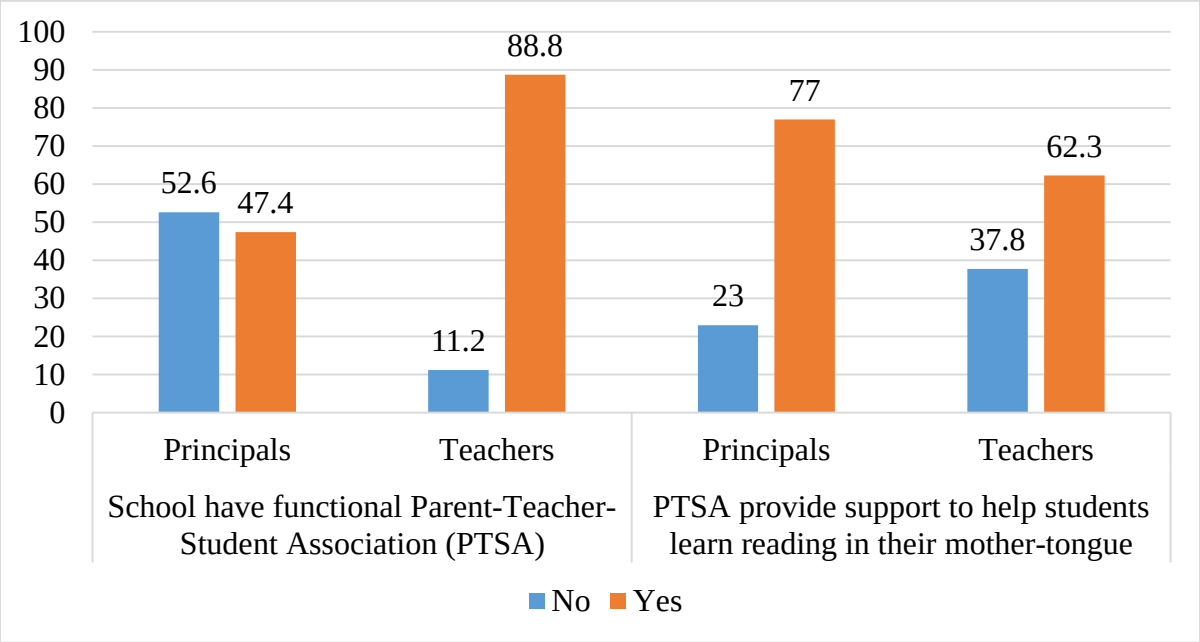


Figure 32 Response of principals and teachers regarding PTSA support in percentage

As shown in figure 33, students were asked about their family reading ability and support from their family while they are reading as well as having enough time to study or reading book. In this case, around 67%, 56%, and 69% of them reported as they had adequate time to study or read a book, received help from family members while reading, and had a family member(s) who can read and write, respectively. However, the remaining percentage stated that they did not have

enough time to study or read a book, did not receive support from their family members while reading, and did not have a family member(s) who could read and write.

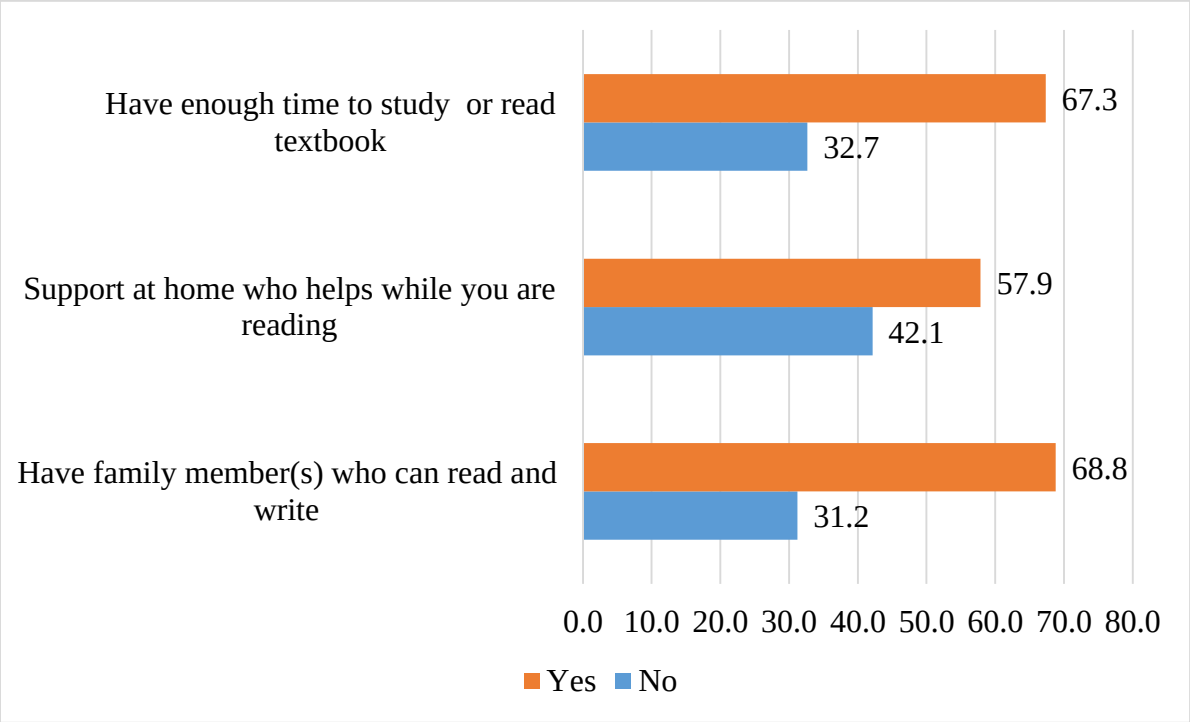


Figure 33 Response of students regarding family's reading and support in percentage

3.9.5 Factors associated with students' oral reading fluency

3.9.5.1 Background Variables

Principals' characteristics that show significant associations with students' ORF are indicated in Table 22. These are age, gender, education level, trainings given to monitor and support MT teachers, as well as the length of trainings. Even though, all of these characteristics show positive correlations with ORF and are found to be significant, the correlations are said to be weak. Regarding the educational level of the principals, the highest level of educations showed a weak but significant association with their students' oral reading fluency. With regard to gender, the results indicate that male principals as compared to females showed positive and significant association with their students' reading performances. Actually, there is a gender imbalance in the number of respondents. In addition, those schools' principals that have got trainings to monitor and support MT teachers showed a statistically significant positive association with reading performances with weak correlation coefficient. Moreover, the length of trainings also showed a weak positive and significant associations with students' reading performances.

Table 22. Correlation between Oral Reading Fluency and Principals' characteristics

Principals' Characteristics	Oral Reading Fluency	
	Correlation coefficient	Sig.
Age	.034**	0.000
Gender	.057**	0.000
Highest level of education	.003**	0.000
Training received that help to monitor and support mother tongue language teachers	.022**	0.000
Length of the training	.079**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

Table 23 shows the correlations of schools' principals variables related to the support provided for MT teachers and students' reading performances. The variables include support for MT teachers while in teaching reading, observations of mother tongue teachers while teaching, frequency of observations, the reviewing of lesson plans prepared by MT teachers. With the exception of support related to pedagogy which showed negative association, the other variables showed positive correlations and with significant values, though, the correlation coefficients are very weak.

Table 23. Relationship between support given to mother tongue teachers and ORF

Principals support to MT- teachers	Oral Reading Fluency	
	Correlation coefficient	Sig.
Support for mother tongue teachers on how to teach reading	-.002**	0.000
Frequency of support for mother tongue teachers on how to teach reading	.087**	0.000
Observing mother tongue teachers while teaching	.017**	0.000
Frequency of observing mother tongue teachers while teaching	.022**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

The teachers' characteristics associated with reading performance are gender, trained as a teacher, the highest level of education, years of experiences as a teacher, whether or not they received training on mother tongue materials, length of the training, and training how to teach mother tongue, as shown in Table 24. As a result shows, teachers' educational level have weak and negative correlation, while all other teachers' characteristics related variables except gender (which has insignificant correlation) have shown weak but positive and statistically significant correlation with the students' oral reading performances.

Table 24. Correlation between Oral Reading Fluency and Teachers' characteristics

Teachers Characteristics	Oral Reading Fluency	
	Correlation coefficient	**Sig.**
Gender (female=1, male=2)	0.002	0.132
Highest education level	-.094**	0.000
Trained as a teacher	.017**	0.000
Highest service year as a teacher	.182**	0.000
Highest service year as a mother tongue teacher	.100**	0.000
Number of days in-service training or CPD	.057**	0.000
Teachers received training on how to teach mother tongue materials either in-service training or CPD,	.148**	0.000

***. Correlation is significant at the 0.01 level (2-tailed).*

Table 25 presents the association between oral reading fluency performance and some students' background characteristics. In this case, although the correlation coefficients are weak, all students' characteristics are significantly associated with oral reading performances. Gender, age, sameness of students' home and instructional languages, students' previous year grade level, having family member(s) who can read and write, having any one at home who can help them while they are reading and having enough time to study or read, have shown positive statistically significant correlations with students oral reading fluency. Students' absenteeism for more than one week and enrollment at preprimary education before Grade 1, on the other hand, were negatively associated to their oral reading performance. It is an amazing finding to have a negative relationships between attending preprimary education and students' oral reading fluency which requires further research.

Table 25. Correlation between oral reading fluency and students' characteristics

Students background characteristics	Oral Reading Fluency	
	Pearson Correlation	Sig.
Gender (female=1, male=2)	.048**	0.000
Age	.092**	0.000
The sameness of students home language and instructional language at school	.021**	0.000
Attending preprimary education before grade 1	-.010**	0.000
Students last year grade level	.155**	0.000
Students' absenteeism from school for more than one week	-.025**	0.000
Having family member(s) who can read and write	.044**	0.000
Having any one at home who helps you while they are reading	.102**	0.000
Have enough time to study or read your textbook	.140**	0.000

3.9.5.2 School resources as principals, teachers and students responses

In this subsection, the correlations between some school resources and students' performance on oral reading fluency are presented based on the response of school principals, MT teachers and students.

Table 26 shows the correlations of schools' resources with students' ORF. Most of the variables related to schools' resources showed positive correlations with students' reading performances. These include availability of MT textbooks and teacher guides for Grades 2 and 3, availability of mother tongue textbooks and teachers' guides for teachers, the presence of reading corners, receiving supplementary reading materials (SRM), presence of functional library and accessibility of supplementary reading materials for students. However, there are items that surprisingly showed negative correlations with students' reading performances. These items include textbook-student ratios for both grades and the number of MT teachers in the school. Students who are sharing textbooks with their peers had relatively better reading performances than those students who own textbooks on 1:1 ratio. Likewise, the presence of more number of MT teachers in the school was also not positively correlated with students' reading performances.

Table 26. School Resources correlated with Students' ORF as principals' response

School Resources as principals' response	Oral Reading Fluency	
	Correlation coefficient	Sig.
Availability of MT- textbooks and teacher guides	.013**	0.000
Textbook – student ratio for grade 2	-.071**	0.000
Textbook – student ratio for Grade 3	-.058**	0.000
Teachers having mother tongue student textbook for the grade level they teach	.021**	0.000
Teachers having mother tongue teachers guide for the grade level they teach	.010**	0.000
Number of mother tongue teachers in the school	-.008**	0.000
Having a reading corner in the classrooms at school	.023**	0.000
Schools received supplementary reading materials	.053**	0.000
Accessibility of supplementary reading materials for students	.002**	0.000
Have a functioning library or a reading room	.049**	0.000
Usability of the library or reading room by students	.056**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

Table 27 shows, the teachers' response on school resources that are associated with oral reading fluency. In this case, the majority of school resource related variables showed a weak but statistically significant positive association with the students' reading fluency. The coefficients of correlation ranged from 0.006 for frequent use of student textbooks to 0.165 for school support for teachers in their efforts to teach students how to read. However, the Grade 3 textbook-to-student ratio and teachers' use of MT student textbooks were found to have a weak and negative relationship with students' performances. In this scenario, this conclusion is consistent with the principals' findings of a negative relationship between Grades 3 textbook - student ratio, however it contradicts with the Grade 2 textbook – student ratio.

Table 27. Correlation between oral reading fluency and school resources as teachers' response

School Resources as Teachers Response	**Oral Reading Fluency**	
	Correlation coefficient	Sig.
Grade 2 Textbook – student ratio	.047**	0.000
Grade 3 Textbook – student ratio	-.141**	0.000
Class size	.028**	0.000
Teachers' having MT student textbook	-.028**	0.000
Frequency of using student textbook	.006**	0.000
Teachers' having MT teacher's Guide	.107**	0.000
Frequency of using Teacher's guide	.031**	0.000
Availability of sufficient supplementary reading materials (SRM)	.107**	0.000
Students borrow SRM and take it home to read	.018**	0.000
School have reading corner for students	.075**	0.000
School have a functioning library	.068**	0.000
Frequency of students using the library	.047**	0.000
Teachers' supervise while pupils are using the library	.072**	0.000
Schools support for the teachers effort to teach students how to read	.165**	0.000
PTSA provide support to help students learn reading in their mother tongue	.124**	0.000

***. Correlation is significant at the 0.01 level (2-tailed)*

Table 28 shows, the correlation between oral reading fluency and school resources as students' response. As the result shows, all the school resource related variables showed a weak but statistically significant positive association with the students' reading fluency. The highest correlation coefficient was for the frequency of bring the MT textbook to the classroom (0.162) and the lowest was for the presence of a reading corner in the classroom (0.008). In this case unlike, the finding under principals, having the MT textbook has also a positive correlation.

Table 28. Correlation between oral reading fluency and school resources as students' response

School Resources as Students Response	Oral Reading Fluency	
	Correlation coefficient	Sig.
Having the mother tongue textbook	.123**	0.000
Frequency of bring the MT textbook to the classroom	.162**	0.000
Having additional books or reading materials at home that support their reading	.129**	0.000
Reading additional books or reading materials other than mother tongue textbook	.148**	0.000
Borrowing supplementary reading materials from school	.123**	0.000
Schools having functional library	.077**	0.000
Presence of a reading corner in their classroom	.008**	0.000

3.9.5.3 Teachers Activities

This subsection presents some activities such as method of instruction and assessment techniques as well as instructional time that teachers perform in the classroom.

The teachers' activities that are significantly associated with reading performance in their schools are listed in Table 29. Students who copied down text from the chalkboard have shown weak and negative correlation while all other teachers' activities related variables have shown weak but positive correlation with the reading fluency of the students.

Table 29. Correlation between Oral Reading Fluency and some teachers' activities

Teachers' Activity	**Oral Reading Fluency**	
	Correlation coefficient	Sig.
Number of periods per week	.045**	0.000
Method of teaching		
• Strictly following the "I do, You do, We do" method	.133**	0.000
• The whole class repeated sentences that you said first	.105**	0.000
• Students copied down text from the chalkboard	-.008**	0.000
• Students retell a story that they read	.062**	0.000
• Students voiced out unfamiliar words	.064**	0.000
• Students learned meanings of new words	.120**	0.000
• Students read aloud to teacher or to other pupils	.180**	0.000
• Students were assigned to read by themselves during school time	.072**	0.000
• Organizing remedial classes for students who lag behind	.068**	0.000
• Discussing with parents of students who lag behind	.129**	0.000
Assessment techniques		
• Written evaluations	.021**	0.000
• Oral Evaluations	.075**	0.000
• Checking classroom exercise	.150**	0.000
• Checking homework	.138**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

4. Conclusions and Recommendations

This section is divided into parts. The first one presents the conclusions of the research and the second part suggests some recommendations based on the results of the study.

4.1 Conclusions

The main objective of this study was to assess the performances of early grade students in mother tongue and identify some of the factors that are related to the reading performances. The results of this study help policy-makers, concerned stakeholders and scholars to have adequate information about the performance of early grade students, and factors that are related to the students' performances. The study adopts a quantitative survey research design, in which the EGRA performance subtasks and contextual questionnaires were used. The participants were 484 primary schools/ school principals, 968 teachers and 19,360 students selected through two-stage cluster sampling techniques from nine languages. Descriptive statistics was used to summarize the findings of the study and some inferential statistics for the purpose of generalization.

According to the results of the EGRA 2021, majority of students were unable to read enough letters or words per minute, which is consistent with past EGRA findings (USAID Ethiopia, 2010, 2014, 2016, 2018). In both Grades 2 and 3, the total mean scores of students in reading comprehension were the lowest across all languages. On the contrary, students in both grades performed better in all language when it comes to listening comprehension. The students also did better in phonemic awareness than the other remaining subtasks, with the exception of students from Afar Af, Berta, and Nuer languages.

When the findings were disaggregated across gender, boys performed relatively better than girls across most languages in each subtasks of EGRA. This is a clear manifestation that, gender inequality is still one of the gray areas that has to be addressed with regard to learning outcomes. The disparity between urban and rural students is still persisting in which case urban students performed better than rural students in most subtasks across languages: Afan Oromo, Afar Af, Amharic, Berta, Hadiyyissa, and Wolayttatto, with the exception of letter recognition in Hadiyyissa and Wolayttatto as well as invented words reading in Afan Oromo. Nevertheless, in Af Somali, Nuer, and Sidaamu Afoo, rural students performed better than urban students in all timed

subtasks, with the exception of letter name recognition and oral reading fluency in Af Somali and Nuer, respectively.

Similarly, in untimed subtasks, urban students outperformed rural students in most language groups. However, rural students outperformed urban students in reading and listening comprehension in Af Somali whereas in phonemic awareness, rural students from Afan Oromo, Nuer and Hadiyyissa also performed better than urban students. In a similar vein, in listening comprehension rural students performed higher than urban students in Hadiyyissa.

Regarding the reading performance of students by the types of program, in all subtasks except in listening comprehension, students from GEQIP phase I sample schools slightly performed better than students from both NEGRA and READ II sample schools. On the other hand, students from READ II schools performed slightly better than students from NEGRA in all timed subtasks except in oral reading fluency. In untimed subtasks, students from NEGRA schools slightly performed better than students from READ II in reading comprehension and listening comprehension except in phonemic awareness.

With regard to reading proficiency, a large number of grade 2 students (68%) at national level were not able to read and categorized as zero readers in oral reading fluency subtask. Across language groups, students in Afar Af and Berta had the highest percentage (92%) of zero readers, while those from Amharic and Af Somali had relatively the lowest (34%) and (35%), respectively. Similarly, in Grade 3, the total percentages of zero readers are about 51% for oral reading fluency. In this case, a large percentage of zero readers were from Berta and the least were from Af Somali language. From the findings, the majority of students lack the skills of oral reading fluency which would be critically important for further interventions.

With regard to the reading proficiency performances of students across grade levels, as the results show, the percentages of students who reached the required highest benchmark or level 1 (reading fluently with full or almost full comprehension) at the national level were found about 5% and 6% for grades 2 and 3, respectively. Similarly, the percentages of students attaining level 2 benchmark which refers to “reading with increasing fluency and comprehension” were 8% and 17% for grades 2 and 3, respectively.

Overall, the percentage of students who attained the two highest levels referring to "full or almost full" and "increasing" reading proficiency known as *functional reading level* were only 13% for Grade 2 and a23% for Grade 3 which almost dropped by half for both grade levels as compared to EGRA 2018. In general, the result of EGRA 2021 was the least for each language as compared to all previously conducted EGRA studies. According to this information, despite the fact that many interventions have been in place since the first EGRA findings were released, the majority of students are still struggling to read at a grade-appropriate level.

The purpose of contextual questionnaires is to provide useful information for policymakers and stakeholders. In this regard, majority of school principals 441(92%) are males and 38 (8%) of them are females. The percentage of female principals in each language ranged from 2% to 19% for Wolayttatto and Berta, respectively. When looking for MT teachers by gender, 46% are females. Among the sampled teachers, only 3% of them possessed a bachelor's degree, and nearly 97% had a diploma or certificate at national level. However, 7% of them were facilitators who had never had any training. Specifically, in Afar Af, the unqualified teachers or teachers without certificate accounts for 17%. Concerning the work experiences, most of MT teachers had 1 to 5 years of experience in teaching mother tongue language.

Regarding professional development, about 61% of the school principals did not receive training on how to support MT teachers during instruction. Particularly, in Afar Af and Berta, large percentage of them did not receive any training. Except for Af Somali, Hadiyyisa, and Nuer, the percentage of principals who did not receive training was higher than the percentage of principals who did receive training in each of the languages. Similarly, at the national level, majority (84%) of teachers responded that they have obtained a teacher training. However, still 16% reported that they did not take any teacher training. The percentage of teachers who did not take any teacher training ranges from 4% to 26% in Sidaamu Afoo and Berta respectively. Particularly, 46% of teachers did not obtain training on how to teach M. In this case, Afar Af (58%), Sidaamu Afoo (55%), and Afan Oromo (55%) have a large percentage of teachers who did not receive any training.

With regard to school facilities and instructional materials, about half of the principals responded that most of the students shared an MT textbook in a 1 to 5 ratio. Likewise, it was also found that

34% and 67% of students replied as they did not have any MT textbooks and additional reading books respectively. This result confirms that textbook scarcity is critical in MT instruction for both Grades levels. According to the response of the teachers, 7% and 27% of them lacked student textbooks and teacher's guides respectively. In addition, 54%, 83%, and 60% of teachers said their school lacked a functional library, reading corner, and supplemental reading materials respectively. Similarly, 58% and 85% of students confirmed that their schools did not have a functional library and reading corners correspondingly.

Moreover, 47% and 77 % of school principals said that Parent-Teacher-Student Association was not functional and did not assist student learning, respectively. About 33% of students stated that they did not have enough time to study/read at home and 42% of them did not get any support at home while reading.

The correlations between some contextual variables and students' oral reading fluency were the other findings obtained from the response of the participants. In this case, despite the fact that most of the variables showed weak correlations: gender, highest education level, provision of trainings on assisting MT teachers, and the length of trainings are some of the school principal's variables that were significantly associated with the students' ORF.

Furthermore, according the response of the participants, variables related to school resources that showed positive correlations with students' reading performances include: the availability of mother tongue textbooks for students, availability of MT textbooks and teachers' guides for teachers, presence of reading corners and functional libraries, existence of functional PTSA and accessibility of supplementary reading materials for students.

According to the findings from the teachers' responses, their teaching methods and assessment techniques were positively associated with the students reading performance, in spite of the fact that, large class size had a negative correlation.

As the current findings indicate, the reading performances of the students is very low. Among the additional factors that contribute to the students' low performances COVID-19 and internal conflicts as well as the drought in certain areas can be mentioned. The impact of the COVID-19 pandemic on schooling could be one of the reasons for the students' low reading performances. In this case, due to the school closure, the formal schooling was disrupted, and students were not actively involved in schoolwork for more than four months. According to a research undertaken

by the NEAEA in partnership with the IEA in 2021, 44% of students were without schoolwork during the COVID-19 schools closure and 61% of teachers did not engage in any schoolwork during the closures.

Internal conflict and drought in some parts of the country could also be the other reason for the students' low performances. In this regard, 20 and 16 sample schools were replaced at the spot of data gathering due to the internal conflict and drought respectively. Yet, reconsidering all these causes, the findings of this study shows the reading performance of early grade students is still in jeopardy, necessitating quick attention and effective reading interventions.

4.2 Recommendations

The overall results of EGRA 2021 indicate that the performance of early grade students in reading is very low as compared to the previous EGRA studies. So, to improve the performance of students' reading skills, all the concerned governmental bodies, non-governmental organizations and relevant stakeholders are required to make concerted efforts so that we can get out of the quagmire of this stalemate. Particularly, this predicament has to be at the front of our national agenda. In addition, the government structures in the education sector from the Ministry of Education to the school levels, teachers, parents and students should play an important role. On top of that, according the suggestions forwarded by the stakeholders who participated in EGRA 2021 dissemination workshop, there is a need to establish system level accountability in general education to promote students' learning. Furthermore, to take actions the following specific recommendations are proposed based on the findings of EGRA, 2021.

- **Minimize the disparity in reading performance across subgroups:** There was a mean difference in reading performances across various subgroups such as gender, location, languages and program types. Therefore, in order to bridging the gaps, special attentions should be given for those disadvantaged and low performing subgroups. Particularly, in most subtasks girls performed lower than boys, even though, there have been efforts made to promote girls' education at all grade levels in Ethiopia. Thus, to alleviate this problem, the Ministry of Education and Regional Educations Bureaus should strengthen the special supports for girls that have been in place for years. In addition, the implementation of different policies,

initiatives and programs relating to the affirmative actions of females should be regularly monitored and evaluated.

Regarding the performance of students by location, urban students' performed better than rural. Similarly, students from the emerging regions (Afar Af, Berta and Nuer) had performed lower than the other language groups. Hence, policy-makers should focus their attention on the low performances of these groups and design special support mechanism. In general, the Ministry of Education and/or the Regional Education Bureaus should track differences in regularly and take preventive measures which can minimize the existing gaps among these various subgroups.

- **Improve the quality of school leadership:** The highest qualification, working experience and short term training of the school principals were related to the students' reading performances, though they had a weak coefficient of correlation. Therefore, MoE and Regional Education Bureaus should recruit, train and assign professionally qualified and experienced school principals, as they are the key actors to provide quality education. Moreover, it is suggested that less qualified principals should be encouraged to attend further high education level, through various short term and long term programs. For those less experienced school principals, the experience sharing programs should be arranged.
- **Strengthen teachers' professional development:** Teachers' competence is the major factor in addressing reading performances of students. Therefore, there is a need to devise mechanisms for professional development of teachers, including revisiting the ongoing teacher education of MT teachers. In this regard, teachers should be trained to teach reading as most of them are still teaching reading without any training. In this case, Teachers' Training Colleges (CTCs) should devise various strategies design on how to train teachers to teach reading in early grades in both pre-service and in-service programs as well as in CPD program. Teachers' upgrading should be strengthened. Particularly, teachers who are at the certificate level should be upgraded as significant percentage of teachers for some languages: Af Somali, Afar Af and Berta are below the required qualification.
- **Encourage teachers to use various methods of teaching and assessment techniques:** Methods of teaching and assessment techniques had positive correlations with the students reading performances. Thus, teachers should be encouraged to use teaching strategies such as "I do, You do, We do" method, let students read aloud to the teacher or other students, and

have students learn the meanings of new words to help students improve their reading skills. READ M&E's formative assessment approach, which permits teachers to monitor and enhance students' reading competency in early grades, should be utilized in schools. It includes both instruction-embedded assessments and periodic EGRA-like assessments that teachers can use for monitoring and fostering reading performance of their students.

- **Provide school facilities and instructional materials:** According to the research findings, most schools lack functional libraries, reading corners, student textbooks, teacher's guides, and SRM. Particularly, shortages of mother tongue students' textbooks and teacher's guides continue to be a serious problems in schools. Thus, these school facilities and instructional materials should be available at all schools. In this case, MOE, REBs should allocate appropriate funds by engaging the school communities to establish functional libraries and reading corners for schools. In addition, the MOE and REBs should set aside sufficient funds to print and distribute the required instructional materials (students' textbooks, teacher's guides and SRM) to all schools. Moreover, Regional Education Bureaus, Zonal and Woreda Education offices should collaborate to monitor where or not schools have appropriate and enough instructional materials for students.
- **Strengthen functional Parent-Teacher-Student-Association (PTSA):** Parental involvement in school activities, commitment to ensuring that students are ready to learn, and pressure on the school to maintain good academic success are all important issues. In this regard, it is critical to strengthen the Parent-Teacher-Student Association in schools. For example, absenteeism, wastage of instructional time, and a lack of funds are some of the problems that affected the performances of students in reading. Thus, educational leaders should be engaged PTSAs regularly to address these issues, as their contribution is invaluable to handle such problems and monitor students' learning progress in schools.
- **Enhance Educational Actors' Commitment and Establish an Accountability System:** Various stakeholders participated in the EGRA report dissemination workshop were forwarded lack of commitment for most educational actors to improve education quality and the absence of accountability in the education system. Therefore, to alleviate these problems, the Ministry of Education should devise the mechanisms how to improve the commitment of the educational actors and design the strategies to implement the accountability system in the education sector.

Since commitment and accountability have been recognized as one of the most critical factors in enhancing the quality of education.

- **Educational leaders should have awareness how to cope with educational disruptions:** The impact of COVID-19, conflict in various parts of the country and drought in some regions of Ethiopia were the major problems contributed for the low performance of students in reading. This was due to a number of instructional times wasted and there were a high rates of absenteeism among both teachers and students. In addition, a substantial number of students were without any schoolwork and large number of teachers did not engage in any teaching during the physical school closures in Ethiopia. Therefore, to manage the education system under certain challenges or crises like COVID-19, educational leaders should be aware ahead of time and respond accordingly by promoting an alternative learning system such as eLearning and distance learning.
- **The impact of various programs for EGRA interventions should be regularly assessed and evaluated:** Even though several remedies and interventions have been in place since the first EGRA results were announced, students' reading performance is still very low, as indicated in the findings of this study and also obviously documented in the EGRA 2018 report. Therefore, the effectiveness of various programs in early grade reading interventions should be seriously assessed and evaluated in order to track the key areas for future planning and design.

Appendices

Appendix A. Timed subtask by Language and Grade Levels

Language – Af Somali

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	103438	44.49	30.283	-2.150	.000	-0.06
	Grade 3	100141	46.64	32.118			
Familiar Words Reading	Grade 2	103438	12.14	12.767	-2.130	.000	-0.09
	Grade 3	100141	14.27	13.274			
Invented Words Reading	Grade 2	103438	14.25	17.947	-3.639	.000	-0.02
	Grade 3	100141	17.89	24.235			
Oral Reading Fluency	Grade 2	103438	22.52	26.145	-7.292	.000	-0.29
	Grade 3	100141	29.81	17.947			

Language -Afan Oromo

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	1464930	31.83	25.973	-21.449	.000	-0.80
	Grade 3	1551067	53.28	27.325			
Familiar Words Reading	Grade 2	1464930	5.69	9.681	-7.907	.000	-0.65
	Grade 3	1551067	13.60	14.253			
Invented Words Reading	Grade 2	1464930	4.47	8.641	-5.727	.000	-0.50
	Grade 3	1551067	10.20	13.620			
Oral Reading Fluency	Grade 2	1464930	5.94	12.038	-9.532	.000	-0.61
	Grade 3	1551067	15.47	18.172			

Language -Afar Af

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	35313	14.06	34.043	-3.951	.000	-0.11
	Grade 3	31963	18.01	28.448			
Familiar Words Reading	Grade 2	35313	2.10	7.657	-1.836	.000	-0.21
	Grade 3	31963	3.94	10.052			
Invented Words Reading	Grade 2	35313	1.77	6.947	-1.810	.000	-0.23
	Grade 3	31963	3.59	9.084			
Oral Reading Fluency	Grade 2	35313	2.24	12.117	-1.916	.000	-0.16
	Grade 3	31963	4.15	12.414			

Language -Amharic

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	947037	22.38	24.176	-10.384	.000	-0.41
	Grade 3	941724	32.76	26.388			
Familiar Words Reading	Grade 2	947037	14.95	15.732	-9.755	.000	-0.55
	Grade 3	941724	24.71	19.526			
Invented Words Reading	Grade 2	947037	12.49	12.530	-6.152	.000	-0.47
	Grade 3	941724	18.64	13.831			
Oral Reading Fluency	Grade 2	947037	17.17	19.111	-12.272	.000	-0.59
	Grade 3	941724	29.45	22.484			

Language -Berta

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	180649	19.03	66.570	0.272	.168	0.00
	Grade 3	190757	18.75	53.357			
Familiar Words Reading	Grade 2	180649	4.51	31.707	-2.869	.000	-0.08
	Grade 3	190757	7.38	37.675			
Invented Words Reading	Grade 2	180649	4.19	36.564	-3.835	.000	-0.10
	Grade 3	190757	8.02	41.518			
Oral Reading Fluency	Grade 2	180649	6.43	53.248	3.503	.000	-0.08
	Grade 3	190757	9.94	37.639			

Language -Hadiyyissa

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	376095	23.50	29.500	-14.349	.000	-0.46
	Grade 3	388068	37.85	33.256			
Familiar Words Reading	Grade 2	376095	4.62	10.211	-5.091	.000	-0.42
	Grade 3	388068	9.71	13.942			
Invented Words Reading	Grade 2	376095	4.11	11.103	-4.546	.000	-0.36
	Grade 3	388068	8.66	13.991			
Oral Reading Fluency	Grade 2	376095	4.65	13.627	-5.003	.000	-0.33
	Grade 3	388068	9.65	16.109			

Language -Nuer

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	236330	9.14	28.754	-7.189	.000	-0.29
	Grade 3	221161	16.33	20.684			
Familiar Words Reading	Grade 2	236330	3.43	13.230	-2.893	.000	-0.24
	Grade 3	221161	6.33	10.453			
Invented Words Reading	Grade 2	236330	2.60	10.257	-2.881	.000	-0.18
	Grade 3	221161	5.48	20.494			
Oral Reading Fluency	Grade 2	236330	8.01	25.183	-3.048	.000	-0.13
	Grade 3	221161	11.05	19.965			

Language -Sidaamu Afoo

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	142750	29.28	26.713	-15.074	.000	-0.53
	Grade 3	146578	44.35	30.024			
Familiar Words Reading	Grade 2	142750	6.07	10.452	-6.251	.000	-0.49
	Grade 3	146578	12.32	14.591			
Invented Words Reading	Grade 2	142750	4.70	9.037	-4.677	.000	-0.42
	Grade 3	146578	9.37	12.723			
Oral Reading Fluency	Grade 2	142750	7.00	13.648	-7.409	.000	-0.46
	Grade 3	146578	14.41	18.410			

Language -Wolyttatto

	Grade	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Grade 2	435351	19.15	25.669	-9.319	.000	-0.35
	Grade 3	422851	28.47	27.075			
Familiar Words Reading	Grade 2	435351	5.00	11.664	-4.106	.000	-0.26
	Grade 3	422851	9.11	19.539			
Invented Words Reading	Grade 2	435351	4.03	10.463	-3.212	.000	-0.27
	Grade 3	422851	7.24	13.140			
Oral Reading Fluency	Grade 2	435351	5.78	15.972	-4.560	.000	-0.25
	Grade 3	422851	10.33	19.734			

Appendix B. Timed Sub Tasks by Language and Gender

Language -Af Somali

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	102393	44.07	30.523	-2.980	.000	-0.12
	Male	101186	47.05	31.834			
Familiar Words Reading	Female	102393	13.55	12.631	0.730	.000	0.06
	Male	101186	12.82	13.475			
Invented Words Reading	Female	102393	16.53	21.736	0.980	.000	0.05
	Male	101186	15.55	20.944			
Oral Reading Fluency	Female	102393	24.57	25.844	-3.090	.000	-0.12
	Male	101186	27.66	24.853			

Language -Afan Oromo

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	1382139	39.14	27.624	-6.883	.000	-0.24
	Male	1633858	46.02	29.301			
Familiar Words Reading	Female	1382139	7.80	10.799	-3.627	.000	-0.28
	Male	1633858	11.42	14.179			
Invented Words Reading	Female	1382139	5.96	9.792	-2.689	.000	-0.23
	Male	1633858	8.65	13.182			
Oral Reading Fluency	Female	1382139	8.31	13.867	-4.666	.000	-0.29
	Male	1633858	12.98	17.682			

Language -Afar Af

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	28547	12.06	28.779	-6.737	.000	-0.21
	Male	38729	18.80	33.191			
Familiar Words Reading	Female	28547	1.57	7.088	-2.443	.000	-0.28
	Male	38729	4.01	9.937			
Invented Words Reading	Female	28547	1.63	7.216	-1.742	.000	-0.22
	Male	38729	3.37	8.594			
Oral Reading Fluency	Female	28547	1.92	9.959	-2.138	.000	-0.71
	Male	38729	4.05	13.697			

Language -Amharic

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	914690	27.06	26.216	-.956	.000	-0.04
	Male	974070	28.02	25.454			
Familiar Words Reading	Female	914690	19.70	18.548	-.224	.000	-0.01
	Male	974070	19.93	18.229			
Invented Words Reading	Female	914690	15.31	13.897	-.493	.000	-0.04
	Male	974070	15.80	13.208			
Oral Reading Fluency	Female	914690	23.21	22.208	-.159	.000	-0.01
	Male	974070	23.37	21.301			

Language -Berta

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	170976	18.53	64.518	-.658	.001	-0.01
	Male	200430	19.19	56.150			
Familiar Words Reading	Female	170976	7.80	41.148	3.378	.000	0.10
	Male	200430	4.43	28.482			
Invented Words Reading	Female	170976	5.00	38.197	-2.146	.000	-0.05
	Male	200430	7.15	40.070			
Oral Reading Fluency	Female	170976	11.85	59.289	6.874	.000	0.15
	Male	200430	4.97	29.818			

Language -Hadiyysissa

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	363729	27.16	28.875	-6.921	.000	-0.22
	Male	400434	34.08	34.745			
Familiar Words Reading	Female	363729	6.14	12.728	-2.016	.000	-0.16
	Male	400434	8.16	12.230			
Invented Words Reading	Female	363729	5.61	13.951	-1.545	.000	-0.12
	Male	400434	7.15	11.723			
Oral Reading Fluency	Female	363729	6.31	15.189	-1.672	.000	-0.11
	Male	400434	7.99	15.065			

Language -Nuer

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	208281	9.02	17.467	-6.610	.000	-0.26
	Male	249209	15.63	30.209			
Familiar Words Reading	Female	208281	3.58	11.766	-2.298	.000	-0.19
	Male	249209	5.88	12.194			
Invented Words Reading	Female	208281	2.93	10.267	-1.938	.000	-0.12
	Male	249209	4.87	19.659			
Oral Reading Fluency	Female	208281	8.14	22.461	-2.450	.000	-0.11
	Male	249209	10.59	23.130			

Language -Sidaamu Afoo

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	145942	30.98	27.028	-11.971	.000	-0.42
	Male	143387	42.95	30.507			
Familiar Words Reading	Female	145942	6.71	10.590	-5.097	.000	-0.41
	Male	143387	11.81	14.794			
Invented Words Reading	Female	145942	5.25	9.754	-3.661	.000	-0.33
	Male	143387	8.91	12.418			
Oral Reading Fluency	Female	145942	7.89	13.963	-5.776	.000	-0.35
	Male	143387	13.67	18.556			

Language -Wolyttatto

	Gender	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Female	415195	19.97	24.797	-7.301	.000	-0.28
	Male	443006	27.27	28.057			
Familiar Words Reading	Female	415195	4.64	10.133	-4.627	.000	-0.29
	Male	443006	9.27	19.991			
Invented Words Reading	Female	415195	3.81	9.853	-3.495	.000	-0.29
	Male	443006	7.30	13.434			
Oral Reading Fluency	Female	415195	5.53	15.777	-4.832	.000	-0.27
	Male	443006	10.36	19.695			

Appendix C. Timed Sub Tasks by Language and Location

Language -Af Somali

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	101724	44.64	31.878	-1.815	0.013	-0.06
	Urban	101855	46.46	30.516			
Familiar Words Reading	Rural	101724	18.52	27.076	4.758	.000	0.22
	Urban	101855	13.76	14.941			
Invented Words Reading	Rural	101724	17.16	23.307	2.238	.000	0.10
	Urban	101855	14.93	19.135			
Oral Reading Fluency	Rural	101724	27.25	25.350	2.285	.000	0.09
	Urban	101855	24.96	25.405			

Language -Afan Oromo

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	2695147	42.42	28.894	-4.143	.000	-0.14
	Urban	320850	46.57	27.232			
Familiar Words Reading	Rural	2695147	9.59	12.894	-1.601	.000	-0.13
	Urban	320850	11.19	12.569			
Invented Words Reading	Rural	2695147	7.43	12.011	0.161	.000	0.01
	Urban	320850	7.27	10.143			
Oral Reading Fluency	Rural	2695147	10.66	16.136	-1.629	.000	-0.10
	Urban	320850	12.29	16.786			

Language -Afar Af

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	60570	15.52	31.695	-4.242	.000	-0.13
	Urban	6706	19.76	30.165			
Familiar Words Reading	Rural	60570	2.82	8.854	-1.488	.000	-0.17
	Urban	6706	4.31	9.419			
Invented Words Reading	Rural	60570	2.51	7.944	-1.264	.000	-0.16
	Urban	6706	3.77	9.176			
Oral Reading Fluency	Rural	60570	2.88	11.229	-2.713	.000	-0.22
	Urban	6706	5.59	19.276			

Language -Amharic

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	1581815	24.85	24.730	-16.665	.000	-0.66
	Urban	306945	41.51	26.858			
Familiar Words Reading	Rural	1581815	18.04	17.528	-10.972	.000	-0.61
	Urban	306945	29.01	19.891			
Invented Words Reading	Rural	1581815	14.31	13.062	-7.664	.000	-0.58
	Urban	306945	21.98	14.179			
Oral Reading Fluency	Rural	1581815	21.05	20.519	-13.774	.000	-0.65
	Urban	306945	34.83	24.104			

Language -Berta

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	264860	14.81	44.437	-14.190	.000	-0.24
	Urban	106546	29.00	86.941			
Familiar Words Reading	Rural	264860	4.75	29.939	-4.292	.000	-0.12
	Urban	106546	9.04	44.851			
Invented Words Reading	Rural	264860	4.19	32.209	-6.849	.000	-0.18
	Urban	106546	11.04	52.472			
Oral Reading Fluency	Rural	264860	5.82	38.876	-8.087	.000	-0.18
	Urban	106546	13.91	59.588			

Language -Hadiyyissa

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	600910	31.39	32.342	2.825	.000	0.09
	Urban	163253	28.56	31.908			
Familiar Words Reading	Rural	600910	6.89	11.475	-1.454	.000	-0.12
	Urban	163253	8.34	15.693			
Invented Words Reading	Rural	600910	5.90	10.214	-2.439	.000	-0.19
	Urban	163253	8.34	19.617			
Oral Reading Fluency	Rural	600910	6.59	13.255	-2.791	.000	-0.18
	Urban	163253	9.38	20.521			

Language -Nuer

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	285009	13.58	22.602	2.560	.000	0.10
	Urban	172482	11.02	29.453			
Familiar Words Reading	Rural	285009	5.11	11.973	0.736	.000	0.06
	Urban	172482	4.37	12.177			
Invented Words Reading	Rural	285009	4.49	19.618	1.316	.000	0.08
	Urban	172482	3.17	7.151			
Oral Reading Fluency	Rural	285009	8.65	21.888	-2.196	.000	-0.10
	Urban	172482	10.85	24.321			

Language -Sidaamu Afoo

	Loca1	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	256844	37.58	29.765	5.946	.000	0.20
	Urban	32485	31.64	25.937			
Familiar Words Reading	Rural	256844	9.33	13.332	0.760	.000	0.06
	Urban	32485	8.57	11.036			
Invented Words Reading	Rural	256844	7.09	11.428	0.191	.000	0.02
	Urban	32485	6.90	10.262			
Oral Reading Fluency	Rural	256844	10.87	16.771	1.025	.000	0.06
	Urban	32485	9.84	15.662			

Language -Wolyttatto

	Location	N	Mean	SD	MD	P	Cohen's d
Letter Name Recognition	Rural	712798	24.03	26.716	1.712	.000	0.06
	Urban	145403	22.32	27.046			
Familiar Words Reading	Rural	712798	6.92	16.375	-0.656	.000	-0.04
	Urban	145403	7.57	15.086			
Invented Words Reading	Rural	712798	5.35	11.164	-1.562	.000	-0.13
	Urban	145403	6.91	15.235			
Oral Reading Fluency	Rural	712798	7.69	16.720	-1.970	.000	-0.11
	Urban	145403	9.66	23.523			

Appendix D. Untimed Sub Tasks by Language and Grade

Language -Af Somali

	Grade	N	Mean	MD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	170715	82.83	25.920	3.199	.000	0.10
	Grade 3	180487	81.73	27.551			
Listening Comprehension	Grade 2	170715	66.45	35.897	-10.714	.000	-0.24
	Grade 3	180487	75.56	29.790			
Reading Comprehension	Grade 2	170715	19.03	31.790	-8.532	.000	-0.32
	Grade 3	180487	28.30	34.343			

Language -Afan Oromo

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	1464930	39.14	41.399	-23.837	.000	-0.58
	Grade 3	1551067	62.97	40.120			
Reading Comprehension	Grade 2	1464930	6.70	18.412	-12.561	.000	-0.52
	Grade 3	1551067	19.26	28.589			
Listening Comprehension	Grade 2	1464930	64.72	30.970	-10.748	.000	-0.38
	Grade 3	1551067	75.47	25.641			

Language -Afar Af

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	35313	15.75	27.850	-7.840	.000	-0.25
	Grade 3	31963	23.59	34.299			
Listening Comprehension	Grade 2	35313	50.94	40.862	-5.555	.000	-0.14
	Grade 3	31963	56.49	39.787			
Reading Comprehension	Grade 2	35313	1.13	6.056	-2.327	.000	-0.26
	Grade 3	31963	3.46	11.109			

Language –Amharic

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	947037	72.60	34.006	-7.500	.000	-0.23
	Grade 3	941724	80.10	29.783			
Listening Comprehension	Grade 2	947037	69.34	26.875	-4.412	.000	-0.17
	Grade 3	941724	73.75	25.318			
Reading Comprehension	Grade 2	947037	16.37	24.026	-14.777	.000	-0.54
	Grade 3	941724	31.15	30.661			

Language -Berta

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	180649	31.16	34.911	-8.843	.000	-0.25
	Grade 3	190757	40.01	36.226			
Listening Comprehension	Grade 2	180649	55.85	31.129	-6.312	.000	-0.20
	Grade 3	190757	62.16	32.958			
Reading Comprehension	Grade 2	180649	2.10	11.577	-0.539	.000	-0.04
	Grade 3	190757	2.64	12.993			

Language -Hadiyyissa

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	376095	73.75	29.064	-4.888	.000	-0.18
	Grade 3	388068	78.63	25.760			
Listening Comprehension	Grade 2	376095	74.21	28.350	-7.263	.000	-0.27
	Grade 3	388068	81.47	25.836			
Reading Comprehension	Grade 2	376095	4.78	13.258	-4.142	.000	-0.28
	Grade 3	388068	8.92	16.530			

Language -Nuer

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	236330	30.30	33.812	-24.311	.000	-0.69
	Grade 3	221161	54.61	37.162			
Reading Comprehension	Grade 2	236330	5.52	14.799	-9.049	.000	-0.44
	Grade 3	221161	14.57	25.037			
Listening Comprehension	Grade 2	236330	76.78	27.289	-3.669	.000	-0.14
	Grade 3	221161	80.45	25.883			

Language -Sidaamu Afoo

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	142750	81.98	20.548	-3.617	.000	-0.20
	Grade 3	146578	85.60	15.447			
Listening Comprehension	Grade 2	142750	75.77	26.400	-6.694	.000	-0.28
	Grade 3	146578	82.46	22.055			
Reading Comprehension	Grade 2	142750	6.18	15.162	-7.614	.000	-0.40
	Grade 3	146578	13.79	21.965			

Language -Wolyttatto

	Grade	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Grade 2	435351	60.01	37.941	-9.350	.000	-0.26
	Grade 3	422851	69.36	33.961			
Listening Comprehension	Grade 2	435351	47.10	24.538	-2.878	.000	-0.11
	Grade 3	422851	49.98	26.868			
Reading Comprehension	Grade 2	435351	6.27	20.753	-4.924	.000	-0.22
	Grade 3	422851	11.19	24.047			

Appendix E. Untimed Sub Tasks by Language and Gender

Language -Af Somali

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	102393	74.32	31.152	-2.250	.000	-0.10
	Male	101186	76.57	31.083			
Listening Comprehension	Female	102393	66.57	35.851	-7.694	.000	-0.21
	Male	101186	74.26	31.815			
Reading Comprehension	Female	102393	20.53	33.109	-7.956	.000	-0.20
	Male	101186	28.49	36.560			

Language -Afan Oromo

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	1382139	44.50	41.314	-12.735	.000	-0.30
	Male	1633858	57.23	42.526			
Listening Comprehension	Female	1382139	68.03	29.620	-4.086	.000	-0.14
	Male	1633858	72.12	28.064			
Reading Comprehension	Female	1382139	9.32	20.429	-7.094	.000	-0.29
	Male	1633858	16.41	27.866			

Language -Afar Af

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	28547	15.73	27.124	-6.51	.000	-0.21
	Male	38729	22.24	33.829			
Listening Comprehension	Female	28547	50.90	40.656	-4.65	.000	-0.12
	Male	38729	55.55	40.184			
Reading Comprehension	Female	28547	1.58	7.903	-1.13	.000	-0.13
	Male	38729	2.71	9.543			

Language -Amharic

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	914690	75.26	33.390	-2.096	.000	-0.06
	Male	974070	77.35	30.986			
Listening Comprehension	Female	914690	70.40	26.815	-2.219	.000	-0.08
	Male	974070	72.62	25.568			
Reading Comprehension	Female	914690	24.16	29.722	0.806	.000	0.03
	Male	974070	23.35	27.315			

Language -Berta

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	170976	37.20	35.997	2.766	.000	0.08
	Male	200430	34.43	35.704			
Listening Comprehension	Female	170976	58.99	32.620	-.176	.097	-0.01
	Male	200430	59.17	31.905			
Reading Comprehension	Female	170976	2.32	12.479	-.109	.007	-0.01
	Male	200430	2.43	12.197			

Language -Hadiyyissa

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	363729	74.09	28.596	-4.075	.000	-0.15
	Male	400434	78.17	26.403			
Listening Comprehension	Female	363729	76.59	27.248	-2.486	.000	-0.09
	Male	400434	79.08	27.378			
Reading Comprehension	Female	363729	6.02	14.802	-1.659	.000	-0.11
	Male	400434	7.67	15.419			

Language -Nuer

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	208281	35.13	35.233	-12.697	.000	-0.34
	Male	249209	47.83	38.338			
Reading Comprehension	Female	208281	6.03	15.959	-7.101	.000	-0.34
	Male	249209	13.13	23.784			
Listening Comprehension	Female	208281	76.78	28.081	-3.267	.000	-0.12
	Male	249209	80.04	25.357			

Language -Sidaamu Afoo

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	145942	82.74	20.006	-2.176	.000	-0.12
	Male	143387	84.91	16.158			
Listening Comprehension	Female	145942	77.61	25.525	-3.125	.000	-0.13
	Male	143387	80.74	23.359			
Reading Comprehension	Female	145942	7.41	16.114	-5.298	.000	-0.28
	Male	143387	12.71	21.745			

Language -Wolyttatto

	Gender	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Female	415195	63.01	36.521	-3.105	.000	-0.09
	Male	443006	66.12	36.099			
Listening Comprehension	Female	415195	47.60	26.140	-1.784	.000	-0.07
	Male	443006	49.39	25.354			
Reading Comprehension	Female	415195	5.44	17.649	-6.307	.000	-0.28
	Male	443006	11.75	25.995			

Appendix F. Untimed Sub Tasks by Language and Location

Language -Af Somali

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	101724	70.32	35.353	-10.248	.000	-0.36
	Urban	101855	80.56	25.242			
Reading Comprehension	Rural	101724	26.64	36.035	4.311	.000	0.14
	Urban	101855	22.33	33.990			
Listening Comprehension	Rural	101724	72.08	32.937	3.367	.000	0.12
	Urban	101855	68.71	35.186			

Language -Afan Oromo

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	2695147	51.61	42.283	2.055	.000	0.05
	Urban	320850	49.56	43.803			
Reading Comprehension	Rural	2695147	12.92	24.881	-2.279	.000	-0.09
	Urban	320850	15.20	25.780			
Listening Comprehension	Rural	2695147	69.94	29.037	-2.918	.000	-0.10
	Urban	320850	72.86	27.183			

Language -Afar Af

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	60570	19.30	31.139	-1.838	.000	-0.06
	Urban	6706	21.13	32.932			
Reading Comprehension	Rural	60570	2.07	8.584	-1.635	.000	-0.18
	Urban	6706	3.71	11.266			
Listening Comprehension	Rural	60570	53.57	40.456	-0.099	.850	0.00
	Urban	6706	53.67	40.394			

Language -Amharic

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	1581815	74.31	33.298	-12.461	.000	-0.39
	Urban	306945	86.77	23.063			
Reading Comprehension	Rural	1581815	20.91	26.942	-17.424	.000	-0.63
	Urban	306945	38.33	31.720			
Listening Comprehension	Rural	1581815	69.43	26.512	-12.997	.000	-0.50
	Urban	306945	82.43	21.476			

Language -Berta

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	264860	32.48	35.775	-11.247	.000	-0.32
	Urban	106546	43.73	34.819			
Reading Comprehension	Rural	264860	2.05	11.388	-1.124	.000	-0.09
	Urban	106546	3.18	14.368			
Listening Comprehension	Rural	264860	59.13	32.022	0.149	.201	0.00
	Urban	106546	58.98	32.762			

Language -Hadiyyissa

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	600910	77.86	27.127	7.644	.000	0.28
	Urban	163253	70.22	28.225			
Reading Comprehension	Rural	600910	6.59	14.701	-1.372	.000	-0.09
	Urban	163253	7.96	16.660			
Listening Comprehension	Rural	600910	78.92	26.807	4.786	.000	0.18
	Urban	163253	74.13	28.930			

Nuer

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	285009	46.79	36.830	12.588	.000	0.34
	Urban	172482	34.21	37.265			
Reading Comprehension	Rural	285009	9.52	19.889	-1.008	.000	-0.05
	Urban	172482	10.52	22.445			
Listening Comprehension	Rural	285009	74.05	29.168	-11.947	.000	-0.46
	Urban	172482	86.00	19.838			

Sidaamu Afoo

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	256844	83.64	18.452	-1.573	.000	-0.09
	Urban	32485	85.21	16.336			
Reading Comprehension	Rural	256844	9.98	19.214	-0.480	.000	-0.02
	Urban	32485	10.46	19.929			
Listening Comprehension	Rural	256844	78.86	24.773	-2.660	.000	-0.11
	Urban	32485	81.52	22.334			

Wolyttatto

	Location	N	Mean	SD	MD	P	Cohen's d
Phonemic Awareness	Rural	712798	63.59	36.867	-6.040	.000	-0.17
	Urban	145403	69.63	33.163			
Reading Comprehension	Rural	712798	8.15	21.526	-3.235	.000	-0.14
	Urban	145403	11.38	26.955			
Listening Comprehension	Rural	712798	47.83	25.624	-4.089	.000	-0.16
	Urban	145403	51.92	26.109			

Appendix G. Multiple Comparisons of students performance by program types

Dependent Variable			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Letter Name Recognition	NEGRA	GEQIP-E	-11.145*	0.208	0.000	-11.63	-10.66
		READ II	-6.324*	0.220	0.000	-6.84	-5.81
	GEQIP-E	NEGRA	11.145*	0.208	0.000	10.66	11.63
		READ II	4.821*	0.281	0.000	4.16	5.48
	READ	NEGRA	6.324*	0.220	0.000	5.81	6.84
		GEQIP-E	-4.821*	0.281	0.000	-5.48	-4.16
Familiar Words Reading	NEGRA	GEQIP-E	-14.007*	0.142	0.000	-14.34	-13.67
		READ II	-14.195*	0.150	0.000	-14.55	-13.84
	GEQIP-E	NEGRA	14.007*	0.142	0.000	13.67	14.34
		READ II	-0.187	0.192	0.593	-0.64	0.26
	READ	NEGRA	14.195*	0.150	0.000	13.84	14.55
		GEQIP-E	0.187	0.192	0.593	-0.26	0.64
Invented Words Reading	NEGRA	GEQIP-E	-7.790*	0.141	0.000	-8.12	-7.46
		READ II	-9.479*	0.149	0.000	-9.83	-9.13
	GEQIP-E	NEGRA	7.790*	0.141	0.000	7.46	8.12
		READ II	-1.689*	0.191	0.000	-2.14	-1.24
	READ	NEGRA	9.479*	0.149	0.000	9.13	9.83
		GEQIP-E	1.689*	0.191	0.000	1.24	2.14
Oral Reading Fluency	NEGRA	GEQIP-E	-0.228	0.170	0.373	-0.63	0.17
		READ II	-6.914*	0.180	0.000	-7.33	-6.49
	GEQIP-E	NEGRA	0.228	0.170	0.373	-0.17	0.63
		READ II	-6.686*	0.230	0.000	-7.22	-6.15
	READ	NEGRA	6.914*	0.180	0.000	6.49	7.33
		GEQIP-E	6.686*	0.230	0.000	6.15	7.22
Phonemic Awareness	NEGRA	GEQIP-E	15.285*	0.206	0.000	14.80	15.77
		READ II	3.535*	0.218	0.000	3.02	4.05
	GEQIP-E	NEGRA	-15.285*	0.206	0.000	-15.77	-14.80
		READ II	-11.751*	0.279	0.000	-12.40	-11.10
	READ	NEGRA	-3.535*	0.218	0.000	-4.05	-3.02
		GEQIP-E	11.751*	0.279	0.000	11.10	12.40
Reading Comprehension	NEGRA	GEQIP-E	-1.945*	0.235	0.000	-2.50	-1.39
		READ II	-6.320*	0.249	0.000	-6.90	-5.74
	GEQIP-E	NEGRA	1.945*	0.235	0.000	1.39	2.50
		READ II	-4.375*	0.318	0.000	-5.12	-3.63
	READ	NEGRA	6.320*	0.249	0.000	5.74	6.90
		GEQIP-E	4.375*	0.318	0.000	3.63	5.12
Listening Comprehension	NEGRA	GEQIP-E	3.994*	0.228	0.000	3.46	4.53
		READ II	-7.313*	0.242	0.000	-7.88	-6.75
	GEQIP-E	NEGRA	-3.994*	0.228	0.000	-4.53	-3.46
		READ II	-11.307*	0.309	0.000	-12.03	-10.58
	READ	NEGRA	7.313*	0.242	0.000	6.75	7.88
		GEQIP-E	11.307*	0.309	0.000	10.58	12.03
*. The mean difference is significant at the 0.05 level.							

Appendix H. Standard set or Benchmark Lebels by Language

Language	Grade	Students Who are Reading		
		With limited fluency and comprehension	With increasing fluency and comprehension	Fluently and with full comprehension
Afan Oromo	Grade 2	1 to 19	20 to 47	48 and above
	Grade 3	1 to 29	30 to 57	58 and above
Afar Af	Grade 2	1 to 18	19 to 45	46 and above
	Grade 3	1 to 23	24 to 50	51 and above
Af Somali	Grade 2	1 to 24	25 to 49	50 and above
	Grade 3	1 to 24	25 to 54	55 and above
Amharic	Grade 2	1 to 29	30 to 49	50 and above
	Grade 3	1 to 34	35 to 59	60 and above
Berta	Grade 2	1 to 15	16 to 44	45 and above
	Grade 3	1 to 20	21 to 49	50 and above
Hadiyyissa	Grade 2	1 to 24	25 to 39	40 and above
	Grade 3	1 to 24	25 to 49	50 and above
Nuer	Grade 2	1 to 19	20 to 44	45 and above
	Grade 3	1 to 24	25 to 47	48 and above
Sidamu Afoo	Grade 2	1 to 19	20 to 44	45 and above
	Grade 3	1 to 24	25 to 52	53 and above
Wolayttatto	Grade 2	1 to 19	20 to 42	43 and above
	Grade 3	1 to 24	25 to 51	52 and above
Tigray	Grade 2	1 to 19	20 to 54	55 and above
	Grade 3	1 to 24	25 to 61	62 and above

Reference

AIR (2018). Early Grade Reading Assessment (EGRA) End line Report

Anteneh, G., Ferede, T., Kelemework, Y., Mikre, F., & Getachew, K. (2016). Early Grade Reading Assessment in the East Wollega Zone of Oromiya: A Study on Mother Tongue Reading Competence of Grade 4 Pupils. *International Journal of Sciences, 27*(3), 20.

Brewer, E. (2009). Conducting survey research in education: DOI:10.4018/978-1-60566-739 3.ch041

Christensen, L., Johnson, B., & Turner, L. (2013). *Research methods, design, and analysis* (12th ed). Boston, Munich: Pearson.

Daniel, M. (2004). Doing Quantitative Research in Education with SPSS, SAGE Publications India Pvt Ltd New Delhi.

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development, 40*, 315–322. https://doi.org/10.1016/j.ijedudev.2014.11.004

Gay,L., Mills, G. & Airasian, P. (20012). Educational Research: Competencies for Analysis and Applications, 10th ed.

IEA & NEAEA (2021.) Responses to Educational Disruption Survey (REDS): The Impact of COVID-19 in Ethiopia.

McMillan, J., & Schumacher, S. (2006). Research in Education: Evidence-based Inquiry (6th ed.). Cape Town: Pearson.

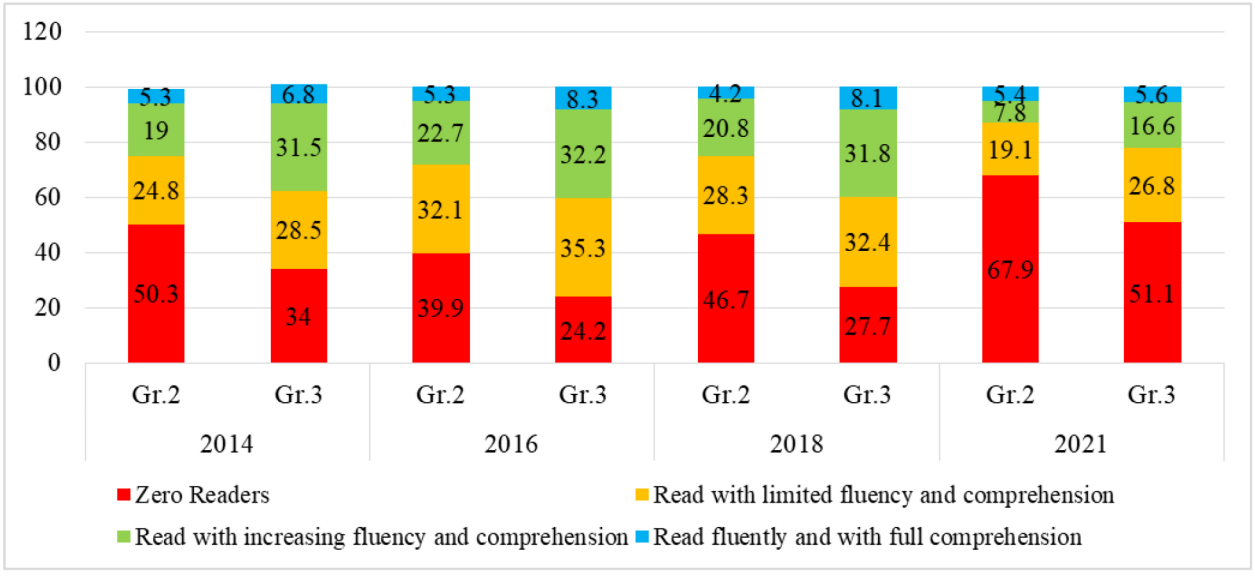
Ross, K. (2007). Sample design for educational survey research, UNESCO International Institute for Educational Planning, Module 3, Paris.

RTI (2016). *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition.* Washington, DC: United States Agency for International Development.

Wolf, F. (1986). Meta-analysis: Quantitative methods for research synthesis. Beverly Hills, CA: Sage.



2021


120
100
80
60
40
20
0
5.3
19
24.8
50.3
6.8
31.5
28.5
34
5.3
22.7
32.1
39.9
8.3
32.2
35.3
24.2
4.2
20.8
28.3
46.7
8.1
31.8
32.4
27.7
5.4
7.8
19.1
67.9
5.6
16.6
26.8
51.1
Gr.2
Gr.3
Gr.2
Gr.3
Gr.2
Gr.3
Gr.2
Gr.3
2014
2016
2018
2021
Zero Readers
Read with limited fluency and comprehension
Read with increasing fluency and comprehension
Read fluently and with full comprehension