



REPUBLIC OF MOZAMBIQUE

MINISTRY OF EDUCATION AND HUMAN DEVELOPMENT



USAID
FROM THE AMERICAN PEOPLE

Effectiveness Evaluation

Early Grade Reading Assessment & Supplementary Tools

Midline Report - Mozambique

March, 2020

This report is made possible by the support of the American People through the United States Agency for International Development (USAID). The contents of this baseline report are the sole responsibility of Creative Associates International and World Education, Inc. and do not necessarily reflect the views of USAID or the United States Government.

This report was prepared by Adam Turney, David Noyes, Flávio Magaia, Trymore Mafucha Dhliwayo, Yuri Machkasov, Euclides Chigogolo Dinis Zacarias, and Haiyan Hua from World Education, Inc.

Contents

List of Figures.....	3
List of Tables.....	3
List of Annexes.....	4
Acknowledgements.....	5
Acronyms.....	6
EXECUTIVE SUMMARY	7
I Introduction	10
1.1 USAID Vamos Ler! Description.....	10
1.2 Purpose of the Midline Effectiveness Evaluation.....	10
1.3 Effectiveness Evaluation Instruments.....	11
1.3.1 Early Grade Reading Assessment (EGRA).....	11
1.3.2 Supplementary Instruments.....	12
2 Design and Methodology	14
2.1 Vamos Ler! Effectiveness Evaluation Design	14
2.2 MEE Implementation Plan	17
2.3 Training for Data Collection	18
2.4 Summary of Field Work.....	20
2.5 Data Quality Control Approach	22
2.5.1 Assessor Accuracy Measurements.....	22
2.5.2 Inter-Rater Reliability	22
2.5.3 Supervision of Data Collection.....	23
2.5.4 Data Cleaning and Real-time Quality Control.....	23
2.5.5 Data Analysis	23
2.6 Limitations of the Study.....	23
3 MEE Results.....	24
3.1 Student Performance by EGRA subtask.....	24
3.1.1 Baseline to Midline Comparison.....	24
3.1.2 EGRA Results by Student Sex.....	26
3.2 Student Characteristics and EGRA Results	27
3.3 Teacher Factors and EGRA Results	28
3.3.1 Teacher Characteristics.....	28
3.3.2 Teacher Classroom Performance.....	30
3.3.3 Teacher LI and School LI Match	32
3.3.4 Teacher ORF and Comprehension.....	32
3.4 School Management Assessment and EGRA Results	33
3.4.1 School Director Profile.....	33
3.4.2 Attendance and Punctuality.....	34
3.4.3 School Management Performance	36

4 Results by District.....	38
5 Qualitative Follow-Up	41
5.1 Rationale	41
5.2 Participant Profiles.....	41
5.3 Summary Responses from Target Groups & Domains	42
5.4 Responses Based on School Performance.....	44
5.4.1 Emerging Trends Among High-Performing Schools.....	44
5.4.2 Emerging Trends Among Low-Performing Schools.....	45
5.5 Individual Cases	46
5.5.1 Example of a Low-Performing School.....	46
5.5.2 Example of a High-Performing School.....	47
6 Conclusions and Recommendations.....	48
6.1 Key Findings	48
6.2 Recommendations	50
Bibliography	52

List of Figures

Figure 1: Sample Size Considerations	17
Figure 2: Student Age Distribution Grade 2	27
Figure 3: Out of School Reading Support Index.....	28
Figure 4: Classroom Observation Index	31
Figure 5: Teacher ORF Distribution	33
Figure 6: Attendance – School Directors, Teachers, Students.....	35
Figure 7: Student Absentee Rates.....	36
Figure 8: School Management Index.....	38

List of Tables

Table 1: EGRA Tool Revision	11
Table 2: EGRA Tool Subtasks for the MEE (all G2).....	12
Table 3: FGD/Interview Script Domains.....	14
Table 4: Data Workshop/Qualitative Training Participants	19
Table 5: Participation List for MEE Quantitative Field Data Collection	20
Table 6: Quantitative Field Data Collection Summary	21
Table 7: Planned vs Completed Quantitative Field Data Collection Summary	21
Table 8: Qualitative Field Data Collection Summary.....	22
Table 9: Letter Name Identification Subtask – Student Percentage Comparison Baseline to Midline by Performance Band	24

Table 10: Letter Sounds Identification Subtask – Student Percentage Comparison Baseline to Midline by Performance Band.....	25
Table 11: Nonsense Word Reading Subtask – Student Percentage Comparison Baseline to Midline by Performance Band	25
Table 12: LI Oral Reading Fluency Subtask – Student Percentage Comparison Baseline to Midline by Performance Band	25
Table 13: EGRA Results by Sex – Including Percentage by which Boys Scored Higher than Girls.	26
Table 14: Student Contextual Interview	27
Table 15: Teacher Characteristics.....	29
Table 16: Student Performance by Student and Teacher Sex	30
Table 17: Teacher Performance on Observation Checklist.....	30
Table 18: Teacher LI Match and ORF Score Impact.....	32
Table 19: Teacher ORF and Comprehension.....	33
Table 20: School Director Profile.....	34
Table 21: School Management Assessment Results	37
Table 22: Results by District – Student-Level.....	39
Table 23: Results by District – Teacher and School-Level	40
Table 24: Qualitative Data Collection Participants	41
Table 25: Summary Discussion Responses by Group & Domain.....	42
Table 26: Programmatic Recommendations for <i>Vamos Ler!</i>	50

List of Annexes

- Annex A: EGRA Tool – Grade 2, Echuwabo
- Annex B: EGRA Tool – Grade 2, Elomwe
- Annex C: EGRA Tool – Grade 2, Emakhuwa
- Annex D: Teacher LI Fluency Assessment – Echuwabo
- Annex E: Teacher LI Fluency Assessment – Elomwe
- Annex F: Teacher LI Fluency Assessment – Emakhuwa
- Annex G: Supplementary Tool – Classroom Observation
- Annex H: Supplementary Tool – School Management
- Annex I: Focus Group Discussion Tool - Community
- Annex J: Focus Group Discussion Tool - Teachers
- Annex K: Interview Tool – School Directors

Acknowledgements

The authors wish to thank the many individuals, organizational partners, and government institutions who made this study possible, both in Mozambique and the US.

Special thanks goes to USAID|Mozambique for their ongoing support and who ultimately make this work possible. They continue to show their commitment year after year to improving education in Mozambique through programs like USAID Vamos Ler!/ Let's Read!.

Thank you to all our partners within the government of Mozambique, particularly within the Ministry of Education and Human Development (MINEDH) and its provincial and district officers who were involved in the data collection and data interpretation workshops. Without their never-ending commitment to improving outcomes for students in Mozambique these types of collaborations would not be possible.

Clearly without the many months, weeks, days, and early morning efforts of the entire **Vamos Ler!** staff in Maputo, Nampula, and Quelimane, this study would not have been realized. From tool revision to data collection through to the final product, teams of staff overcame daily challenges to help answer the questions this research aims to address.

Lastly, and certainly not least, thank you to the school councils, school directors, teachers, parents, and students throughout Nampula and Zambézia for welcoming our teams, assisting with the data collection and being patient throughout the surveys and observations.

Acronyms

AAM	Assessor Accuracy Measures
ADE	Direct Support to Schools (Apoio Directo às Escolas)
ANOVA	Analysis of Variance
ApaL	Aprender a Ler (Learn to Read)
BE	Bilingual Education
SD	School Director
DSD	Assistant School Director
DO	District Officer
DPEDH	Provincial Directorate of Education and Human Development
EE	Effectiveness Evaluation
EGR	Early Grade Reading
EGRA	Early Grade Reading Assessment
FGD	Focus Group Discussion
G1/G2/G3	Grade One / Grade Two / Grade Three
HLM	Hierarchical Linear Modeling
ICC	Intraclass Correlation Coefficient
IFP	Teacher Training Institutes (<i>Institutos de Formação de Professores</i>)
INDE	National Institute for the Development of Education
INSET	In-Service Teacher Training
L1	Language One (local language)
L2	Language Two (Portuguese)
LEMA	Local Education Monitoring Approach
MEE	Midline Effectiveness Evaluation
M&E	Monitoring and Evaluation
MINEDH	Ministry of Education and Human Development
NGO	Non-Governmental Organization
ORF	Oral Reading Fluency
PB	Performance Band
PD	Pedagogical Director
PMEP	Project Monitoring and Evaluation Plan
SC	School Council
SD	School Director
SDEJT	District Education Offices (<i>Servico de Educação, Juventude e Tecnologia</i>)
SPD	Schools Per District
UEM	Universidade Eduardo Mondlane
UP	Pedagogical University (<i>Universidade Pedagogica</i>)
USAID	United States Agency for International Development
WEI	World Education, Inc.
ZIP	School Clusters (<i>Zona de influencia pedagogica</i>)

EXECUTIVE SUMMARY

After two years of implementation in the schools, USAID *Vamos Ler!* conducted a midline evaluation with the goal of assessing student learning progress made to date and identifying factors influencing student learning outcomes so that final strategy adjustments could be made prior to the final full year of implementation. The evaluation covered all 21 program districts in Nampula and Zambézia provinces in a total of 536 schools. The evaluation assessed the early grade reading skills of 10,801 second grade students, oral reading fluency and classroom practices of 478 second grade teachers, and school management practices in each of the schools. The evaluation included qualitative follow-up visits based on preliminary quantitative findings in 15 high-performing and 14 low-performing schools. These visits were conducted to validate the initial quantitative findings observed and allowed documentation of additional factors may influence student learning outcomes that were not captured by the quantitative piece of the evaluation.

The field work followed a rigorous process that began with a highly competitive enumerator recruitment, training, evaluation, and selection. The field work, preliminary data analysis and interpretation, and qualitative follow-up visits were completed during the period of 02 September to 04 October of 2019. Data cleaning was conducted simultaneously during the data collection, allowing detection of any issues during the collection and timely production of the quantitative results to facilitate selection of high and low-performing schools immediately following the end of the quantitative part of the evaluation. About 80% of the student sample, nearly 90% of the teacher sample, and 100% of the school management sample was achieved. The full sample was not achieved due to high levels of absenteeism, including non-functioning schools (no lessons given on the day visit due to absenteeism of teachers and directors) in 8% of the schools. In 31% of the schools, the school director was absent. In 11% of the schools, no G2 grade teacher was present. In 8% of the schools, fewer than 10 students were present. This resulted in several challenges, including scheduling return visits to several schools that were found non-functioning or where fewer than 10 students were present on the day of the visit.

In addition to conducting frequency analyses and disaggregation of results by language, grade, and sex, additional analysis focused on comparison of the midline results to the baseline and multiple regression analysis to identify key factors influencing student learning outcomes. We summarize here the key findings noted in the quantitative analysis and further confirmed through the qualitative follow-up visits.

1. **Improved student reading skills compared to baseline:** Compared to the program baseline, student performance improved substantially for all EGRA subtasks. The was true for all three Mozambican languages assessed. The improvements were in large part due to a substantial reduction in the percentage of students registering zero scores on the EGRA subtasks. The largest improvements were seen in letter name and letter sound identification subtasks, with more modest gains in word and text reading.
2. **Positive reception of bilingual education in classrooms and communities:** A clear finding from the classroom observations and further validated in the qualitative follow-up visits was the very visible improvements in teacher-student interactions and overall student participation brought about through the introduction of bilingual education. This has the potential to produce large gains in student learning outcomes, but we find several factors related to school management and relationships between the school and the community that limited the results overall.

3. **Reduction of instruction time through school closures, absenteeism and tardiness:** The evaluation noted continued high levels of school director, teacher, and student absenteeism and lateness resulting in substantial reductions of instructional time. A total of 41 schools (8% of the sample) visited were completely non-functional (no lessons given) on the day of the data collection visit. Both quantitative and qualitative results point to the high rates of student absenteeism (58% average student absentee rate overall) being the result of irregular teacher attendance. In addition to absenteeism, the suggest that a late start to the school day could result in as much as a 15% reduction in instruction time over the course of a year. Factoring in the high absentee rates and other factors such as delayed start of lessons after school holidays, the overall amount of exposure to reading instruction that students have is quite limited.
4. **Gender reading performance gap:** Boys outperformed girls on all EGRA subtasks, with the gaps in performance growing as the difficulty of the task increases (i.e. moving from letters to text reading). The gender performance gap also showed a regional variation, with districts in the south having a smaller gap compared to districts in the north, suggesting that cultural factors specific to the north (e.g. initiation rites and other traditions that are more common in the north) may help explain this gap. It was also observed that the gender performance gap was reduced when students had a female teacher. While this was encouraging, the fact that rural schools have very few female teachers limited the overall impact that the female teachers had in support greater equity in reading performance. We also noted that female teachers generally had better performance as measured through the classroom observation instrument.
5. **Identification of factors that influence student learning outcomes:** A high degree of variability was noted both within schools (students in the same school with same teacher, but drastically different outcomes) and between schools (average performance of one school drastically different from another school). Several factors were identified through multiple regression analysis as being significantly correlated with the student learning outcomes. The results (standardized coefficients of the factors) clearly point out that not only are these factors associated with the learning improvement, but also, they rank by the order of influence on student ORF scores as listed below.
 - a. **Out-of-school support**, including access to books at home, having someone at home to read with, participation in a reading club, and having regular support at home to complete homework assignments. Students with high levels of out of school support for reading, on average are reading **6.2 more letters and 2.7 more words** compared to students with low levels of support. Out-of-school support, being the most influential factor, inform us that early children's learning in literacy must have conducive learning environment at home and parental engagement in and support to their children. The fact that out-of-school support seems to be more critical in young children's literacy learning than school and classroom factors deserves strategic policy attention and program planning.
 - b. **Teacher classroom instruction practice** was observed to be largely positive, though with a large degree of variability. Female teachers generally performed better than their male counterparts, but all teachers had difficulties implementing some of the practices specific to reading instruction (students reading aloud, students reading together in pairs, asking students questions about the text they read, and discussing the meaning of words in the text that was read). We found

that students with high performing teachers, on average read **4.1 more letters and 1.4 more words** compared to students with low performing teachers.

- c. **Teacher oral reading fluency levels** showed a large degree of variability for all languages. The introduction of the program may have been the first time that many teachers began to read the Mozambican languages covered by the program, and while in some cases the teacher's LI did not match the LI of the school, this did not have a significant impact on student performance. A teacher's overall fluency level did have a significant impact, with students of teachers with high fluency levels, on average reading **3.0 more letters and 1.2 more words** compared to students of teachers with low fluency levels.
- d. **Implementation of effective school management practices** such as management of student attendance and provision of in-school pedagogical support proved the most difficult for schools. More regular provision of in-school coaching could help eliminate some of the variance in teacher performance as observed during the classroom observation and ORF assessments of teachers. We found that schools implementing more of the recommended school management practices had students who read on average **3.2 more letters and 1.1 more words** compared to low-performing schools.

With the sample design, we were able to conduct an analysis of the variance components in student learning outcomes, focusing specifically on ORF. We found out that 73% of the total variance in ORF is within individual students. In other words, the individual level variance is likely associated with unique individual characteristics – factors outside of the school environment and more related to parental support. The remaining variance lies in the school/classroom level (24%) and 4% at the district level, suggesting that teacher improvement program activities and school management enhancement may only impact 28% of the variance in student learning outcomes.

Based on the quantitative findings, and additional data from the qualitative follow-up visits, several recommendations for the program were developed and considered for integration into the 2020 program implementation strategy.

- The program should expand efforts to promote greater involvement of communities, e.g.:
 - Expand reading clubs
 - Strengthen the use of complementary reading material at home
 - Promote reading competitions and other events to stimulate community involvement
- Teachers should integrate practices to reduce differences in results for boys and girls
- Teachers should receive support to improve their oral reading fluency skills, with special focus on teachers who are not native speakers of the school's LI
- Teachers should place greater focus on building the decoding skills of students
- The program should find ways to address teacher motivation
- Teachers need to receive more in-school coaching
- Greater involvement of local authorities (administrators, district director, local leaders) as well as the provisional inspectorate to address absenteeism

I Introduction

1.1 USAID Vamos Ler! Description

USAID Vamos Ler!//Let's Read! is a five-year program, funded by the U.S. Agency for International Development (USAID), with the goal to strengthen the Mozambican government's ability to ensure that students in the target regions of Nampula and Zambézia can achieve grade-level fluency and comprehension in the local language in the first cycle of primary education (grades 1 – 3). **Vamos Ler!** provides evidence-based technical and material assistance to improve early grade literacy instruction in three local languages, Emakhuwa, Elomwe, and Echuwabo, along with second language acquisition and literacy support to prepare children for transition to Portuguese. Through a three-fold, results-based approach **Vamos Ler!** supports the Ministry of Education and Human Development (MINEDH) to strengthen classroom and systems support to lay the foundation for sustaining and expanding bilingual early grade literacy throughout the country. The program provides support in three core areas:

- Improved quality of instruction through provision of teaching and learning materials, teacher training and coaching, school director (SD) training and coaching in support of improved reading outcomes, and district supervision for fidelity of implementation;
- Improved support systems for reading improvement through the coordination and integration of programming within existing MINEDH structures and development of coordination mechanisms, support for evidence-based decision-making, and use of an improved Early Grade Reading Assessment (EGRA); and
- Improved community support for reading improvement through mobilization of communities in support of bilingual education (BE), strengthening of school councils, and greater involvement of parents and community members in their child's learning process.

The program will reach 1,950 schools, reaching over 600,000 children, and training more than 10,000 teachers and school directors by 2021.

1.2 Purpose of the Midline Effectiveness Evaluation

The Midline Effectiveness Evaluation (MEE) was conducted at the end of the second year of the implementation of **Vamos Ler!** program activities in schools with the following objectives:

1. Measure gains in overall reading and comprehension skills among grade two students in the target districts;
2. Identify relevant factors that support or hinder student learning in the target districts.
3. Based on the first two points, make specific recommendations for improvements in program interventions and approaches prior to the start of the final full year of program implementation in the schools.

These objectives informed the sample design and the data analysis approach described below. For measuring gains in student outcomes, comparisons were made with the 2017 **Vamos Ler!** Baseline Effectiveness Evaluation (EE)¹. To assess gains after one and half years of implementation in the schools, the focus of the assessment was on Grade Two students (G2). G2 Teachers as well as overall school

¹ USAID Vamos Ler! Effectiveness Evaluation Baseline Report, 2018

management and community involvement were also assessed to provide additional contextual and explanatory variables.

1.3 Effectiveness Evaluation Instruments

The primary indicators for the MEE are the student learning outcomes as measured by the Early Grade Reading Assessment (EGRA) instruments developed by the program in collaboration with MINEDH. Supplementary instruments were also developed with MINEDH counterparts to provide additional quantitative and qualitative data on aspects related to the **Vamos Ler!** intervention, which may serve to both contextualize the intervention and provide potential explanatory variables to better understand variation in the student outcomes. All tools are included in the report annexes.

1.3.1 Early Grade Reading Assessment (EGRA)

For the 2017 baseline evaluation, EGRA instruments for G1, G2, G3 were adapted following guidance found in USAID’s EGRA Toolkit 2.0 for each of the three Mozambican languages with which the program works and Portuguese (G3 only). In addition to the reading assessment itself, a student interview questionnaire was administered during this session with the students to gather information about reading behaviors, attitudes, and household situations. The baseline tools were revised during a workshop led by Dr. Sylvia Linan-Thompson in Maputo from the 17th to the 21st of June with participation of Vamos Ler Monitoring and Evaluation (M&E) staff and reading specialists, MINEDH bilingual education experts (from Nampula, Zambézia, and the central ministry), and University of Eduardo Mondlane bilingual education experts. The table below summarizes the revisions made to the EGRA instruments.

Table 1: EGRA Tool Revision

Revision	Justification
Elimination of subtasks for initial sound recognition, listening comprehension and L2 vocabulary based on images	The baseline EGRA instruments were deemed lengthy, potentially requiring up to 30 minutes to apply per student. Subtasks were eliminated based on high correlation with other subtasks or lack of predictive power with respect to reading fluency and comprehension.
Minor linguistic revision to L1 reading passages and comprehension questions	After two years of experience with both bilingual education and the official orthography, the teams were able to recognize minor orthographic errors and correct them. This did not substantially change the length or complexity of the text or questions.
Addition of new L2 Vocabulary task	The baseline L2 vocabulary task was based on identification of images and was deemed as not well aligned with the methodology used in the bilingual program. The instrument maintained an L2 vocabulary assessment by adapting the oral Portuguese subtask from the USAID Aprender a Ler (ApaL) EGRA. This subtask had already been validated by the previous program and was better aligned to the instructional strategy used by the program (e.g. “ <i>apontar e nomear</i> ” or point and name). The subtask required students to point to various parts of their body (head, foot, arm, etc.) or to demonstrate knowledge of concepts like “ <i>in front of</i> ” or “ <i>behind</i> ” by placing a pencil in the correct position relative to a blank sheet of paper.

Simplification of the contextual interview questions	The baseline contextual interview was deemed too long and thus was revised to focus on the core variables that workshop participants identified as most relevant to student learning outcomes.
--	--

After revision, the final EGRA instrument included the subtasks outlined in Table 2 below. The instruments were then adapted for integration into the Tangerine platform for testing and eventual implementation.

Table 2: EGRA Tool Subtasks for the MEE (all G2)

L1 Subtasks -Local Language	L2 Subtasks - Portuguese
Letter Names Letter Sounds Invented Words Oral Reading Fluency Reading Comprehension	Vocabulary
Contextual interview	

The revised instruments were tested in the field with approximately 20 students per language across a total of five schools. For the instrument testing, program staff and Ministry District Education Office (*Servico de Educaçao, Juventude e Tecnologia (SDEJT)*) bilingual education focal points asked teachers to identify their most fluent readers. This was done to avoid zero scores and guarantee that a large portion of the text would be read.

The testing allowed the teams to identify a few remaining linguistic corrections and, in the case of Elomwe, an improved alignment of the comprehension questions with the text. The testing also revealed that the language for some of the contextual interview questions needed additional revision.

1.3.2 Supplementary Instruments

Supplementary instruments were vastly simplified from the initial baseline. This was done to focus on the core variables that were deemed relevant by the program and MINEDH counterparts and was a deliberate decision to streamline the evaluation process for greater cost-effectiveness and increased chance for MINEDH ownership. In addition to supplementary data collected during the primary EGRA data collection visits, initial results were used to identify a subset of schools for follow-up visits to collect qualitative data, as outlined throughout the report below.

Classroom Observation Tool

A classroom observation instrument was adapted from existing tools developed by the program² to evaluate the extent to which teachers implement best practices for early grade reading instruction. The instrument is aligned to the reading and writing instruction practices introduced by the program

² The program used the guidance included USAID Toolkit for the Local Education Monitoring Approach (LEMA) to develop a district-based supervision approach that was implemented in all 21 program districts twice in 2019. As part of this approach, the program worked with MINEDH to develop classroom observation and school management assessment instruments in alignment with existing MINEDH tools and priorities.

through the district-level teacher training program. The instrument consists of an inventory of 33 practices divided into seven domains:

- Basic steps of instruction (5 items)
- Reading instruction (5 items)
- Writing instruction (5 items)
- Student-teacher interaction (6 items)
- Classroom environment (6 items)
- Classroom management (3 items)
- Teacher preparation (3 items)

Enumerators observe one 90-minute LI reading and writing lesson block in the same G2 classroom from which students were selected for the EGRA. Each item in the instrument is dichotomous. If the teacher used the indicated practice, the enumerator marked “YES”. By the end of the lesson, practices not observed are marked “NO”. This allows computation of an overall index variable as well as index variables by the domains noted above. All items are formulated as positive practices such that the greater number of practices observed is indicative of a higher quality of instruction practice observed overall and within each domain. The instrument includes a short teacher interview to provide additional contextual data.

Teacher LI Fluency and Comprehension Assessment

Based on prior observations that some teachers faced difficulties in reading fluently in the school’s LI, the program included an oral reading fluency (ORF) and comprehension assessment for teachers. In this case the G3 EGRA subtasks from the baseline were applied to the teachers following the same revision and testing process described above.

School Management Assessment

A school management assessment instrument was adapted from existing supervision instruments (principally the LEMA tools) development by the program with MINEDH counterparts and aligned to MINEDH quality indicators³. The instrument consists of 52 directly observable school management practices divided into 10 domains:

- Attendance and punctuality of school director and assistant director (3 items)
- Attendance and punctuality of teachers (4 items)
- Attendance and punctuality of students (8 items)
- School timetable (2 items)
- Participatory management (3 items)
- Financial management (5 items)
- School planning (4 items)
- Management and conservation of teaching and learning material (4 items)
- Use of school management tools (5 items)
- Coaching and pedagogical support (14 items)

³ Manual of Standards and Quality Indicators for Primary Schools and Adult Literacy and Education Centers (*Manual de Padrões e Indicadores de Qualidade para as Escolas Primárias e Centros de Alfabetização e Educação de Adultos*), MINEDH, 2018

Some of the 52 items were intermediary items used to calculate the item of interest (e.g. student attendance required the number of students enrolled and the number of students present) and these were then transformed so that all final items were dichotomous, allowing the construction of composite indicators by subdomain and overall. The total number of items used was then refined after conducting reliability analysis. This is described further in the analysis section. The instrument also includes a brief interview with the school director and assistant director.

Qualitative Tools for Follow-Up Visits

Immediately after the initial EGRA data collection visits were completed, the Vamos Ler! M&E team convened in Quelimane to host a 3-day data workshop & qualitative data collection training. The sessions prepared EGRA supervisors for follow-up visits to a subset of high-performing and low-performing schools. The purpose of the follow-up visits was to conduct semi-structured individual interviews (*school directors*) and Focus Group Discussions (FGDs) (*teachers and community members, separately*) to understand more about how various factors may have affected student performance as seen in cases with statistical outliers when looking at key outcome variables in the quantitative data.

EGRA Supervisors (i.e. SDEJT and DPEDH officials from each VL target district) were guided in their analysis of preliminary quantitative MEE data grouped by 9 key domains (see Table 3 below) and organized in tables. Based on their understanding of the results from these discussions, participants worked on Focus Group Discussion (FGD) and individual interview instruments (i.e. scripts) for the target groups.

Scripts for each group were aligned by the same domains, tailored for each target group, and conducted in the appropriate local language of the community (see tools attached in the report annexes).

Table 3: FGD/Interview Script Domains

Absenteeism	Language Factors	Instruction Quality
School Quality	Gender	School Management
Vamos Ler! Inputs	External Student Support	Community

Insights from the qualitative data collected with these tools have been included in Section 5 to supplement the quantitative findings in this report.

2 Design and Methodology

2.1 Vamos Ler! Effectiveness Evaluation Design

When the USAID Vamos Ler! Program began in 2016 with the initial goal of reaching over 3,000 schools across 21 program districts, we considered whether it would be possible to use a randomized control trial design. Given that nearly all the schools where the program would eventually be implemented were at the time fully monolingual, a true control group was not possible. Attempting to designate some schools as controls would require comparing LI reading scores for children taught to read in LI under the Vamos Ler! program with children never taught to read in their mother tongue. This is not a valid comparison since most children will be highly unlikely to read in a language in which they have never received instruction. This is clear from the 2017 baseline assessment conducted in

394 schools across the 2I program districts where over 90% of G2 students were unable to read a single word of a grade-level text in their LI.

The original evaluation design was based on conducting the EE at the end of every program year with a fixed sample of 120 schools in each of the three originally planned cohorts. Following this approach, the baseline was conducted in 360 schools (120 from each cohort), selected proportionally by district population. An additional sample of 34 existing bilingual schools from the 2I program districts were also included to see if these showed different characteristics from the monolingual schools (no significant differences were observed).

Prior to the 2019 MEE, some programmatic considerations influenced the evaluation design for the midline:

1. Elimination of the 3rd cohort of schools to allow the program financial capacity to respond to several requests from MINEDH (e.g. printing of books for bilingual schools outside of the **Vamos Ler!** intervention districts but using the same LI).
2. Elimination of the 1st midline evaluation, originally planned for 2018. Given the program and country context, it was clear that G1 students would likely not have had enough exposure to the intervention to see appreciable gains. This is also more in alignment with MINEDH's approach where the National Assessment is conducted with G3 students to assess G2 skill levels.
3. Given that 2020 would be the final full year of implementation in schools, the 2019 EE would be the last opportunity to identify relevant factors that could be incorporated into the program strategic approach to support improved student outcomes.

Based on these considerations, the MEE design was altered in the following ways:

1. Only G2 students are assessed for the midline.
2. As an implication of the first point, only cohort one schools are assessed (since cohort two was only introduced in 2019 and only had covered G1 so far).
3. Increased number of schools and students assessed to allow a more robust analysis of the variation in student outcomes. The additional school sample was added on top of the schools from the original baseline sample.
4. Qualitative follow-up visits were added to schools where statistical outliers had emerged to conduct interviews and Focus Group Discussions with school directors, teachers, and community members. These additional data have helped to identify unique characteristics of high-performing and low-performing school communities.

The design for the MEE considered the fact that we collect data on multiple, nested levels:

1. **Student:** reading scores (main outcome variable) and self-reported student data (e.g. whether they have books at home);
2. **Classroom/teacher:** classroom observation data and teacher data (e.g. LI reading proficiency) from same teacher of students from whom EGRA data was collected;
3. **School:** director, teacher, student absentee rates, whether the school received support from local NGOs supported by the program, and other school management indicators.
4. **District:** while no specific variables are collected at the district level, for analysis and reporting purposes, results are analyzed by district as this is the level of analysis that MINEDH will want to use to help focus subsequent supervision activities.

Sample Size Determination

Given that the Vamos Ler! program has interventions at both the level of the classroom (teacher training and coaching) and at the level of the school (school management training and coaching), we may expect a higher level of correlation of reading scores between students within the same school and classroom than between students from different schools/classrooms. This is measured by the Intraclass Correlation Coefficient (ICC), calculated through the following formula:

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}$$

where τ_{00} is the variance between schools, and σ^2 is the variance within schools. We can use existing data to estimate the ICC for our context. Given that the 2017 Vamos Ler baseline data was dominated by zero scores, we can look at 2016 USAID Aprender a Ler (ApaL) program G2 data (collected in 2016 in the same provinces in Mozambique as Vamos Ler), where we estimate an ICC of 0.23. In the literature this is typical (e.g. in the EGRA Toolkit, examples are given ranging from 0.17 to 0.48). An ICC of this magnitude is considered sufficiently large to justify a hierarchical design to which we might apply techniques of Hierarchical Linear Modelling (HLM). In the same manner we could consider to what extent students in the same district are more likely to have similar results than compared to students from a different district. This would then be a 3-level model (district-school-students). We could add a further level by considering multiple teachers within the same school, but this would further increase the sample size and we know that many schools may only have one G2 classroom, particularly in the more rural areas where most of our schools are located.

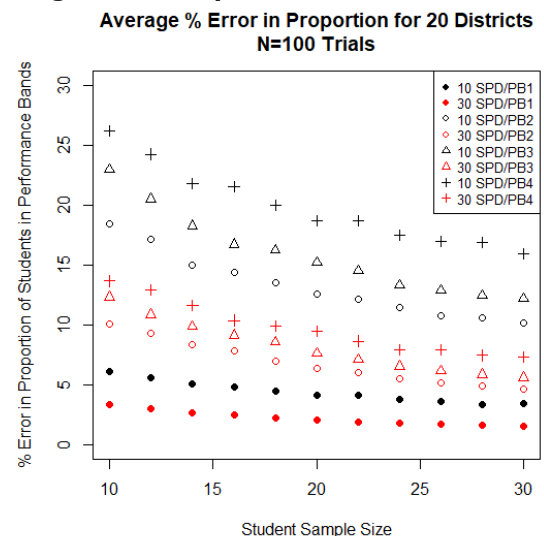
To determine a required minimum sample size, we could make several assumptions about effect size, population variance, and other factors, to use standard statistical formulas. Depending on the assumptions, we may get quite different results in terms of the suggested sample. We can also use simulations to study the effect of different sample designs on error in estimation of the mean or proportion of students within a certain fluency band. Using a model based on G2 ORF scores observed in ApaL, we produced simulated student results based on the actual number of school and students in all **Vamos Ler!** program districts. Using this approach, we could then test different sampling scenarios and compare calculated sample means to “actual” means of the full sample of simulated data. Given that the data is highly non-normal, the results obtain can be quite sensitive to the sample size. Figure 1 (below) shows the effects of different school and student sample sizes on the percentage error in the mean estimation. As can be seen, large gains are made in increasing the number of schools sampled from 10 to 30 at any level of student sample. The added benefit of adding more students per school levels off after 20-25 students.

The effect of larger sample size is even greater when ones looks at classifying students into different reading performance bands (PB). Consider the following categories as an example.

PB1: Non-Reader	PB2: Emergent Reader	PB3: Beginning Reader	PB4: Fluent Reader
0 – 5 CWPM	6 – 15 CWPM	16 – 30 CWPM	31+ CWPM

The graph below shows the percentage error in the calculation of percentage of students in each performance band. This is done for different school and student samples and for each of the four performance bands (e.g. PBI = % students with 0-5 CWPM). Two extremes of school samples are used - 10 (black marks) and 30 (red marks) schools per district (SPD). The horizontal axis is the number of students per school. Consider the case of classification of percentage of non-readers (PBI). These are the black and red filled circles on the graph. As can be seen, there is relatively little gain in increasing the school sample from 10 to 30 or in increasing the student sample. This tells us that if there are a high percentage of zero scores (assumed this will still be the case in the midline), you do not need a large sample to correctly classify the students in the lowest performance band. The large difference comes from the higher performance band. The black and red crosses in the graph show the effect of sample size on the highest performance band (31+ CWPM). Just going from 10 to 30 schools reduces the error by almost half. Going from 10 to 25 students also results in large reduction in the error, though the effect levels off above this. We feel this is a significant enough reduction to justify the resources required to utilize this sampling approach. Based on this analysis, we applied a sample of 25 schools per district and 25 students per school. The actual number of schools per district was determined in proportion to the total number of schools in the district. Given that we have 21 program districts, this resulted in an initial projection of 525 schools. After applying the random selection procedure, the final target population was 535 schools and 13,375 students.

Figure 1: Sample Size Considerations



2.2 MEE Implementation Plan

The main processes for the MEE followed the schedule below.

Date	Activity	Observation
17-21 June	EE instrument revision	Revision of EGRA and supplementary instruments.
21-24 August	EE supervisor training	Training of supervisors in central location (Mocuba).
26-30 August	EE enumerator training	Training of enumerators simultaneously in both provinces.
02-24 September	EE data collection	Data collection in all districts.
26-28 September	Initial results presentation with EE supervisors	Validation of results and preparation for qualitative follow-up visits.
30 September – 2 October	Qualitative follow-up visits	Qualitative data collected in 15 high-performing and 14 low-performing schools.
18 October	Preliminary EGRA results shared with USAID	Initial results on EGRA scores shared, not including supplementary tool analysis.

Data collection teams for the primary, quantitative data collection were composed of two enumerators and one supervisor per school. Supervisors collected supplementary data while enumerators collected EGRA data. Teams worked simultaneous in three to four districts per week, completing each district in one week. The logic for the team composition is outlined in the table below.

Team Composition

EE Data Collection	Number
Schools	525
Students per school	25
Number of data collection days planned	15
Enumerators per school (max 15 students per enumerator)	2
Districts	21
Days per district	5
Schools per district	25
Teams per district	5
Districts per week (Nampula – wk1, wk2, wk3) - 13 districts	5, 4, 4
Districts per week (Zambézia – wk1, wk2, wk3) – 8 districts	3, 3, 2
Teams/Supervisors	40
Total Enumerators	80

Additional roaming supervisors from senior program staff and senior MINEDH officials circulated between teams in each district to ensure overall quality and consistency of application of data collection procedures.

For the qualitative follow-up, a total of 29 schools were visited in 3 days, spread across 15 districts (7 Zambézia, 9 Nampula). The teams consisted of three members (supervisor, interviewer, note-taker). A total of 10 teams completed the work in three days. An additional day at the end of the process was included so that teams could meet to consolidate and submit their final notes.

2.3 Training for Data Collection

Building on previous experience, the MEE followed a rigorous enumerator recruitment process that included an open call for applicants with a rigorous screening process that included written and oral evaluations to guarantee adequate skill in the LI for the region in consideration. The open call resulted in over 4000 applicants for 80 enumerator positions.

Data collection supervisors were identified separately and included MINEDH district (SDEJT) and provincial (DPEDH) staff who had demonstrated skill and commitment throughout the previous years of implementation of the Local Education Monitoring Approach (LEMA) used by the program which collected similar data. Using MINEDH staff at district and provincial levels is one way the program seeks to gain buy-in for this type of evaluation.

Supervisor Training on Effectiveness Evaluation Tools

Given that supervisors already had substantial experience with the core aspects of the supplementary tools, a 3-day supervisor training was held with a total of 59 MINEDH staff from district, provincial, and central levels. The training included review of the supplementary tools, supervision of the EGRA

data collection, the data collection plan, use of tablets with the KoBo Collect application for data collection, and all data collection protocols. The training also covered supervisor responsibilities during the enumerator training.

Enumerator Training on EGRA Tools

Immediately following the supervisor training, enumerator workshops focused on collection of EGRA data using the Tangerine application were held simultaneously in the two provinces. In each province, 43 enumerators were trained together with 20 supervisors. The training included two days of classroom training, two days of field practice, and one day of preparation for the actual field data collection. At the end of this training, the enumerators were evaluated based on the observations made by supervisors in relation to attendance, punctuality, performance in the field practice, performance on Assessor Accuracy Measures (AAM, see Section 2.5.1 below), and a short multiple-choice quiz. Based on these results, 80 (24 women, 56 men) of the original 86 enumerators were selected (40 per each province).

Supervisor Training on Qualitative Follow-Up Tools

Immediately following the quantitative data collection activities, EGRA Supervisors met for 3 days in Quelimane for a data interpretation workshop and qualitative data collection training. The Vamos Ler! M&E team had been cleaning data in real-time as it had been entered (See 2.5.4 below) which allowed for a clean preliminary dataset that could be validated during this workshop and used as a basis for preparing for field interviews and FGDs in high- and low-performing outlier school communities.

Having just completed the initial data collection, EGRA Supervisors came with an in-depth knowledge of the tools and protocols used to collect the EGRA & Supplementary data, as well as knowing the schools in each district to be visited.⁴ For this reason, they served as enumerators (FGD facilitators, interviewers, and note-takers) for the qualitative data collection follow up visits. A breakdown of participants is included in Table 4 below.

Table 4: Data Workshop/Qualitative Training Participants

Institution	Nampula	Zambézia	Maputo	USA	Total
SDEJT	13	8	--	--	21
DPEDH	2	3	--	--	5
M&E – VL	2	4	3	1	10
Total	17	15	3	1	36

The first half of the 3-day workshop/training was reserved for interpreting the preliminary results from the MEE data collection (e.g. EGRA, SD interview, and Classroom Observation). Participants started in groups with looking at a specific district's data. They were given data grouped by *student* level variables, then *classroom* level variables, and finally *school* level variables. Participants noted observations on a worksheet and were asked to share thoughts with the group. In this way, participants were able to see trends in the data and make connections between factors related to students, teachers, and school management.

⁴ Note that not all EGRA Supervisors participated in this follow-up workshop/training. One (1) EGRA Supervisor from each of the 21 VL districts was invited to the follow-up visit training to ensure coverage from each district as outliers from the overall sample were still being identified during the workshop planning stage.

After looking at the school level data in this way, participants then looked at overall results by domain. For example, groups were given a sheet of data related only to ‘gender’ factors, or to ‘instruction quality’ factors from the overall sample across all districts. These domains were pre-defined as the topics to be discussed during the FGDs/interviews during the follow-up visits. With a deeper understanding of these factors, participants were more prepared to tease out the qualitative data in the semi-structured discussions they would be having during the fieldwork.

The second half of the 3-day training focused on preparing participants for the qualitative follow up data collection. This involved practical exercises focused on formulating follow-on questions in real time, note taking, and group simulations.

2.4 Summary of Field Work

The quantitative piece of the fieldwork ran from September 2-24 and covered multiple districts simultaneously in each province. A total of 40 teams covered on average one school per day, each with the goal of assessing the reading skills of 25 G2 students, observing one G2 LI reading and writing lesson, assessing the ORF and comprehension of G2 teachers, assessing school management practices, and conducting interviews with both teachers and school directors. The table below provides a summary of participation in the data collection process.

Table 5: Participation List for MEE Quantitative Field Data Collection

Sector	M	F	Total
SDEJT	33	5	38
DPEDH	4	1	5
INDE	0	2	2
DGGQ	1	1	2
DINEP	0	2	2
Enumerators	56	24	80

All program districts were informed in advance of the data collection period, but specific schools and dates were not shared. Given that the data collection coincided with the run-up to the 2019 presidential elections and campaigns in Mozambique, it was important to inform all local authorities as well so that data collection teams were not mistaken for part of a political campaign.

Even though districts were informed about the data collection period, successful completion of the field work faced many challenges due to extremely high rates of school director, teacher, and student absenteeism. In 42 schools (8% of the total sample), schools were completely non-operational. Due to lack of teachers or students in the school, no lessons were given on the day of the visit and no data was collected by the evaluation team. When a school was found to be non-operational, a return visit was planned for a later date, this time with SDEJT coordinating directly with the school director on the exact timing of the next visit. Even with this coordination, 19 of the 42 schools remained non-operational and were substituted with another randomly selected school in the district. The table below provides a summary by province. Note that one additional school was included in the sample due to inclusion of one additional substitution school that was visited without need. Since the data was collected, we maintained it in the sample.

Table 6: Quantitative Field Data Collection Summary

Province	Schools Planned	1 st Visit		2 nd Visit		Schools Replaced	Total Completed
		Schools Completed	Non-operating Schools	Schools Completed	Non-operating Schools		
Nampula	263	242	21 (8%)	9	12 (5%)	12	263
Zambézia	272	251	21 (8%)	11	11 (4%)	11	273
Total	535	493	42 (8%)	19	23 (4%)	23	536

In addition to non-operational schools, high-levels of absenteeism on the part of school directors, teachers, and students limited the ability to collect 100% of the data sample planned. We observed the following trends overall:

- In 31% of the schools, the school director was absent.
- In 11% of the schools, no G2 grade teacher was present.
- In 8% of the schools, fewer than 10 students were present.

Table 7 provides a summary of the field work in terms of percentage of planned data collected. As can be seen, about 80% of the student sample, nearly 90% of the teacher sample, and 100% of the school management sample was achieved. We note that Zambézia had somewhat higher rates of student and teacher absenteeism, resulting in slightly lower sample achievement.

Table 7: Planned vs Completed Quantitative Field Data Collection Summary

Province	EGRA		Classroom Observation		School Management	
	Planned	Completed	Planned	Completed	Planned	Completed
Nampula	6,575	5,587 (85%)	263	240 (91%)	263	263 (100%)
Zambézia	6,800	5,214 (77%)	272	238 (88%)	272	272 (100%)
Overall	13,375	10,801 (81%)	535	478 (89%)	535	535 (100%)

After completing the quantitative data collection process (e.g. EGRA, SD interview, and Classroom Observations) ending on September 24, the M&E team organized a three-day workshop from September 26-28. Participants analyzed preliminary data grouped by key domains and variables and prepared qualitative follow-up visits in 15 high-performing and 14-low performing schools. These visits to schools, identified as statistical outliers, took place from September 30-October 2 with the trained SDEJT/DPEDH supervisors from the EGRA data collection (see Section 2.3 above).

Teams were divided by language group (as community focus group discussions were conducted in local languages) and set out to visit 29 schools across 15 districts. Based on the MEE quantitative data, the preliminary results were sorted by average ORF scores, teacher absenteeism rates, and average letter

reading scores. Target schools for follow-up visits were then chosen to understand what is going on in the classrooms, schools, and communities where we find the outliers.

Data collection lasted 3 days (1 school/team/day given the distances between schools and districts) with 1 final day for bringing respective teams together in Nampula and Quelimane to verify materials and gather feedback from the data collection.

Table 8: Qualitative Field Data Collection Summary

Province	Districts	Schools	High / Low	# of Teams	# of Members/Team	# of Days
Zambézia	7	11	6 / 5	4	3	3
Nampula	9	18	9 / 9	6	3	3
TOTAL	15	29	15 / 14	10	30	3

2.5 Data Quality Control Approach

The EE used several approaches to guarantee the quality of the data collected. This began during the enumerator training and continued throughout the data collection process. These approaches are summarized below.

2.5.1 Assessor Accuracy Measurements

During the training of enumerators, assessor accuracy measurements (AAM) were conducted. For each EGRA subtask, the team created videos with the camera focused on the student stimulus sheet for the given subtask. A supervisor played the role of an enumerator, prompting a program staff member (in the role of a student) through the EGRA subtask. The program staff member would make occasional mistakes of the type expected in the field (e.g. incorrect letter read or skipping letters or lines). Enumerators, using tablets with the Tangerine application, marked the simulated student response. After passing through all EGRA subtasks, the enumerators uploaded the data, allowing program M&E staff to quickly analyze the data. The percentage agreement with the most frequent response for the group was determined for each enumerator and each subtask to calculate an overall percentage agreement per enumerator. The average agreement was over 90% overall, though notably, some enumerators received quite low scores. These results were used as part of the final criteria for selection of enumerators to continue with the field work.

2.5.2 Inter-Rater Reliability

Inter-rater reliability was assessed by having, in each school, one student assessed simultaneously by two enumerators. This was done always on the first student assessed and with enumerators swapping roles each day between enumerator and observer. The enumerator had the responsibility to guide the student through the EGRA stimuli and record their responses on the tablet. The observer simply recorded the responses on the tablet without interacting with the student. This allowed tracking of percentage agreement between enumerators throughout the entire data collection process.

For each student assessed, the number of items marked differently by the two enumerators was counted for each subtask. Looking at the ORF subtask, in most cases (94%), there was no difference between the two enumerators. When there was a difference, it was usually only on a small number of items. Given the relatively large number of items, the small differences seen in individual items resulted in marginal difference in overall score for the subtask. For the 6% of cases where the difference

between the two enumerators' final ORF score for the same student was non-zero, the average difference between enumerators was 1.4 words. While this difference is large compared to the mean ORF scores, it is marginal in terms of practical significance and would have marginal impact on the overall results.

2.5.3 Supervision of Data Collection

Supervisors from MINEDH, DPEDH, SDEJT, and Vamos Ler! accompanied the data collection to ensure that teams arrived at schools on-time, followed the data collection protocol, and communicated any challenges encountered. Supervisors used an EGRA supervision checklist to guide the process. At the end of each day all teams met to discuss challenges encountered to ensure common understanding of the solutions.

2.5.4 Data Cleaning and Real-time Quality Control

EGRA data collection with the Tangerine software and classroom observation and school management data collection using the KoBo Collect software allowed data cleaning and quality checks while the data collection was taking place. Data cleaning involved identification of incomplete cases, enumerator errors in the identification of basic school data (school name, cluster name, etc.). These were checked by comparing daily reports from field supervisors with summary reports generated by program M&E staff. Data cleaning was carried out by processing the raw data using analysis routines running on the R statistical software platform. In this way the raw EGRA data was never manually edited, it is simply processed by the program-developed software to produce a new cleaned output data file for analysis. This eliminates potential errors introduced by manual editing of the raw data. In this way, the final cleaned data sets were available within 24 hours of the final data collection.

2.5.5 Data Analysis

Immediately following data collection and cleaning, the EGRA data were available for analysis. With support from World Education, Inc (WEI) Vice President and Senior Monitoring and Evaluation Specialist Dr. Haiyan Hua, data were analyzed using SPSS and R software platforms. Where relevant, data were disaggregated by language, grade, and sex. In some cases, additional analysis using ANOVA, T-tests, and item analysis methods were used.

2.6 Limitations of the Study

The primary limitations of this study are the following:

1. **Overall low reading scores:** given that over 75% of the students could not read a single word of grade level text, the extent to which more advanced techniques such as Hierarchical Linear Modeling (HLM) was limited. Simple linear regression models were used to examine the effect of individual variables and this allowed us to estimate the overall relevance of key variables.
2. **High student absenteeism:** school directors and teachers were substantially involved in political campaign activities in the time leading up to and during the data collection. This resulted in high rates of absenteeism during the data collection and likely in the period prior to the data collection. This limits the study in two ways at least:
 - a. High rates of student absenteeism prior to the data collection implies overall lower exposure of children to the program interventions. This means that the student performance may not be fully indicative of the ability of the program approach to improve student reading scores, but instead points to systemic issues in delivery of education services in Mozambique.

- b. Students assessed may not be representative of the general student population. When teachers are not consistently present in the schools, students who live far away may be more frequently absent. The students living close to the school will often come when they see the teacher arriving on their motorcycle and word gradually spreads through the community. While this is the case, it is our opinion that the results would likely be lower if the more frequently absent students were present, as they are likely to have less overall exposure to instruction.

3 MEE Results

This section presents the results of the MEE in detail. We note that overall, significant gains in means scores were observed from baseline to midline in all student outcome measures, with zero score percentages substantially reduced for all student outcome measures. Gains were most visible in letter name and letter sound identification, with over 75% of students still unable to read a single word of grade-level text in their LI. In the following subsections we delve further into these results and identify some of the key factors that influence student performance.

3.1 Student Performance by EGRA subtask

3.1.1 Baseline to Midline Comparison

In 2017, a baseline assessment was conducted prior to the introduction of the program in schools. As most of the schools assessed in the baseline were monolingual, even though students were assessed in a language they understood (the LI of the school), they had never been taught to read in that language. Predictably, the results showed a high proportion of students with zero scores in all subtasks. Despite this, comparison to the baseline provides a good measure of progress achieved to date and can point to where improvements are still needed.

The tables below present the proportion of students by performance range (0, 1-5, 6-15, and 16+) for different subtasks, by language, for both the baseline and the midline. For all subtasks and languages, the percentage of zero scores was substantially reduced, in some cases by more than 50%. Mean scores increase significantly for all subtasks.

**Table 9: Letter Name Identification Subtask – Student Percentage Comparison
Baseline to Midline by Performance Band**

Letter Name Identification Subtask		Echuwabo		Elomwe		Emakhuwa	
		Baseline%	Midline%	Baseline%	Midline%	Baseline%	Midline%
Correct Letters Names Read Per Minute	0	72	32	80	37	84	48
	1-5	16	34	10	16	9	18
	6-15	11	24	9	27	6	22
	16+	1	10	1	21	0	13
	MEAN	1.6	6.2	1.4	8.9	2.1	6.1

Table 10: Letter Sounds Identification Subtask – Student Percentage Comparison Baseline to Midline by Performance Band

Letter Sound Identification Subtask		Echuwabo		Elomwe		Emakhuwa	
		Baseline%	Midline%	Baseline%	Midline%	Baseline%	Midline%
Correct Letters Sounds Read Per Minute	0	78	35	81	45	89	60
	1-5	11	30	10	15	5	15
	6-15	8	25	9	26	6	19
	16+	2	10	1	14	0	6
	MEAN	1.7	5.8	1.3	6.4	0.8	3.7

Table 11: Nonsense Word Reading Subtask – Student Percentage Comparison Baseline to Midline by Performance Band

Nonsense Word Reading Subtask		Echuwabo		Elomwe		Emakhuwa	
		Baseline%	Midline%	Baseline%	Midline%	Baseline%	Midline%
Correct Words Read Per Minute	0	96	85	97	81	95	85
	1-5	1	5	0	8	2	4
	6-15	2	6	1	7	2	6
	16+	1	3	1	4	1	6
	MEAN	0.4	1.5	0.4	1.7	0.4	1.9

Table 12: LI Oral Reading Fluency Subtask – Student Percentage Comparison Baseline to Midline by Performance Band

LI ORF Subtask		Echuwabo		Elomwe		Emakhuwa	
		Baseline%	Midline%	Baseline%	Midline%	Baseline%	Midline%
Correct Words Read Per Minute	0	96	84	98	77	93	82
	1-5	1	6	0	12	4	7
	6-15	2	6	1	7	3	7
	16+	1	4	0	4	0	4
	MEAN	0.4	1.7	0.2	1.8	0.6	1.8

Note that these performance bands are not referenced to any specific standard or benchmark. They were chosen to allow more insight into variability of the student performance data. Mozambique has yet to establish LI oral reading fluency benchmarks, but this work is planned for 2020.

Examining these data, we can make the following observations:

1. Across all subtasks and for all languages, zero scores were substantially reduced, with gains spread across the remaining performance bands to various degrees.
2. Largest gains were observed for letter name and sound identification subtasks (e.g. letter name zero scores for Elomwe went from 80% to 37%).
3. Gains for word reading tasks were relatively modest. The largest improvement was for Elomwe, yet still 77% of students could not read a single word of text. This result together with observations made during the lessons, indicate that teachers may need more support to improve their instruction on word decoding. It seems children can identify the letters and sound them out, but they have difficulty joining the sounds together to form the words.
4. Generally, we do not carry out comparisons between the languages, especially for reading fluency since each language has its own characteristics (e.g. average word length). However, we note that for Emakhuwa the gains and overall performance in the letter sound identification subtask are quite low compared to that for the letter names subtask. Follow-up should be made to determine if there are any difficulties noted in Emakhuwa or in the trainings carried out in Nampula regarding teaching the letter sounds.

3.1.2 EGRA Results by Student Sex

Throughout the analysis we consider student performance disaggregated by student sex. The table below presents the mean scores by sex and language for letter recognition and ORF subtasks. In each case we also include the percentage by which boys scored higher than girls on average.

Table 13: EGRA Results by Sex – Including Percentage by which Boys Scored Higher than Girls.

Subtask	Echuwabo			Elomwe			Emakhuwa		
	Boys	Girls	%Diff	Boys	Girls	%Diff	Boys	Girls	%Diff
Letter Names (letters per minute)	6.7	5.5	22%	10.0	8.0	25%	7.4	4.8	54%
ORF (words per minute)	1.9	1.4	36%	2.2	1.5	47%	2.6	1.0	160%

From these data we can see that:

1. Boys outperform girls on all subtasks (the same trend holds for other subtasks not shown here).
2. The gap is much larger for ORF compared to letter name identification, suggesting that the gap grows as the difficulty of the task increases.
3. The gap increases as we move from more southern districts (Echuwabo) to the north (Elomwe in upper Zambézia, followed by Emakhuwa in the north in Nampula), suggesting that cultural factors specific to the north (e.g. initiation rites and other traditions that are more common in the north) may help explain this gap.

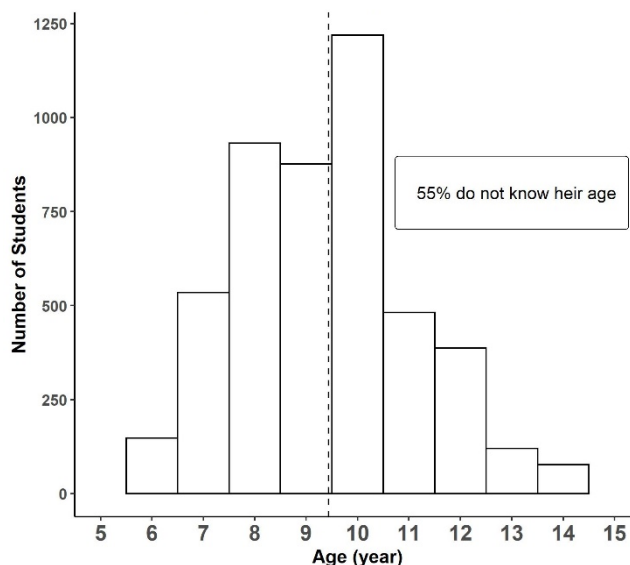
3.2 Student Characteristics and EGRA Results

As part of the EGRA, students are asked their age, their sex is noted, and several questions related to support for learning to read outside of school were asked.

As was noted in the baseline, the student age distribution is characterized by two notable features:

1. **A majority of students do not know their own age.** In this case, 55% of students were unable to answer this question. When examining student registration documentation, it was rare to find complete information on student birthdate and other required information. Regression analysis was done with student performance aggregated to the school level with explanatory variable of percentage of students that knew their age, and large positive correlations were found. We did not use this variable in our detailed modeling as it was difficult to articulate specific policy recommendations for the program or MINEDH. However, it is likely that this is a proxy for various socio-economic factors that have a large impact on student performance.
2. **Most students are over-age.** G2 students, assuming they entered the education system at six years of age, would be seven or eight years old at the time of the assessment. Amongst students who knew their age, the average was over nine years old. This could be indicative of both late entry and low efficiency of the system, with students repeating the same grades multiple times due to high absenteeism over the years, mobility of families in relation to seasonal labor opportunities, or other factors.

Figure 2: Student Age Distribution Grade 2



The table below is a summary of student responses to several contextual questions related to factors outside of school which may have positive impact on student learning outcomes. Many of these factors were practices that the program called on teachers and schools to implement (e.g. asking parents to read to their children at home). The data show that while most students had books and writing materials, only a small percentage of students reported having additional reading books at home or additional opportunities to practice reading outside of school (with someone at home or in a reading club).

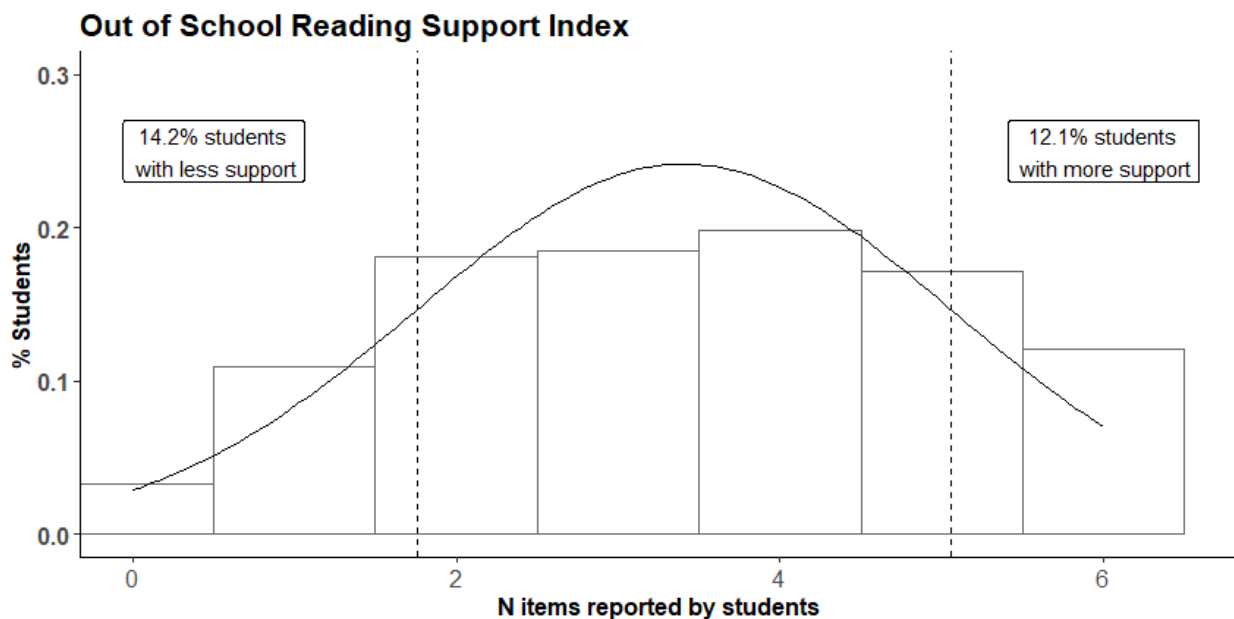
Table 14: Student Contextual Interview

N	Question	% YES
1	Which language do you speak most at home? (% indicates a match with the school's LI).	96%
2	Do you have school books?	93%
3	Do you have writing material (pencil or pen)?	87%
4	Do you have books at home to read (other than school books)?	34%
5	Do you ever read to someone at home?	50%

6	Does someone at home ever read to you?	58%
7	Do you participate in a reading club?	37%
8	Does your teacher give you homework every day?	89%
9	Does someone at home help you with your homework?	73%
10	Do you have to work when you are not in school?	91%
11	Do you live far from school?	44%

Item analysis was conducted with the 11 items to examine whether a reliable composite indicator for “Out of School Support for Reading” could be used. After running the inter-item correlation tests, we found that a total of six items (numbers four to nine in the table above) could be used to construct this indicator (the composite indicator had a Cronbach alpha of 0.7). The histogram of this indicator is shown in the figure below. We can see that it has a broad, normal distribution which we fit as indicated by the curve overlay. From this we can group students as having low, medium, and high levels of out of school support for reading. Using this approach, we find that **students with high levels of out of school support for reading, on average are reading 6.2 more letters and 2.7 more words compared to students with low levels of support**. After analysis of various other factors, this variable had the largest impact on student performance.

Figure 3: Out of School Reading Support Index



3.3 Teacher Factors and EGRA Results

3.3.1 Teacher Characteristics

In addition to the EGRA, field teams used a 33-item instructional practice checklist (e.g. classroom observation) to evaluate the teacher’s performance during a LI reading and writing lesson, assessed the teacher’s ORF and comprehension skills, and applied a short interview. The field teams worked with the same teacher of the students who were assessed, though in some cases (11% of the schools) it was not possible to collect this data due to teacher absences. In these cases, if the school was

operational and the school director or other authority from the school was present, the teams collected the EGRA data with the students present. The table below presents a summary of the teacher characteristics obtain as part of the interview section of the instrument.

Table 15: Teacher Characteristics

Teacher Characteristic	Male	Female	Overall
Teacher sex	77%	23%	100%
Teacher LI matches LI of school	84%	81%	83%
Teacher is ZIP-level trainer	27%	24%	26%
Teacher is district-level trainer	12%	13%	13%
Teacher is school director or deputy director	24%	10%	21%
Teacher teaches more than one class	58%	41%	54%
Teacher teaches more than one class during same shift	4%	4%	4%
Teacher has a house in the community	78%	60%	74%
Teacher has their own means of transportation	69%	28%	60%

The data permit the following observations:

1. Most teachers are male. It is well known that in Mozambique, schools in the rural areas (where most of **Vamos Ler!** intervention schools are located) are staffed by very few women. We also note that a larger percentage of female teachers do not reside in the community where the school is located and depend on someone else for transport to the school.
2. There are many challenges in terms of human resource management in the schools: due to insufficient teacher numbers many school directors or assistant directors are teaching in the early grades, many teachers teach multiple classes, and 4% of those teaching multiple classes do so during the same shift. This last point reflects a situation that is occasionally encountered: a school with only one or two staff who are required to teach multiple classes simultaneously. It is difficult to expect positive outcomes in these schools.

While female teachers are few in number, they have an important role in reducing the performance gap between boys and girls noted earlier. The data suggest that the reading achievement gap between boys and girls is reduced when students have a female teacher. The table below shows the means scores for boys and girls and the percentage by which boys outperformed girls as a function of LI and teacher sex.

Table 16: Student Performance by Student and Teacher Sex

Subtask		Echuwabo				Elomwe				Emakuwa			
		Male Teacher N=38		Female Teacher N=20		Male Teacher N=138		Female Teacher N=42		Male Teacher N=193		Female Teacher N=47	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Letter Names	Mean	6.2	5.0	7.7	6.8	10.7	8.2	9.4	9.0	7.7	4.7	6.8	5.5
	% DIFF	24%		13%		30%		4%		64%		24%	
ORF	Mean	1.7	1.1	2.5	2.1	2.5	1.6	1.9	1.9	2.7	1.0	2.3	1.3
	% DIFF	54%		19%		56%		0%		170%		77%	

From the table we can see that while the overall performance of both boys and girls remains low, **the performance gap between boys and girls is substantially reduced when the teacher is a woman.** Data on teacher hiring in Mozambique indicate that there are nearly equal numbers of male and female teachers currently employed. Given that so many of the schools targeted by the program are in more rural areas, this discrepancy suggests that there is an unequal allocation of female teachers to the more urban schools. This is confirmed anecdotally by many MINEDH partners. Further study should be done to understand why the pattern exists in order to provide clear policy recommendations to MINEDH.

3.3.2 Teacher Classroom Performance

On the assumption that the teacher’s performance during the observed lesson is indicative of their daily practice, and that the more items from the checklist that were observed during the lesson, the higher the quality of the lesson, we would expect to see higher performing teachers to have higher performing students. The 33-item checklist was grouped into seven domains and included the practices promoted by the program. The table below displays the average scores by domain and teacher sex.

Table 17: Teacher Performance on Observation Checklist

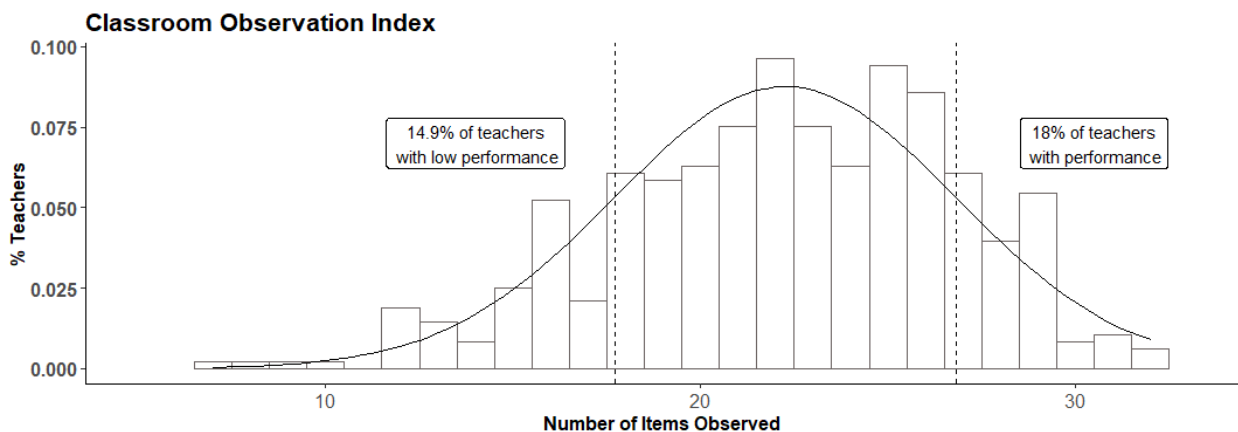
Instruction Domain	Male Teachers	Female Teachers	Overall
Basic steps of instruction	64%	74%	66%
Reading instruction	55%	65%	57%
Writing instruction	76%	77%	76%
Student-teacher interaction	77%	82%	78%
Learning environment	66%	67%	67%
Classroom management	51%	57%	52%
Teacher preparation	67%	72%	68%
Overall (33 items)	66%	72%	67%

Some observations can be made from these data:

1. Female teachers were observed to implement more of the instruction practices than male teachers on average. This may help explain the reduction in the student performance gender gap observed when the teacher is a woman.
2. The domain with the most positive performance was student-teacher interaction, and this was quite visible to the entire data collection team. Many MINEDH supervisors commented that is one of the very clear positive changes seen when comparing previous experience observing mono-lingual classrooms with bilingual ones – the students are clearly much more engaged.
3. The reading instruction practices was a domain with low overall performance. Items included whether students read aloud, whether they read together in pairs, whether they were asked questions about the text read, and if they discussed the meaning of words in the text that was read. Given that the student reading scores are so low, it could be that teachers were not able to get to very many of these practices. This is something future training activities should seek to address.

The 33 items in the classroom observation instrument can be used to construct an overall composite indicator for teacher classroom performance. Item analysis was conducted with all items and was found to have high internal consistency with a Cronbach alpha of 0.76. The histogram for all 478 teachers who were observed is shown in the figure below, along with a normal curve overlay. We again use this data to create a categorical variable of high, medium, and low performing teachers as indicated on the graph.

Figure 4: Classroom Observation Index



We note the following:

1. The distribution is well fit by a broad, normal distribution with a mean of 22 items observed (67%), indicating that in general terms a large proportion of teachers were able to implement many of the recommended practices.
2. We group the teachers into low, medium, and high classroom performance levels as indicated by the dashed vertical lines on the graph. We find that **students with high performing teachers, on average are reading 4.1 more letters and 1.4 more words compared to students with low performing teachers.**

3.3.3 Teacher LI and School LI Match

Teachers were also given an oral reading fluency assessment based on the G3 EGRA used in the 2017 program baseline. This task was included after several observations throughout the last two years of program implementation that some teachers had difficulty reading aloud with fluency. For most teachers in the program, the introduction of bilingual education was the first time they began to read and write in their own mother tongue or in any Mozambican national language. Teachers may also simply have low literacy levels even in Portuguese, making reading in a language they have never read before more difficult. In some cases, the teacher’s LI was not the same as the LI for the school. The program worked with SDEJT to identify teachers who were native speakers of the language of instruction of the school, but due to human resource limitations in the rural schools where the program is implemented, it was not always possible to allocate a teacher with the same LI as the school. In these cases, teachers had often already picked up a bit of the local language and then were given extra support by the school-based reading coach.

Given these considerations, we conducted additional analysis on the impact of teacher ORF on student ORF scores. The table below presents the percentage of teachers having the same LI as the school and breaks down the teacher and student mean ORF scores for native LI speakers and non-native LI speakers.

Table 18: Teacher LI Match and ORF Score Impact

Teacher LI match	Echuwabo		Elomwe		Emakhuwa	
	YES	NO	YES	NO	YES	NO
Teacher LI match	91%	9%	70%	30%	91%	9%
Mean ORF Teachers	52.4	48.8	50.1	44.1	39.1	35.6
Mean ORF Students	1.7	1.7	1.9	1.8	1.8	1.4

From these data we note a few points:

1. Schools with Elomwe as the LI had the highest proportion of non-native LI speakers.
2. The largest difference in ORF scores between native LI and non-native LI speakers were among teachers whose LI was not Elomwe but who were teaching in a school with Elomwe as the LI.
3. Comparing the ORF scores of students in these two cases, we see no significant differences. This was true even in the case of Elomwe schools where there were a higher proportion of non-native speakers. This suggests that having **a teacher that is not a native LI speaker is not a major factor limiting student performance.**

3.3.4 Teacher ORF and Comprehension

We also examined if the overall teacher ORF score was a relevant factor for predicting student performance. The figure and table below depict the spread in teacher ORF scores by language. Note that we do not read into the absolute values of the ORF as no benchmarking activity has yet to be carried out for these language in Mozambique. The relevant aspect to note is that there is quite a large variation in the scores. We also note that relatively few teachers were able to respond to at least 80% of the comprehension questions that were asked immediately after they read the text.

Table 19: Teacher ORF and Comprehension

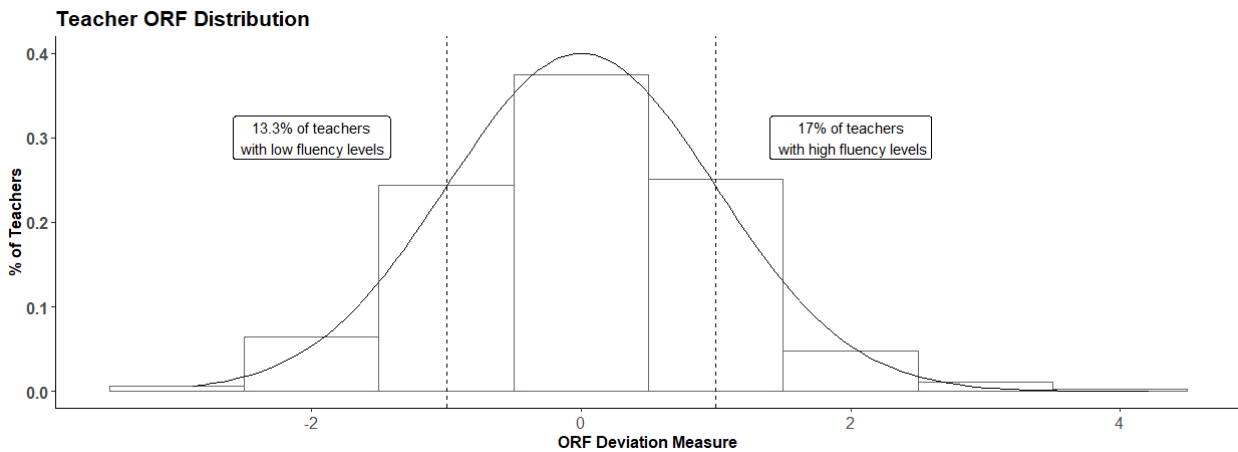
	Echuwabo	Elomwe	Emakhuwa
Mean ORF	52.1	48.1	38.8
Standard Deviation	16.0	12.6	10.8
Teachers achieving 80% comprehension	67%	41%	21%

To evaluate the impact of teacher ORF on student scores across the sample, we standardized the teacher ORF score by converting the ORF score in units of standard deviation for each language. We use the formula below.

$$ORFDIST_{ij} = \frac{MeanORF_i - ORF_j}{Sdev_i}$$

MeanORF_i is the overall mean ORF for teachers for a given LI = i, ORF_j is the ORF of the jth teacher, and Sdev_i is the standard deviation for a given LI. The distribution of this measure is shown in the figure below with a normal curve fit overlay.

Figure 5: Teacher ORF Distribution



Following the same approach as above, we can group teachers as having low, medium, and high levels of ORF. Using this approach, we find that **students with teachers who have high fluency levels, on average are reading 3.0 more letters and 1.2 more words compared to students with teachers who have low fluency levels.**

3.4 School Management Assessment and EGRA Results

3.4.1 School Director Profile

As part of the MEE, teams planned to meet with the school director (SD) and the deputy school director (DSD) to assess aspects related to school management. Smaller rural schools often do not have DSDs, in which case only the SD was interviewed. However, in a high percentage of schools, the SD was absent at the time of the visit, and in some cases both the SD and DSD (if the school had one) were both absent. The table below provides a summary of the situation along with other contextual data for SDs and DSDs.

Table 20: School Director Profile

School Managers	Schools with SD only	Schools with SD and DSD
Schools	33%	67%
SD absentee rate	24%	34%
DSD absentee rate	NA	32%
Both SD and DSD absent	NA	6%
% Female SD	10%	5%
% Female DSD	NA	15%
LI match SD	88%	73%
LI match DSD	NA	66%
SD has home in community	88%	77%
DSD has home in community	NA	77%
SD has own transport	82%	76%
DSD has own transport	NA	66%

Based on these data, we make the following observations:

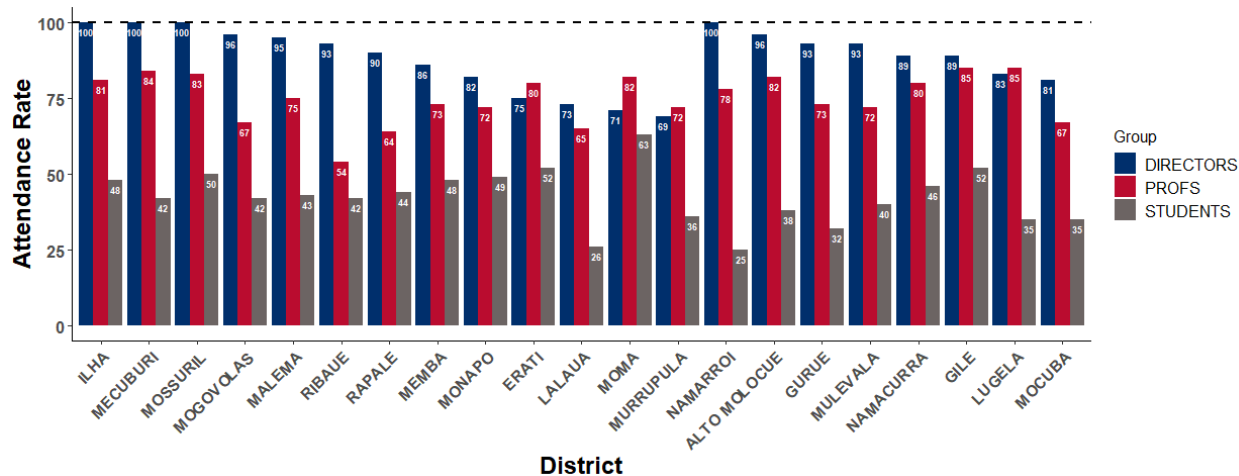
1. SD and DSD absentee rates are generally high at over 30% in schools that have DSDs and nearly 25% for schools with only DEs. We note that when the school has a DSD, the SD is more often found to be absent.
2. We generally find very few female SDs or DSDs. When the school has a DSD, there is a greater tendency for the SD to be male, and slightly higher percentage of female DSDs.
3. Schools only having a SD, which tend to be in more remote areas, show a higher percentage of SDs having a residence in the community and having their own means of transport as compared to both SDs and DSDs in schools having both positions. When SDs and DSDs do not live near the schools in which they work and when they lack transportation, they may be more likely to arrive late or be absent from the school.
4. We note that, compared to teachers, a relatively higher percentage of SDs and DSDs are not native speakers of the LI of the school. This could limit to some degree their ability to provide coaching support to the teachers in their school. In these cases, the program emphasizes additional peer coaching with fellow teachers also implementing bilingual instruction in the school.

3.4.2 Attendance and Punctuality

Many studies carried out in Mozambique have pointed to the high rates of absenteeism and lateness of school directors, teachers, and students and the negative impact that this has on student

performance⁵. The MEE showed that this trend continues. In each school, the teams recorded whether the school director was present during the visit, the number of teachers present, the number of students expected during the school shift observed, the number of students present in the classroom observed and where EGRA was applied, and the number of students enrolled in that classroom. These data were used to produce the graph below showing the average attendance rates for directors, teachers, and students on the day of the data collection visit.

Figure 6: Attendance – School Directors, Teachers, Students



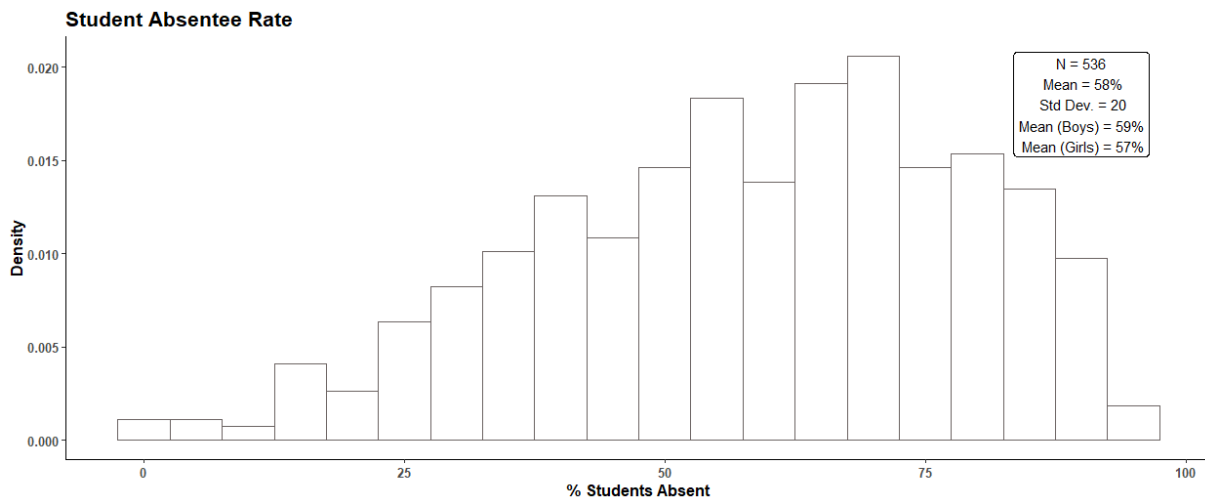
From this graph we make the following observations:

1. There is a clear domino effect – when school directors are absent, the teacher absentee rate is more likely to be higher; when more teachers are absent, the student absentee rate is more likely to be higher. **High student absentee rates are a result of regular absenteeism of the teachers.** This was a finding confirmed in the qualitative follow-up visits. Additional analysis shows a significant positive correlation between overall absenteeism of teachers in the school and the absentee rate in the observed classroom.
2. There is substantial variation from one district to another. For example, Lalaua with only 26% of students present, compared to Moma with a 63% average attendance rate.

While schools were not informed as to the exact day for the evaluation team to visit their schools, once the teams started working in a given districts, both school directors and teachers may make an extra effort to be at the school since they know their schools might be visited. For this reason, school director and teacher absentee rates may not be reflective of the typical situation. On the other hand, it would take a lot of extra effort to mobilize students to be present for the day of the evaluation visit. For this reason, we take student absentee rates as more reflective of the typical situation in the school, and more likely to be correlated with student outcomes.

⁵ Most recently from the World Bank: Bassi, M, et al., Education Service Delivery in Mozambique: A Second Round of the Service Delivery Indicators Survey, 2019.

Figure 7: Student Absentee Rates



The distribution of student absentee rates for all schools is shown above. We note the following:

1. The student absentee rate distribution is quite broad and skewed towards the higher end, with the overall average absentee rate 58%.
2. There was no significant difference in absentee rates between boys and girls. Given that boys and girls appear to have nearly equal attendance rates, we exclude this as a potential explanation of the differences in reading outcomes between boys and girls.
3. A separate analysis shows a significant correlation between student absenteeism rates and average performance for the school. **Schools with higher rates of student absenteeism, presumably reflective of regular teacher absenteeism, achieved lower average performance, with students reading 1.8 fewer letters and 0.9 fewer words on average.**

Absenteeism is a significant barrier to improving student performance. It is a major source of reduced instructional time for the students. Another factor reducing instructional time is late start of the school day. Evaluation teams arrived in the schools before the official start of the school day to observe the time at which instruction started. As a proxy for start time, supervisors recorded the time at which at least one teacher began instruction, independently of the grade level or class. Out of the 536 schools visited, the start time was observed in 515 schools. In the remaining 21 schools, the teams arrived late to the school due to difficulties of access by road. For the schools observed, **the average late start time was 39 minutes**. In Mozambican primary schools with one or two shifts, for grades one through three, a total of 28 periods of 45 minutes each per week (21 hours total)⁶ are planned. **A late start of 39 minutes each day of the week would result in a 15% reduction of instructional time per week**. The program held many discussions with MINEDH staff at all levels to explore how to address these systemic issues which have plagued the education sector for many years now. In the qualitative follow-up section below, we explore this further.

3.4.3 School Management Performance

While absenteeism and late start of the school day were a major focus of the school management assessment, the teams assessed several additional factors. This assessment considered multiple

⁶ Obligatory Orientations and Tasks for Primary Schools (*Orientações e Tarefas Escolares Obrigatórias*), 2015, MINEDH

elements related to effective school management practices that were emphasized in the trainings delivered by the program and which are part of official MINEDH policy for schools. We assume that these effective school management practices lead to increased instructional time, greater involvement of parents, increased pedagogical support for teachers, and effective use of school resources (including teaching and learning materials). With these elements all in place, we assume that this will lead to better student performance. The full instrument consisted of 52 items divided into 10 domains. The items were evaluated for internal consistency and a total of 44 items divided into seven domains were used. One of the original domains (existence of school timetable) was eliminated as nearly 100% of schools had a positive response for all items. An additional three domains (participatory management, financial management, school planning) were joined under one domain of participatory management as many of the items were conceptually related to effective planning and financial management with involvement of the school council in all aspects. The final 44-item composite indicator had high internal consistency with a Cronbach alpha of 0.87. The table below provides an overall summary of the results of the assessment.

Table 21: School Management Assessment Results

School Management Domain	Mean Performance
Attendance and punctuality of school director and assistant director (3 items)	75%
Attendance and punctuality of teachers (4 items)	51%
Attendance and punctuality of students (8 items)	46%
Participatory management (12 items)	57%
Management and conservation of teaching and learning material (4 items)	64%
Use of school management tools (5 items)	66%
Coaching and pedagogical support (14 items)	52%
Overall	57%

Based on these results, we note the following:

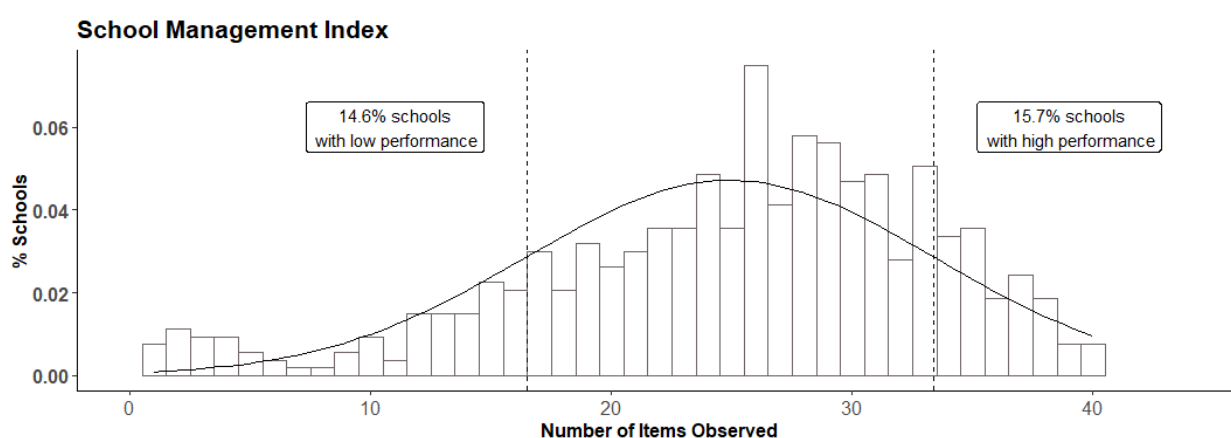
- I. The poorest performing domain was student attendance and punctuality (46%). In addition to the overall student attendance rate, this domain included items related to whether those students in attendance arrived before the official start time of the school, whether the students who were absent were actually “real” students⁷, and whether the school was implementing an “early warning system” promoted by the program to identify students with high absentee rates and conduct follow-up actions for support. Performance on this domain indicates that many schools were unable to follow program recommendations in this area, likely resulting in low levels of instruction time.

⁷ The student registration process in Mozambique is known to have many issues, one of which is inflation of student enrollment numbers. Teams requested the register of students and called each name. When a student was absent, the enumerator asked if anyone knew the absent student. In several cases, the student was unknown and likely never attended the school. See USAID Vamos Ler! March 3rd Action Research, 2018.

2. After attendance and punctuality of students and teachers, coaching and pedagogical support for teachers was the next lowest domain. Items under this domain included practices related to planning of lessons (including supervision of the planning process by school directors), provision of classroom coaching by school directors, and supervision visits by ZIP Coordinators, SDEJT, and program staff. The performance on this domain indicates that teachers may not have received adequate support to improve the quality of their instruction throughout the year. This may explain the variance in teacher performance noted in section 3.3.2.

As in previous sections, we look at how these factors affect student performance by looking at the school management index of 44 items and creating a three-level categorical variable of high, medium, and low performing schools to allow comparison of average reading scores among the cases.

Figure 8: School Management Index



From the distribution shown in the figure above, we note the following points:

1. The distribution has a small cluster of extremely low performing schools. These were mostly schools in which the school director and/or the assistant director were absent.
2. We find that **implementing more of the recommended school management practices had students who read on average 3.2 more letters and 1.1 more words compared to low-performing schools.**

4 Results by District

To support future supervision actions, we compile here the MEE key results by district. In the tables below, we present student-level variables in one table and teacher and school-level variables in another table. For letter identification and oral reading fluency, we include the overall mean score, the ratio of means comparing girls to boys (a score of 100% indicates complete equity, while a score closer to 0% implies boys substantially out-performing girls), and the percentage of zero scores. We also include student absentee rates and performance on the out-of-school support index as these two variables were found to have more influence on student learning outcomes. We note some districts had particularly low performance and should be prioritized for additional support in the following year of intervention. In Nampula Province priority districts are Erati, Lalaua, Malema, Mecuburi, Monapo, Murrupula, and Rapale. In Zambézia province, the priority districts are Gile, Lugela, Namacurra, and Namarroi. The strongest performing districts in each province were those with the fewest number of schools: Ilha de Moçambique in Nampula and Mulevala in Zambézia.

Table 22: Results by District – Student-Level

Province	District	Schools	Students	Letter Names			Oral Reading Fluency			Student Absentee Rate			Out of School Support Index
				Mean	Mean ratio (F/M)	% Zero Scores	Mean	Mean ratio (F/M)	% Zero Scores	M	F	Overall	
NAMPULA PROVINCE	ERATI	20	465	3.9	60%	54%	1.3	41%	78%	39%	53%	47%	60%
	ILHA DE MOC.	8	164	13.6	75%	21%	4.7	55%	65%	49%	44%	47%	59%
	LALAU	11	181	6.4	45%	45%	0.9	10%	92%	80%	64%	73%	63%
	MALEMA	22	452	4.5	52%	55%	1.3	50%	90%	55%	49%	52%	56%
	MECUBURI	19	337	3.1	72%	65%	0.4	13%	94%	60%	46%	55%	56%
	MEMBA	28	611	8.3	74%	35%	2.7	61%	77%	49%	51%	50%	64%
	MOGOVOLAS	24	517	9.1	74%	37%	2.2	39%	80%	50%	57%	55%	59%
	MOMA	28	694	6	55%	52%	2.3	25%	79%	37%	36%	37%	58%
	MONAPO	22	506	5.2	60%	48%	1.3	44%	88%	52%	49%	50%	65%
	MOSSURIL	20	431	6.4	60%	44%	2.2	36%	72%	50%	47%	49%	49%
	MURRUPULA	13	212	4.8	78%	58%	1.1	33%	84%	61%	58%	60%	62%
	RAPALE	21	454	3.8	63%	57%	0.8	23%	89%	56%	51%	54%	62%
RIBAU	27	563	6.4	65%	47%	2.1	56%	80%	57%	52%	55%	57%	
ZAMBÉZIA PROVINCE	ALTO MOLOCUE	45	946	9.4	75%	32%	1.6	55%	75%	62%	56%	59%	61%
	GILE	36	818	7.8	83%	41%	1.4	81%	83%	47%	45%	46%	57%
	GURUE	43	768	9.1	84%	32%	1.9	61%	81%	62%	64%	64%	57%
	LUGELA	41	628	6.3	94%	40%	2	86%	82%	61%	63%	63%	51%
	MOCUBA	43	856	9.5	82%	36%	2.6	86%	69%	60%	60%	61%	53%
	MULEVALA	14	279	12.5	77%	34%	3.2	72%	70%	59%	54%	57%	60%
	NAMACURRA	19	401	5.5	68%	18%	0.9	33%	91%	50%	50%	50%	38%
	NAMARROI	32	518	6.3	64%	49%	0.9	27%	83%	71%	73%	72%	49%

Table 23: Results by District – Teacher and School-Level

Province	District	Schools	Students	Classroom Observation Index	Teacher ORF	Teacher Absentee Rate	SD Absentee Rate	Avg. Minutes Late Start	School Management Index	Coaching Index
NAMPULA PROVINCE	ERATI	20	465	59%	42.6	19%	47%	27	51%	38%
	ILHA DE MOC.	8	164	72%	42.1	21%	24%	30	71%	72%
	LALAU	11	181	74%	43.9	32%	42%	21	64%	56%
	MALEMA	22	452	65%	35.4	24%	42%	20	65%	55%
	MECUBURI	19	337	62%	33.9	15%	7%	31	59%	58%
	MEMBA	28	611	67%	38.9	28%	14%	27	58%	51%
	MOGOVOLAS	24	517	67%	37.5	32%	68%	40	58%	50%
	MOMA	28	694	64%	39.0	18%	14%	31	51%	37%
	MONAPO	22	506	62%	41.3	28%	38%	24	58%	53%
	MOSSURIL	20	431	67%	38.1	17%	13%	37	62%	53%
	MURRUPULA	13	212	73%	42.0	27%	58%	12	69%	67%
	RAPALE	21	454	63%	34.5	35%	23%	49	59%	57%
RIBAUE	27	563	63%	41.6	45%	48%	33	56%	53%	
ZAMBÉZIA PROVINCE	ALTO MOLOCUE	45	946	67%	47.9	19%	33%	48	58%	53%
	GILE	36	818	69%	49.2	16%	23%	21	60%	54%
	GURUE	43	768	65%	48.9	25%	31%	36	58%	58%
	LUGELA	41	628	77%	51.1	15%	25%	46	52%	49%
	MOCUBA	43	856	73%	44.9	32%	49%	68	56%	49%
	MULEVALA	14	279	73%	50.2	25%	26%	28	69%	66%
	NAMACURRA	19	401	79%	58.3	21%	34%	28	56%	53%
	NAMARROI	32	518	66%	49.7	21%	25%	45	53%	58%

5 Qualitative Follow-Up

As outlined in sections 2.3 and 2.4 above, quantitative MEE preliminary results were used to identify a subsample of high- and low-performing schools for qualitative follow-up visits. Highlights from those discussions have been analyzed by target group and domain below.

5.1 Rationale

While not designed to be a statistically representative sample of beneficiaries and collaborators, the qualitative follow-up data was collected to:

1. Validate and provide clarity for some of the trends seen in the MEE results;
2. Identify additional factors not captured by the quantitative analysis that could help explain the high degree of variability between schools.

We note that qualitative data of this nature are not intended to answer questions of “how much” or “to what degree” a result or opinion exists in the whole population – this is the role of the quantitative data. Talking directly to outliers midway through the program, while they are receiving the VL package of support, provides opportunities for course corrections before activities may be scaled-up elsewhere while also providing clarity for some of the trends seen in the EGRA results – which have been outlined in Section 3 above.⁸

5.2 Participant Profiles

After preliminary data were analyzed, 29 schools were selected based on their performance with respect to student letter reading and ORF scores as well as the teacher absentee rates. A total of 15 high-performing and 14 low-performing schools were identified, and supervisors (e.g. SDEJT technical officers) notified school directors of the date of the visit to ensure they could mobilize community members and teachers to participate.

While the format for discussions with community members was originally envisioned to be an FGD (6-10 individuals), many of the discussions with community members turned out to be Community Interviews because of the large numbers of parents and other community members who showed up on the morning of the visit. Community Interviews are often used for gathering data from a relatively larger group of people (25-50 individuals).⁹ In this case, the nature of the semi-structured interview scripts (i.e. basic questions about community conditions and school services) lent themselves to this format and was allowed where numbers exceeded expectations. The breakdown of total and average numbers of respondents is detailed in Table 24 below.

Table 24: Qualitative Data Collection Participants

<i>Type</i>	<i>Group Interviews</i>		<i>Individual Interviews</i>
	Community	Teachers	School Directors
<i>Target Group</i>			
<i>Average Size</i>	20	3	1
<i>Total Participants</i>	570	82	29

⁸ USAID, Technical Note on Focus Group Interviews: Monitoring and Evaluation Series. USAID, 2013.

⁹ USAID, Technical Note on Focus Group Interviews: Monitoring and Evaluation Series. USAID, 2013. (page 2)

Community members engaged in discussions included - but were not limited to - School Council (SC) members and were held without the school director or teachers present. For the teacher discussions, the teachers who are trained and implementing the VL program were the primary focus, but any other teachers present who were not teaching at the time of the FGD were invited to participate. Schools directors were interviewed on an individual basis and all were present.

5.3 Summary Responses from Target Groups & Domains

The table below includes summarized responses condensed from all 29 interview sites. As questions from the domains were tailored to each target group in the domains, responses have been presented as such. This aims to provide a comprehensive understanding of the experience of teachers, school directors, and community members as we focus in on which factors help explain some of the quantitative results. Note that more (or less) information has been included in line with the structure of the interview scripts. For example, discussions with teachers focused more time on issues of absenteeism, while community members had more to say about external support and community involvement in the school.

Table 25: Summary Discussion Responses by Group & Domain

	Teachers	School Directors	Community Members
Absenteeism	<ul style="list-style-type: none"> • Most teachers say there are no teacher absenteeism problems, blame late students for missed classes/late starts. • Classes covered for absent teachers, but usually unprepared/don't teach full class. • Most teachers live close to or at the school, use their own transport. • Weather, distance to school, domestic chores, seasonal work, and lack of support from parents cause student absences. • For absent students, teachers visit homes/call parents to school, send Early Warning System cards, visit religious institutions. • When school day starts late, it's not possible to follow the lesson plans for the day, teachers often skip some of the lessons; some try to recuperate lost lessons on Fridays or weekends. • During rains students come late or are absent. Many schools have flooding/leaks because of weak infrastructure. • Schools often start late after holidays, with the first full week of classes often lost. Teachers spend first week going house to house, churches, etc. to let them know school is starting again. 	<ul style="list-style-type: none"> • In most cases where there is no teacher absenteeism issue (according to the SD) it is because teachers live close to/at the school. • No specific protocol exists for when teachers are absent, but usually teachers combine classes so students don't go home or disrupt other classes. • Reasons for student absenteeism are long distances to schools, harvesting activities, and parents not giving priority to school issues, activities, punctuality, etc. 	<ul style="list-style-type: none"> • Inconsistent teacher attendance and late arrivals of students have big impacts on quality of school. • Parents generally do not accompany their kids to school, so they are often late even when they leave the house on time (e.g. playing on the way).

	Teachers	School Directors	Community Members
Language	<ul style="list-style-type: none"> Across the board teachers note improvements in student-teacher interactions and student participation when compared to L2 classes, since starting bilingual education. 	<ul style="list-style-type: none"> Their teachers report noticeable improvements in student engagement compared to their previous monolingual experiences. 	<ul style="list-style-type: none"> Enthusiastic support from communities when talking about differences in student attitudes toward school since starting using LI in classrooms. Examples given included teaching others at home, reading in church, and being able to explain what they did in class.
Teaching Quality	<ul style="list-style-type: none"> Teachers noted the support received via district trainings and ZIP-level trainings, with fewer noting classroom support at the school level 	<ul style="list-style-type: none"> Support for improved school management mostly through district level trainings, ZIP-level trainings, and in some cases ZIP Coordinator visits 	<ul style="list-style-type: none"> N/A
School Quality	<ul style="list-style-type: none"> Teachers noted several factors that negatively impact school quality. These included the need to assign group work when classes are too big to help students with individual work; when class sizes are too large, students often must share materials; students often lose writing materials and parents do not replace them. 	<ul style="list-style-type: none"> Many utilize ADE funds for students in need, as well as mobilize parents to provide school materials for their students 	<ul style="list-style-type: none"> Students bring home exercise books and in some cases, they use complimentary books at home Parents confirm they know that students should be bringing pencils, pens, notebooks, and bags home, but that young students often lose their material.
Gender	<ul style="list-style-type: none"> Teachers note that they arrange students to work in mixed-gender groups, and call on them equally in the classroom to allow the same opportunities. Teachers note that girls are shy in the classroom and are disadvantaged because they have too much domestic work, leaving little time for homework. 	<ul style="list-style-type: none"> Widespread gender imbalance in school staff, with far more men than women. In some cases, SDs note that this affected girl's attendance and participation. SD gives guidance to teachers to be inclusive in classrooms; some give community lectures promoting girls' education. 	<ul style="list-style-type: none"> All community members agree that girls and boys have the same opportunities in school, regardless of how well they perform. In several cases, issues of premature marriages were cited as reasons for girls dropping out before completion, while boys drop out to find work in local markets (e.g. manual labor)
School Management	<ul style="list-style-type: none"> When there are meetings about absenteeism in the school, it is almost exclusively focused on <i>student</i> absenteeism. Nearly all teachers confirmed that teachers are marked absent, with some noting it only happens in extreme cases (e.g. multiple consecutive absences) Almost all teachers confirm that they know of other schools where salaries are discounted, but not in their own. 	<ul style="list-style-type: none"> SDs track teacher attendance and send to SDEJT for salary discounts, but SDEJT doesn't always follow through. Most report regular meetings with teachers, students, and communities to discuss tardiness/ absenteeism. 	<ul style="list-style-type: none"> Most schools have community meetings quarterly, but in most cases, meetings are initiated and organized by the school.

	Teachers	School Directors	Community Members
Vamos Ler!	<ul style="list-style-type: none"> • Improvements noted in school management since beginning of VL • Much improvement in planning, following a curriculum, and teaching strategies to use in the classroom • Teachers find the VL materials provided to be very helpful, but consistently note the importance of students being allowed to take books home to practice. 	<ul style="list-style-type: none"> • Schools most benefited from VL materials and trainings 	<ul style="list-style-type: none"> • N/A
External Support	<ul style="list-style-type: none"> • In the absence of reading clubs - which is most cases - students receive help at home, but difficult to verify because many parents are illiterate 	<ul style="list-style-type: none"> • N/A 	<ul style="list-style-type: none"> • Parents and siblings help students with homework where it's possible (i.e. if they know how to read) • Most communities do not have functioning reading clubs or other opportunities to read outside school
Community Involvement	<ul style="list-style-type: none"> • In most cases teachers meet with parents quarterly, and on top of that mostly when it is related to absences, sometimes communicated through School Council members 	<ul style="list-style-type: none"> • Frequencies differ (quarterly, monthly, bi-weekly), but each SD reports holding meetings with School Council members 	<ul style="list-style-type: none"> • Members generally see their role as supporting infrastructure and helping to reduce student absenteeism • Members build school structures, such as teacher housing, and help to mobilize community for matriculation and attendance throughout the year. • Most communication is through the SD, but in some cases community members interact with teachers who live nearby.

5.4 Responses Based on School Performance

While Section 5.3 aims to more generally convey the experiences of school communities, in this section we identify common characteristics of higher and lower performing schools, respectively, which emerge and align across the different respondents.

It is important to remember that these points have been mentioned in communities considered to be outliers from the overall sample of target VL schools (when looking at key indicators of performance such as ORF, teacher absenteeism, and letter reading). We cannot confirm that they hold up as broad generalizations across all VL schools. They are provided to highlight best practices and potential areas for follow up and support in schools with exceptional performance data. These common characteristics are not representative of trends across *all* high or low performing schools. Rather, responses appearing in multiple distinct conversations have been considered here.

5.4.1 Emerging Trends Among High-Performing Schools

Community Ownership: When the community feels responsible for the management of the school, they expect certain standards of performance from teachers and students. This has the added benefit of improving teacher motivation, as their job is facilitated by a supportive working environment with allies in their communities to help combat absenteeism, tardiness, etc. In one of the schools visited, teachers described meeting with community members every two weeks (verified in both SD interview and the community discussion) and noted the “good working environment” in the school. This enables the teachers to give regular feedback to parents on what they should be looking for in student notebooks at home, address absenteeism problems, and inquire about students who have dropped

out to quickly encourage them to come back to school. This type of environment lays the groundwork for improved teacher performance, which (see Section 3.3.2) the classroom observation data showed us is improving rates of word and letter reading directly in the classroom.

Teacher Attendance: As seen in Table 25 (Summary Discussion Responses by Group & Domain) above, teachers from many of the schools visited for follow-up interviews denied any issues of teacher absenteeism at their own schools. Yet in many of the low-performing schools, teacher responses did not align well with SD responses. Regardless of the nature of the response (i.e. confirming or denying an issue of teacher absenteeism), this misalignment in responses was not the case with the higher-performing schools. We found that teacher and SD responses were aligned in the same way (while interviews were held separately), and in many cases both mentioned the reasons for low absenteeism rates was that teachers lived near/at the school. When asked about absent teachers, one SD noted that, “teacher absenteeism is not a problem in our school, because there is participatory management that motivates them. Teachers and directors enjoy all the same rights and duties in this school.”

Open Lines of Communication: One of the more notable differences in the tone of the conversations in high-performing schools during the visits - evident in many of the recorded responses and confirmed through debriefs with the interview teams - was the positive and open relationships the teachers, SDs, and community members had with each other. In one school a parent commented that “we contact teachers and the SD every week to know if our students are coming to class.” This was quite different to many schools where teachers blame parents for not staying informed on school activities and schedules. In another case, a parent claimed “we interact daily with the SD and the teachers through School Council members. Some SC members come to school every day to control student and teacher absenteeism.”

5.4.2 Emerging Trends Among Low-Performing Schools

Inadequate Infrastructure Weak structures and insufficient classroom space are frequently mentioned as problems during times of extreme weather and affects the ability of the school to function consistently. One teacher explained that “there are some classrooms which offer learning conditions, and others do not.” In some places the school does not even open during rains. “Any time it rains, the school is closed,” as another school confirmed.

Issues with school structures are compounded by late and/or absent even in times of normal weather. Even schools where enough classrooms exist, issues of teacher absenteeism strain existing resources and have an adverse effect on the classroom environment. As one teacher confirmed, “when classes are too big, we have to teach under the cashew trees outside.” This is the case when teachers are missing at the start of the day and classes need to be combined in classrooms lacking the capacity for such high numbers of students. In these same schools, teachers report that when classes are too big they revert to giving individual work because there are too many students to manage. In these cases, the pedagogical approaches teachers are trained to use – such as ‘I do, we do, you do’, or reading practice in pairs with appropriate numbers of books – are not possible, particularly when forced to use a makeshift space outside the classroom.

Conflicting Feedback: Interviews for each group were organized using a structured set of domains but with semi-structured questions to allow for more in-depth and open discussions on the chosen topics. When comparing these responses in some low-performing schools the interviewees contradicted one another, highlighting strained relationships and misunderstandings on issues related to roles and responsibilities in the school. For example, when one SD was asked about absenteeism he claimed that “teacher absenteeism is a problem” and that “student absenteeism is related to the

absence of the teachers, but also due to negligence on the part of the parents who do not send their children to school.” In the same school, teachers in their FGD independently claimed that “teacher absenteeism in this school is inspired by the irresponsible and absent school director.” Furthermore, when community members were asked to explain why students may be late or absent, they stated that it is “because of the absent SD and teachers. They sometimes teach only 2 days/week and don’t live near the school.” Expanding on this frustration, this tied into the community’s observations around the benefits of the bilingual education program. “Students and parents are motivated by bilingual education, but it is difficult to perceive the importance of the program, because they rarely have classes in the school.”

Communication Barriers: In addition to inconsistencies across interviewees in low-performing school communities, interviews also highlight cases of limited direct communication among the different members of the school community. For example, one group of teachers described ways they disseminate information about the start of the school year. “Teachers have to go house to house to recruit students to come back after holidays, which takes a full week.” Another teacher, specifically referring to absent students, noted that “we speak with students who are present to talk to their absent classmates, but we also sometimes visit their homes.” Noticeably there were no specific references to communication with the SD or the School Council about these matters. A system where teachers going house to house to get absent students depends largely on the commitment of an individual teacher and likely only benefits those students living close to the school. Reinforcing a strong network of communication between teachers/SD/SCs has the benefit of keeping parents informed while helping teachers to be more effective.

5.5 Individual Cases

Here we showcase a select few schools to understand experiences of the different members in a particular school community and how they relate to each other. Quantitative results from the school have been included as well to attempt to understand the effects of the school environment on student performance.

5.5.1 Example of a Low-Performing School

In one of the low-performing schools, community members were vocal about problems with teacher absenteeism and asked the team to help them by getting SDEJT to replace all the teachers and the school director. Clearly there was distrust between the community and the teachers, which was evident when comparing responses on the topic of absenteeism (i.e. teachers blamed students, community blamed teachers). The SD also noted in this community that not everyone has bought into the bilingual education approach. During the discussion, situated across the schoolyard from all the classrooms, one community member commented:

“You can see, even when the teacher is there, they don't make the kids ask permission to come into the classroom [pointing at a student walking in late to one of the classrooms] if they are late, so the kids don't even know when they should be on time and know that it doesn't matter if they are late. These teachers have been here too long. All the kids are their friends and they don't enforce rules. The director also needs to be transferred. He doesn't have the energy or motivation to run the school well.”

Community members at this school seemed to have little interaction with teachers (quarterly meetings only) and speak mostly with the SD:

“People interact mostly with the director. We only meet with teachers in the beginning of the year, quarterly, and then once at the end of the year to hear results.”

As seen in the table below, the school also registered high levels of student absenteeism, which was partly explained by the School Director who blamed “negligence” on the part of the parents who don’t send their kids to school. He also commented that teacher gender contributes to this problem, which reinforces disparities in learning outcomes by teacher gender:

“There is an unequal gender balance among teachers in the school. One female teacher and four male teachers. This has influenced how frequently girls come to school. Girls come more when they have a female teacher.”

While there is presumably a mix of factors leading to high absenteeism in the schools, the impacts of such low amounts of classroom time for most students is reflected in the majority zero scores on both letter and word reading found during the EGRA (see table below).

	Correct Letters per Minute (boys)	Correct Letters per Minute (girls)	Correct Letters per Minute (overall)	Oral Reading Fluency (overall)	Students have books at home to read (%)	Someone reads to students at home (%)	Student/Teacher Interaction Index	Student Absenteeism (%)	Teacher Absenteeism (%)
School	0	0	0	0	50%	57%	67%	85%	40%
Overall	6.7	5.5	6.1	1.8	34%	58%	78%	58%	25%

5.5.2 Example of a High-Performing School

In one high-performing school, the community members emphasized that they ownership of the school and importance of taking care of the school and its staff. One member commented:

"It's our school, the teachers and director only came to work here, so we need to take care of the school so one day when they leave, they leave it in good condition for the next people who come to teach our kids. And if the kids will become someone someday maybe they'll come fix the road here for us since they'll be from here."

Another community member explained the community’s approach when they have an issue with a teacher, as others nodded unanimously in agreement:

“When we see something strange in the behavior of a teacher, we invite them to come talk with us. But overall these teachers are great.”

Teachers, in their FGD separate from the community members and the SD, also talked of accountability for not missing classes because of the community members.

“What motivates us to be on time and present in the school is the positive environment in the school. Working in the school itself is nice, but also it’s a supportive community.”

Teachers, community members, and the SD all separately acknowledged the supportive relationship between the school and the community and talked about the meetings they have with the School Council every two weeks. High overall performance among G2 students on the EGRA and low levels of teacher absenteeism have resulted in this community, as seen in the results below.

	Correct Letters per Minute (boys)	Correct Letters per Minute (girls)	Correct Letters per Minute (overall)	Oral Reading Fluency (overall)	Students have books at home to read (%)	Someone reads to students at home (%)	Student/Teacher Interaction Index	Student Absenteeism (%)	Teacher Absenteeism (%)
School	23.1	17.4	20.6	7.5	84%	84%	100%	62%	0%
Overall	6.7	5.5	6.1	1.8	34%	58%	78%	58%	25%

6 Conclusions and Recommendations

6.1 Key Findings

As a summary, we include here the key findings noted in the quantitative analysis and further confirmed through the qualitative follow-up visits.

1. **Improved student reading skills compared to baseline:** Compared to the program baseline, student performance improved substantially for all EGRA subtasks. This was true for all three Mozambican languages assessed. The improvements were in large part due to a substantial reduction in the percentage of students registering zero scores on the EGRA subtasks. The largest improvements were seen in letter name and letter sound identification subtasks, with more modest gains in word and text reading.
2. **Positive reception of bilingual education in classrooms and communities:** A clear finding from the classroom observations and further validated in the qualitative follow-up visits was the very visible improvements in teacher-student interactions and overall student participation brought about through the introduction of bilingual education. This has the potential to produce large gains in student learning outcomes, but we find several factors related to school management and relationships between the school and the community that limited the results overall.
3. **Reduction of instruction time through school closures, absenteeism and tardiness:** The evaluation noted continued high levels of school director, teacher, and student absenteeism and lateness resulting in substantial reductions of instructional time. A total of 41 schools (8% of the sample) visited were completely non-functional (no lessons given) on the day of the data collection visit. Both quantitative and qualitative results point to the high rates of student absenteeism (58% average student absentee rate overall) being the result of irregular teacher attendance. In addition to absenteeism, the results suggest that a late start to the school day could result in as much as a 15% reduction in instruction time over the course of a year. Factoring in the high absentee rates and other factors such as delayed start of lessons after school holidays, the overall amount of exposure to reading instruction that students have is quite limited.
4. **Gender reading performance gap:** Boys outperformed girls on all EGRA subtasks, with the gaps in performance growing as the difficulty of the task increases (i.e. moving from letters to text reading). The gender performance gap also showed a regional variation, with districts in the south having a smaller gap compared to districts in the north, suggesting that cultural factors specific to the north (e.g. initiation rites and other traditions that are more common in the north) may help explain this gap. It was also observed that the gender performance gap was reduced when students had a female teacher. While this was encouraging, the fact that rural schools have very few female

teachers limited the overall impact that the female teachers had in support greater equity in reading performance. We also noted that female teachers generally had better performance as measured through the classroom observation instrument.

5. **Identification of factors that influence student learning outcomes:** A high degree of variability was noted both within schools (students in the same school with same teacher, but drastically different outcomes) and between schools (average performance of one school drastically different from another school). Several factors were identified through multiple regression analysis as being significantly correlated with the student learning outcomes. The results (standardized coefficients of the factors) clearly point out that not only are these factors associated with the learning improvement, but also, they rank by the order of influence on student ORF scores as listed below.
 - a. **Out-of-school support**, including access to books at home, having someone at home to read with, participation in a reading club, and having regular support at home to complete homework assignments. Students with high levels of out of school support for reading, on average are reading **6.2 more letters and 2.7 more words** compared to students with low levels of support. Out-of-school support, being the most influential factor, inform us that early children's learning in literacy must have conducive learning environment at home and parental engagement in and support to their children. The fact that out-of-school support seems to be more critical in young children's literacy learning than school and classroom factors deserves strategic policy attention and program planning.
 - b. **Teacher classroom instruction practice** was observed to be largely positive, though with a large degree of variability. Female teachers generally performed better than their male counterparts, but all teachers had difficulties implementing some of the practices specific to reading instruction (students reading aloud, students reading together in pairs, asking students questions about the text they read, and discussing the meaning of words in the text that was read). We found that students with high performing teachers, on average read **4.1 more letters and 1.4 more words** compared to students with low performing teachers.
 - c. **Teacher oral reading fluency levels** showed a large degree of variability for all languages. The introduction of the program may have been the first time that many teachers began to read the Mozambican languages covered by the program, and while in some cases the teacher's LI did not match the LI of the school, this did not have a significant impact on student performance. A teacher's overall fluency level did have a significant impact, with students of teachers with high fluency levels, on average reading **3.0 more letters and 1.2 more words** compared to students of teachers with low fluency levels.
 - d. **Implementation of effective school management practices** such as management of student attendance and provision of in-school pedagogical support proved the most difficult for schools. More regular provision of in-school coaching could help eliminate some of the variance in teacher performance as observed during the classroom observation and ORF assessments of teachers. We found that schools implementing more of the recommended school management practices had students who read on average **3.2 more letters and 1.1 more words** compared to low-performing schools.

With the sample design, we were able to conduct an analysis of the variance components in student learning outcomes, focusing specifically on ORF. We found out that 73% of the total variance in ORF is within individual students. In other words, the individual level variance is likely associated with unique individual characteristics – factors outside of the school environment and more related to parental support. The remaining variance lies in the school/classroom level (24%) and 4% at the district level, suggesting that teacher improvement program activities and school management enhancement may only impact 28% of the variance in student learning outcomes¹⁰. We note that statistical interaction terms may happen as teacher and school factors become positively influential to home environment factors to young children’s learning of literacy. Although the ratios of the three levels may vary from context to context, worldwide literature on all three levels of variance component analysis has always been that the variance of the individual children in learning takes more than half of all the variance in young children’s learning outcomes, particularly for student reading outcomes. This again indicates that program like **Vamos Ler!** should continue to implement strategies that support both families and communities as well as teachers and schools.

While the overall text reading results remain low, we note several bright spots. In one of the top performing schools, with 25 students assessed, students on average identified 33 letters per minute and read 16 words per minute, with no students having a zero score on the oral reading fluency subtask. This was in a very rural school with relatively poor infrastructure, quite far from the main town of the district. In this school, we noted a high level of support by parents and community members both as reported by the students during the EGRA and confirmed during the qualitative follow-up visits. This is an encouraging finding because it shows that much better results are possible, even in the remote, rural schools where the program is implemented and even after only a year and a half of intervention.

6.2 Recommendations

Based on the quantitative findings, and additional data from the qualitative follow-up visits, several recommendations for the program were developed and considered for integration into the 2020 program implementation strategy. These are presented below, broken down by target population.

Table 26: Programmatic Recommendations for Vamos Ler!

Level	Recommendation	MEE Evidence Base	Strategic Approach
Classroom	Teachers should integrate instruction practices that reduce differences in results for boys and girls	Substantial differences between boys and girls, with boy outperforming girls in all EGRA subtasks.	Improved approaches integrated into all program teacher training and coaching activities.
	Provide support to teachers to improve their oral reading fluency skills, with special focus on teachers who are not native speakers of the school’s LI	Teacher oral reading fluency assessments showed a large degree of variability in teacher fluency levels. Teachers with lower fluency also had lower performing students.	The program training approach should build in time for teachers to practice reading aloud. In-school coaching support should also focus on strengthening teachers’ skills in this area.
	Teachers should place greater focus on building	Larger gains in letter reading compared to oral reading	As the phonemic approach is new for most teachers in the

¹⁰ Bryk, Anthony and Raudenbush, Stephen. Hierarchical Linear Modeling: Application and Data Analysis Methods (2nd Edition). Sage Publications, 2002.

	the decoding skills of students.	fluency, suggest that students have difficulty linking the letter sounds together to read words and phrases.	program, greater emphasis and practice should be devoted to this area during program training and coaching activities.
	The program should find ways to address teacher motivation.	High rates of teacher absenteeism, which qualitative and quantitative findings indicate are provoking higher student absenteeism rates, are reducing the amount of time in the classroom and diminishing the quality of instruction where teachers ARE present (due to disruptions, student overflow, etc).	Teacher certificates and other forms of recognition for strong performance
	Teachers need to receive more in-school coaching	Classroom observations showed that some teachers were unable to effectively implement the instruction practices recommended by the program. Students of the lower performing teachers also had lower reading scores.	Training and supervision activities should reinforce the provision of in-school coaching.
School	Greater involvement of local authorities (administrators, district director, local leaders) as well as the provisional inspectorate to address absenteeism	Both quantitative and qualitative results point to the negative impact absenteeism has on student learning. The qualitative findings suggest that a lack of consequences for recurrent absenteeism is a large part of the problem. Local government has the authority to take disciplinary action to address these issues. The involvement of the inspectorate could help add weight to any recommendations that result.	Following all supervision activities, the program should seek to report out to the local authorities with details specific to the schools that were visited.
Community	Expand reading clubs Strengthen use of complementary reading material at home Promote reading competitions and other events to stimulate community involvement	Out-of school support was the largest factor influencing student outcomes.	These activities align with efforts already undertaken by the program. Efforts should be made to strengthen the approach and scale up to all schools covered by the program.

Bibliography

Bassi, M, et al., Education Service Delivery in Mozambique: A Second Round of the Service Delivery Indicators Survey, World Bank, 2019.

Bryk, Anthony and Raudenbush, Stephen. Hierarchical Linear Modeling: Application and Data Analysis Methods (2nd Edition). Sage Publications, 2002.

Dhliwayo, Trymore Mafucha, Haiyan Hua, Yuri Machkasov, David Noyes, Adam Turney, and Willem van de Waal. Vamos Ler! Effectiveness Evaluation Baseline Report. USAID, 2018.

MINEDH, Manual of Standards and Quality Indicators for Primary Schools and Adult Literacy and Education Centers (Manual de Padrões e Indicadores de Qualidade para as Escolas Primárias e Centros de Alfabetização e Educação de Adultos), MINEDH, 2018.

RTI, Toolkit for the Local Education Monitoring Approach (LEMA), USAID, 2016.

USAID, Technical Note on Focus Group Interviews: Monitoring and Evaluation Series. USAID, 2013.