USAID
FROM THE AMERICAN PEOPLE

Impact Evaluation of the USAID/ Aprender A Ler Project in Mozambique

Year 3 IE/RCT

Final Report

May 23, 2016

This report was prepared for USAID/Mozambique by Magda Raupp, Bruce Newman, Luis Revés, Carlos Lauchande and Edward Jay Allan under Contract AID- 656-C-15-00002 awarded to International Business & Technical Consultants, Inc. (IBTCI). The authors' views expressed in this report do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

IMPACT EVALUATION OF THE USAID/*APRENDER A LER* PROJECT IN MOZAMBIQUE

IE/RCT Report
Final Report

International Business & Technical Consultants Inc.

Magda Raupp, Team Leader
Bruce Newman, Statistician/Data Analyst
Luis Revés, GSC Research, Deputy Team Leader
Carlos Lauchande, Statistician
Edward Jay Allan, Project Director
Merrill Anne Jordan, Program Associate

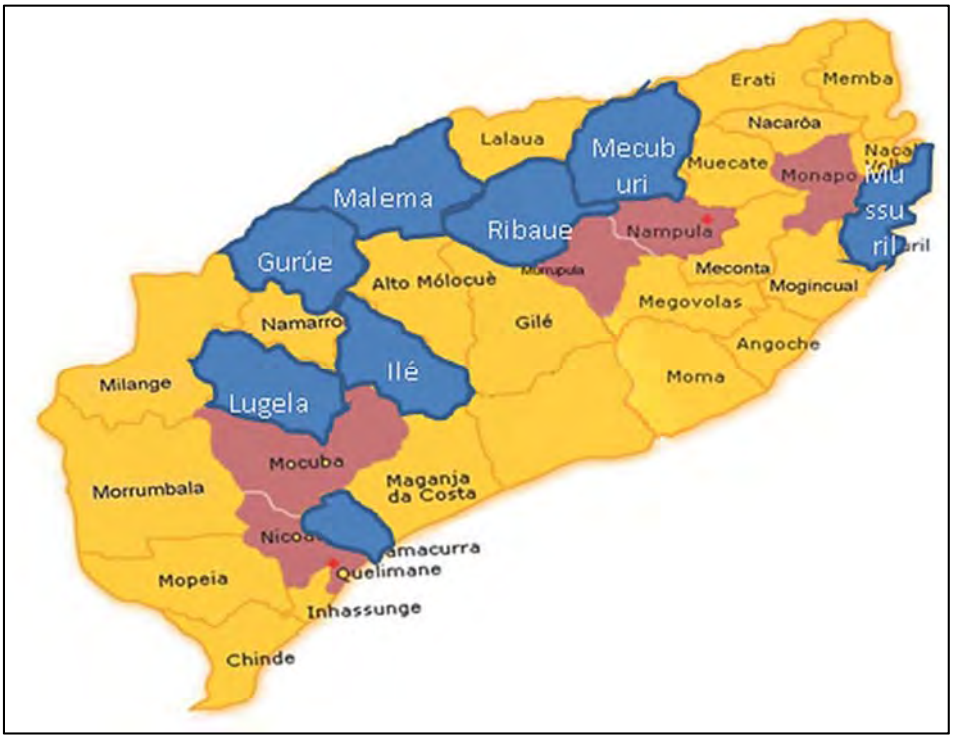
May 23, 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

MAP OF APAL



2012-2014 ApaL Program Districts

Shown in red

Nampula	Zambézia
Murrupula	Mocuba
Nampula City	Quelimane
Monapo	Nicoadala

Source: ApaL

ACRONYMS

ApaL	*Aprender a Ler* (Learn to Read)
CAP	Concepts about Print
CEA	Cost Effectiveness Analysis
clpm	Correct Letters Per Minute
cwpm	Correct Words Per Minute
df	Degrees of Freedom
DPEC	*Direcção Provincial de Educação e Cultura* (Provincial Directorate of Education and Culture)
EGRA	Early Grade Reading Assessment
IBTCI	International Business &Technical Consultants, Inc.
IE	Impact Evaluation
IFP	*Instituto de Formação de Professores* (Teacher Training College)
INSET	In-Service Teaching
LEI	Local Education Institution
M&E	Monitoring and Evaluation
MINEDH	*Ministério da Educação e Desenvolvimento Humano* (Ministry of Education and Human Development)
NS	Not Significant
PD	Pedagogical Director
RCT	Randomized Controlled Trial
RSA	Rapid School Assessment
SD	School Director
SDEJT	*Serviços Distritais de Educação, Juventude e Tecnologias (District Services for Education, Youth, and Technology*
SMA	School Management Assessment
SMT	School Management Tool
TIMSS	Trends in International Mathematics and Science Study
TLA	Teaching-Learning Aid
TOT	Training of Trainer
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
USAID	United States Agency for International Development
WEI	World Education, Inc.
ZIP	*Zona de Influência Pedagógica*

EXECUTIVE SUMMARY

Sustainability is frequently mentioned as a development project goal but efforts to sustain interventions beyond the conclusion of funding have only recently emerged as a growing area of emphasis for evaluation. And yet, the sustainability of interventions following the conclusion of formal support should be viewed as an essential element of programs that attempt to bring about meaningful behavior change — especially in the variety of settings with severe resource limitations, as is the case with development projects. Despite widespread acknowledgment that sustainability should be central to intervention efforts[1] (e.g., August et al. 2006; Johnson et al. 2004), specific attention to the process of assessing the sustainability of development interventions remains relatively sparse.

In 2012, the United States Agency for International Development (USAID), in collaboration with the Mozambique Ministry of Education and Human Development (MINEDH), designed the USAID/Aprender a Ler (ApaL) project to improve early grade reading skills and selected World Education, Inc. (WEI) as the implementer and International Business & Technical Consultants Inc. (IBTCI) to conduct an Impact Evaluation (IE) of the project. Both contracts, the one to implement ApaL and the one to conduct the IE, were awarded at essentially the same time. USAID believes that this Impact Evaluation of ApaL is the first external and independent IE of its kind for USAID in Africa. It is also one of the relatively few IEs designed from the beginning to measure the relative impacts of the interventions not only at the conclusion of the interventions but also to measure the lasting impacts of the interventions. Previous reports submitted to USAID/ Mozambique in 2013 and 2014 by IBTCI, under the Evaluation Services IQC task order, have focused on the effects of the ApaL program during implementation. This report specifically addresses the extent to which the effects observed at the conclusion of ApaL interventions in 2014 were sustained once direct support from the implementer ceased.

ApaL began with delivery of the program in 35 *Zonas de Influência Pedagógica* (ZIPs) [2] in six districts along the economic corridors of the provinces of Nampula and Zambézia. Based on experience during initial implementation, expansion to close to five hundred schools in the area was anticipated. [3] USAID selected these provinces because at the time of project design, they contained 42 percent of the school-age population of Mozambique. Furthermore, these heavily populated and rural provinces had posted weak education performance results compared to national averages. At the beginning of the 2013 school year, 180 schools clustered around the 35 ZIPs in the two provinces were randomly assigned by the IE to three groups—60 schools to Full treatment, 60 to Medium treatment, and 60 to a no-treatment or Control group. The Randomized Control Trial (RCT) methodology utilized by the IE ensured that the three groups were equivalent at Baseline. Thus, any differences in reading outcomes could be ascribed to the intervention, and the findings for Control schools would reflect what would have happened in the absence of any intervention.

This report presents the findings, conclusions, and recommendations of an IE of the USAID/ApaL project and constitutes the final phase of this three-year effort to determine the magnitude of the effects of the project on its beneficiaries. Specifically, the analyses (i) examined the degree to which gains obtained by students on the Early Grade Reading Assessment (EGRA) in 2014 have been sustained; (ii) documented which activities, processes and procedures implemented by ApaL in 2014 have remained after the cessation of direct support to the schools; and, (iii) explored whether school directors have the technical skills and the schools have the resources needed to continue activities implemented by ApaL.

---

[1] Johnson K, Hayes C, Center H, Daley C. Building capacity and sustainable prevention innovations: A sustainability planning model. Evaluation and Program Planning. 2004; 27:135–149.

[2] In Mozambique, a head school and surrounding 5-10 schools form a Zona de Influência Pedagógica (ZIP).

[3] ApaL expanded to 538 schools in 2015.

Detailed reports of the work done in 2013 and 2014 under the Evaluation Services IQC task order are on file on the Development Experience Clearinghouse (DEC); Annex B presents a list with links. Here we refer to the main findings of the evaluation, summarize findings observed in 2015, and discuss the conclusions reached, and the implications and recommendations for future direction. The information provided could assist MINEDH, USAID, and other donors in the allocation of additional resources both towards improving early grade reading outcomes per se and the broader learning environment. The findings included in this report can be grouped under three main areas.

Did the ApaL reading intervention improve reading outcomes?

Measurements using the Early Grade Reading Assessment (EGRA) methodology had compared Midline 1 (2013), Midline 2 (2014) reading scores obtained by Grade 2 and Grade 3 students at the treatment schools to scores of a Control group that received no intervention. The results showed that while overall performance remains low, large absolute and relative gains in reading performance were achieved in the intervention schools that were not achieved in Control schools. It must be noted that the EGRA scores at Baseline had shown no significant statistical differences between the three groups, thus reaffirming the success of the randomized sampling strategy.[4] In all comparisons made between 2013 and 2014, children in treatment schools, especially those in the Full treatment schools, outperformed their counterparts in the Control schools. Table 0-1 shows scores obtained by students by year and by treatment group for four EGRA subtests. All differences observed between Full and Medium treatment and Control groups are statistically significant.

Table 0-1. Scores at Midline 1 (Sep 2013), Midline 2 (Sep 2014) and Endline (Sep 2015)

EGRA Subtests and Maximums	Midline 1- 2013			Midline 2 - 2014			Endline – 2015		
	Full	Med.	Cont.	Full	Med.	Cont.	Full	Med.	Cont.
	Grade 2								
Letter Recognition (100 clpm)	9.5	7.4	4.7	19.9	17.2	5.4	10.7	7.2	5.9
Familiar Words (30 cwpm)	1.9	1.0	0.8	3.3	2.6	1.1	2.4	1.5	1.3
Reading Fluency (120 cwpm)	2.2	1.0	1.0	5.2	4.2	1.7	3.4	2.4	2.0
Reading Comprehension (4 items)	.07	.02	.02	.21	.12	.03	.25	.22	.14
	Grade 3								
Letter Recognition (100 clpm)	16.6	15.3	12.0	29.6	27.8	12.3	20.1	19.2	12.6
Familiar Words (30 cwpm)	3.9	3.3	2.8	8.0	6.0	3.2	5.7	4.8	3.3
Reading Fluency (120 cwpm)	5.3	4.4	4.3	14.6	12.0	5.2	8.8	7.4	5.0
Reading Comprehension (4 items)	.16	.12	.12	.53	.43	.15	.50	.36	.30

clpm = Correct Letters Per Minute; cwpm=Correct Words Per Minute

'Better' than Control or statistically significant, however, still does not mean "Good." While these results were encouraging, despite the training provided to teachers and to school and pedagogical directors and the influx of materials to improve reading instruction and learning, the great majority of children did not demonstrate the skills on oral reading fluency needed to read with comprehension and were far from reading the number of words per minute that have been established as benchmarks for developing countries such as Mozambique. Although the effects of the intervention were clear, it is not realistic to think that a six-month intervention could bring children close to benchmark levels. ApaL alone is not enough to overcome many of the underlying recognized constraints to quality education that exist in Mozambican schools—high absenteeism rates among students, teachers, and school directors, teachers with limited training in teaching reading, a very short school day, and persistent

[4] *Please note that comparisons between Baseline and Endline results are not meaningful. Due to delays in the implementer's work plan, the Baseline study was not conducted at the end of 2012 as planned but in the beginning of the 2013 school year. That means that 3rd graders assessed at Baseline were 2nd graders in 2012 who after 2/3 months of vacation were arriving in February 2013 to start Grade 3. Same with 2nd graders, who were in fact 1st graders returning after the summer vacation to start Grade 2.*

shortages in teaching and learning materials.

Among Grade 2 students there was no statistically significant performance difference between boys and girls in Full treatment schools. In the Medium and Control groups significant differences were found on the Letter Recognition and Familiar Word Reading subtests. Although individual subtest scores were not always significant, overall there is a clear tendency for favoring boys. In none of the 6 subtests across the three groups do girls outperform boys and this gap increases with greater subtest complexity. In Grade 3 these patterns crystalize. In all six subtests in Control schools, girls lag behind reaching about 60% of boys' performance on reading skills. In Full and Medium schools although boys outperform girls on four out of the six subtests, the differences favoring boys are much smaller than in Control schools. Table 0-2 compares scores obtained by boys and girls in three EGRA subtests: Letter Recognition, Familiar Word Reading and Fluency.

Table 0-2. Differences in performance of girls and boys by treatment group and grade (2015)

EGRA Subtest	Student Sex	SECOND GRADE			THIRD GRADE		
		Means by Treatment Group			Means by Treatment Group		
		Full	Medium	Control	Full	Medium	Control
Letter Recognition (clpm)	Girls	10.0	6.1	4.5	17.6	16.7	9.9
	Boys	11.4	8.4	7.4	22.7	21.7	15.7
	Sig. t-test	0.305 NS	**0.030**	**0.005**	**0.007**	**0.004**	**0.000**
	Girls/Boys	87.7%	73.3%	60.8%	77.7%	76.8%	62.8%
Familiar Words Read Correctly (cwpm)	Girls	2.2	1.1	1.0	4.8	4.4	2.5
	Boys	2.7	1.8	1.6	6.6	5.3	4.2
	Sig. t-test	0.241 NS	**0.015**	0.066 NS	**0.011**	0.128 NS	**0.003**
	Girls/Boys	81.9%	61.7%	63.8%	71.9%	81.8%	59.7%
Text Reading Fluency (wpm)	Girls	3.0	2.0	1.8	7.4	6.3	3.8
	Boys	3.8	2.8	2.3	10.3	8.5	6.3
	Sig. t-test	0.214 NS	0.137 NS	0.340 NS	**0.028**	**0.043**	**0.010**
	Girls/Boys	78.6%	73.6%	80.0%	71.7%	74.0%	60.7%

Thus IBTCI concludes that the ApaL program contributed significantly to reducing the increasing performance gap between boys and girls in Grade 3 and that there may be other factors that must be addressed in order to eliminate the continuing disadvantage of girls in school performance. We should also note that most of the Grade 3 students received support from ApaL in 2014, which explains the higher performance of boys and girls in the treatment groups when compared with that of boys and girls in Control schools.

On every single comparison of results from the EGRA, students in urban schools strongly out-performed their rural counterparts as shown in Table 0-2.

Table 0-3. Comparison of 2015 Grade 3 EGRA Mean Scores by Urban/Rural Locality

EGRA Subtest - Grade 3	Urban	Rural
Oral Comprehension	9.7	7.6
Concepts about Print	7.4	5.8
Letter Recognition (clpm)	25.2	14.4
Familiar Words Read Correctly (wpm)	7.2	3.7
Text Reading Fluency (cwpm)	10.8	5.7
Text Reading Comprehension	0.58	0.32

The Student Assessment Coversheet identifies some of the factors that could be related to these performance differences.

Table 0-4. Variables that affect performance of students in rural settings

Other Variables	Urban	Rural
Class Days Offered in July	20.9	11.8
Student Days Attendance in July	11.9	8.3
Frequent Portuguese use with Parents (%)	73%	48%

Given these large urban-rural differences, IBTCI also examined the relative effects of each of the two interventions within Urban and within Rural schools in 2015. In Urban schools Grade 2, we detected no performance differences between intervention and control groups. In Grade 3 small (12% – 14%) but statistically significant differences were observed only on the Oral Comprehension and the Concepts about Print subtests.

In Rural schools, on the other hand, large differences, of 50% or more, were found between the two intervention groups and the Control schools. In Grade 2 Full schools outperformed Control on all six subtests often by 85% or more—Medium schools did not outperform Control on any of the subtests. In Grade 3 both Full and Medium groups outperformed the Control group on Concepts about Print, Letter Recognition, Familiar Word Reading, Reading Fluency and, for Full treatment, also on Reading Comprehension.

But even then, Full treatment schools in rural areas generally do not perform as well as Control schools in Urban areas. From the data available, we see that Full treatment Rural schools had a significantly higher (50% more) number of student days attended; in Urban schools no differences were found between any of the groups.

What contributed to this achievement?

The RCT methodology utilized allows for direct attribution of the USAID/ApaL intervention to improved outcomes because the random sampling component eliminates the effects of potential unobservable differences between treatment and control groups on the outcomes. This allows one to conclude that being in a school receiving the ApaL reading program is the only significant factor that explains the differences between the performance of students in treatment and in Control schools. Thus, the significant statistical differences between scores observed between 2013 and 2014 are a result of the ApaL program.

Due to delays in start-up, the intervention was implemented for two months in the 2013 school year. Full implementation only occurred in 2014. ApaL was designed to integrate into Grades 2 and 3 classrooms of treatment schools major inputs that have been shown effective in improving reading scores: increased time to read, appropriate and sufficient teaching and learning materials, and improved methods to teach reading.[5] While both the Medium and the Full treatment schools received these inputs, a component specifically designed to improve school management processes and routines that supported these initiatives was added to the Full treatment schools. The reading program includes four major result areas related to early grade reading: strengthening the teaching of reading, improving teacher performance, increasing the availability of teaching and learning materials (TLAs), and enhancing school management processes to support reading instruction.

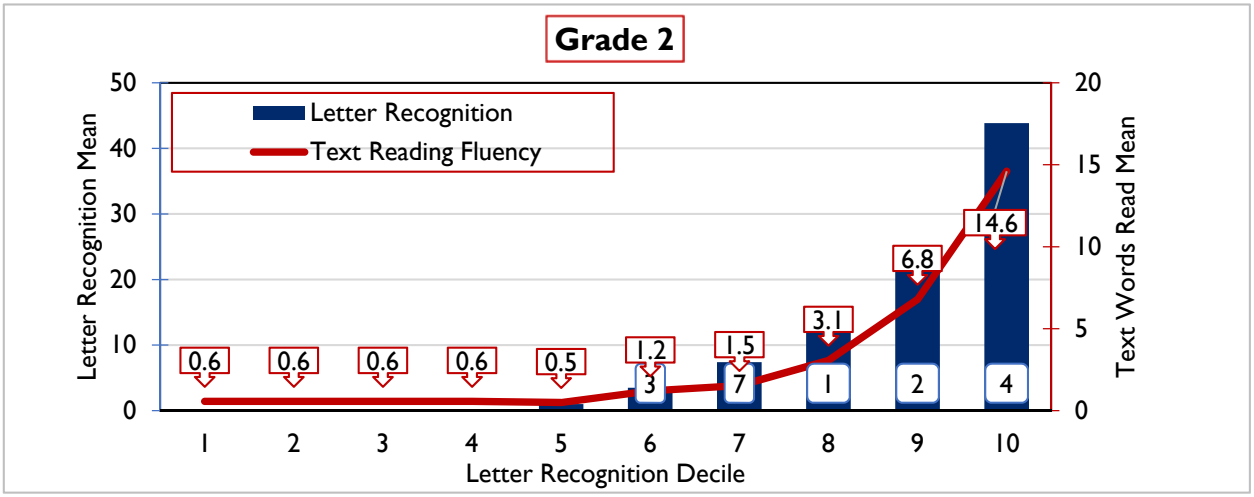
The results obtained allow us to conclude that by addressing fundamental reading skills such as phonological awareness and alphabetic principle in the early grades, in a systematic and sustained way, ApaL can and does improve student reading performance. Figure 0-1 shows the results of the quantitative analysis confirming that children who succeed in basic skills such as letter recognition and

[5] *Gove, A. and P. Cvelich (2011). Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute www.eddataglobal.org*

familiar word reading have higher oral reading fluency that allows them to read with comprehension—the ultimate objective of reading.

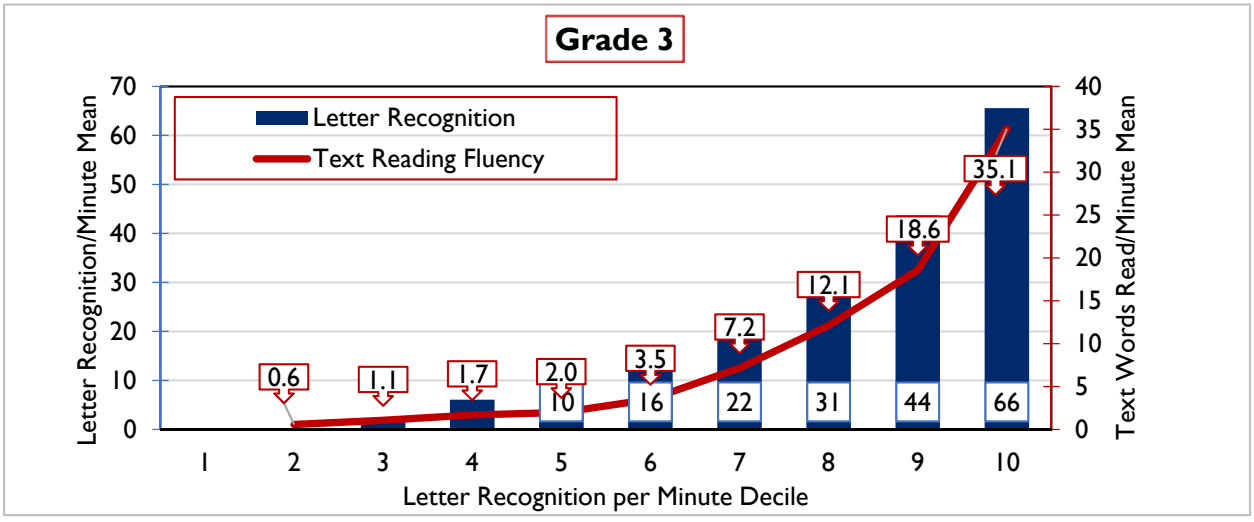
In Grade 2, for example, Letter Recognition is highly associated with Reading Fluency. The correlation between Letter Recognition and Text Reading Fluency is quite high (r = 0.65), which means that 42% of all variance in reading fluency is explained by letter recognition scores alone. As shown in Figure 0-1, extremely low levels of performance on this task mean that few students can recognize enough letters to be able to read, even slowly, words in a text passage.

Figure 0-1 Letter Recognition deciles as a predictor of Reading Fluency performance - Grade 2



This pattern is even stronger in Grade 3, where more students perform higher on both the Letter Recognition and Text Reading Fluency subtests as shown in Figure 0-2. Students who performed in the seventh decile (average 22 clpm) of the Letter Recognition subtest read 7 words per minute. Those students in the top decile (average 66 clpm) read five times as many words—35 cwpm. The correlation between Letter Recognition and Text Reading Fluency is stronger than in Grade 2 (r = 0.77), accounting for 59% of variance in reading fluency scores.

Figure 0-2 Letter Recognition deciles as a predictor of Reading Fluency performance - Grade 3



The relationships observed between all the pre-reading and reading skills measured by the EGRA for Grade 3 are displayed in Table 0-5. The information presented leads to the conclusion that (1) pre-

reading skills (Oral Comprehension and Concepts about Print are generally not strong predictors of reading skills (the variance explained ranges from 14 to 20%; (2) Letter Recognition is a strong predictor of Familiar Word Reading (69%), of Text Reading Fluency (59%), and contributes somewhat (30%) to Reading Comprehension; (3) Familiar Word reading is an excellent predictor of Text Reading Fluency (70%) and strongly associated with Reading Comprehension (43%); (4) Text Reading Fluency predicts fully 49% of Reading Comprehension scores.

Table 0-5. Correlations between Subtests and Percentage of Variance Explained

Grade 3	Concepts about Print	Letter Recognition	Familiar Word Reading	Text Reading Fluency	Reading Compre-hension
Oral Comprehension	**r = 0.58** (R^2 = 34%)	**r = 0.38** (R^2 = 44%)	**r = 0.38** (R^2 = 14%)	**r = 0.34** (R^2 = 11%)	**r = 0.45** (R^2 = 20%)
Concepts about Print		**r = 0.54** (R^2 = 29%)	**r = 0.49** (R^2 = 24%)	**r = 0.43** (R^2 = 18%)	**r = 0.45** (R^2 = 20%)
Letter Recognition			**r = 0.83** (R^2 = 69%)	**r = 0.77** (R^2 = 59%)	**r = 0.55** (R^2 = 30%)
Familiar Word Reading				**r = 0.89** (R^2 = 70%)	**r = 0.65** (R^2 = 43%)
Text Reading Fluency					**r = 0.70** (R^2 = 49%)

The importance of learning letters and their sounds, which facilitate the reading of words in order to reach the fluency necessary to read with comprehension—the ultimate goal of reading—cannot be overstated and must be at the core of any reading program. The inputs provided by ApaL in 2014 had a clear impact on student reading skills as shown by the consistent higher performance observed in the treatment groups, especially in the Full schools. However, scripted lessons, teacher and school director training, coaching and providing TLAs and reading materials are only one aspect of the solution to a persistent reading problem as observed in schools in Mozambique. Our findings suggest that additional efforts need to be made to reduce student, teacher and school director tardiness and absenteeism and to expand the time students spend learning to read in order to profit from interventions such as the ApaL reading program. This is not an easy job given the variety of factors and the many challenges surrounding school absenteeism.

Are the effects sustainable after the cessation of the intervention?

The USAID/ApaL reading program was able to significantly improve foundational reading skills in 2014, and some results persist in 2015 after the intervention has ceased for one year. Table 0-5 reflects the percent of retention of learning (significant differences are **bolded**) obtained at the end of 2014 after one school year of implementation and retained at the end of 2015 one year after ApaL had ceased its involvement at the schools selected for the IE sample.

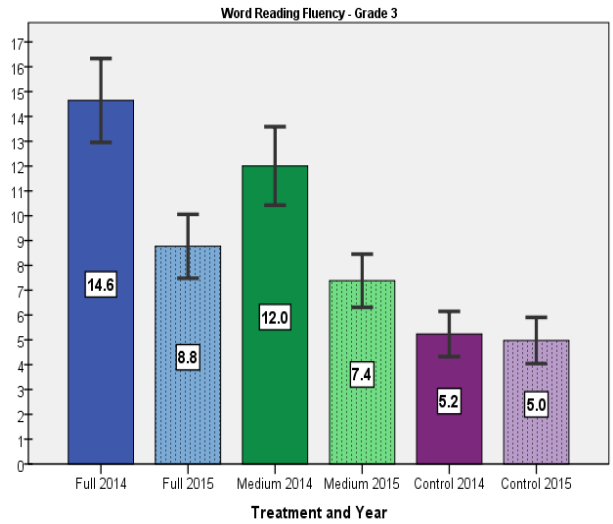
Table 0-6. Results obtained in 2014 vs. 2015

EGRA Subtests 2015 as % of 2014	Grade 2			Grade 3		
	Full	Medium	Control	Full	Medium	Control
Oral Comprehension	**92%**	**90%**	95%	**92%**	**64%**	98%
Concepts about Print	**88%**	**80%**	94%	**89%**	**92%**	97%
Letter Recognition	**54%**	**42%**	109%	**68%**	**69%**	102%
Familiar Word Reading	**71%**	**55%**	109%	**70%**	**74%**	102%
Text Word Reading	**65%**	**57%**	118%	**60%**	**61%**	95%

Figure 0-3 illustrates the sustainability of 2014 scores for text reading fluency in Grade 3, by treatment group. In the treatment groups, scores significantly declined in the no-intervention year: 60% of 2014

scores remained in 2015 for Full school students; 61% for Medium. Control groups did not change. While Full schools maintain a lead over Control schools, Medium schools have become almost indistinguishable from Control schools.

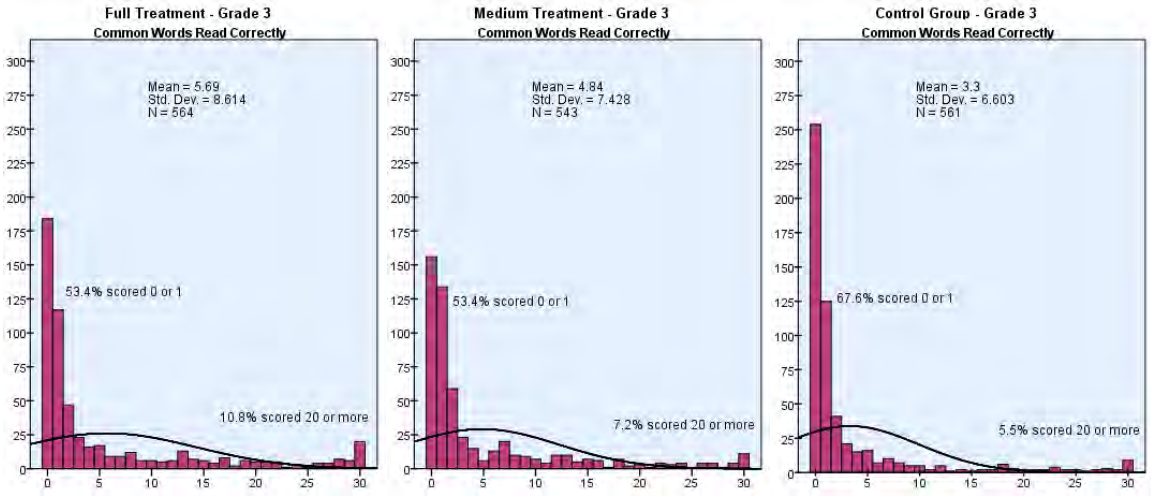
Figure 0-3 Text Reading Fluency by treatment group and year – Grade 3



Although as a result of the cessation of the program scores went down—as expected—on all EGRA subtests in both grades, the Full treatment schools continued to significantly out-perform Control schools in both Grade 2 and Grade 3. Medium schools generally did not perform better than Control schools in Grade 2, but did show better results relative to the Control schools in Grade 3 on four of the six EGRA subtasks. Grade 3 Medium schools were much more similar to the Full schools. The reason Grade 3 Medium treatment students performed better than Grade 2 students vis-à-vis Control students is probably due to the fact that most of Grade 3 students and most of their teachers would already have experienced a full year of ApaL intervention the year before. In 2015, Grade 2 students (except those who repeated) were not directly exposed to the ApaL program.

In short, from a technical standpoint, the intervention benefits students both at the conclusion of the intervention and also, to a lesser extent, one year after the intervention has ceased and schools have been functioning without further support. However, note that scores on the sub-test Text Reading Comprehension, the ultimate goal of learning to read, continue to be extremely low, as reading fluency remains a serious limitation for almost all students. Research shows that there is a strong correlation between fluency and reading comprehension and for this reason Oral Reading Fluency is often used as the best "composite" indicator of the ability to read. This is behind the establishment of the benchmarks of 20 words correctly read per minute at the end of Grade 2 and 40 at the end of Grade 3 for developing countries such as Mozambique. Students who do not reach these benchmarks will not be able to read with comprehension. Figure 0-4 shows the distribution of scores per treatment group and the percentage of students who scores 0-1 or 20 or more words correctly read per minute.

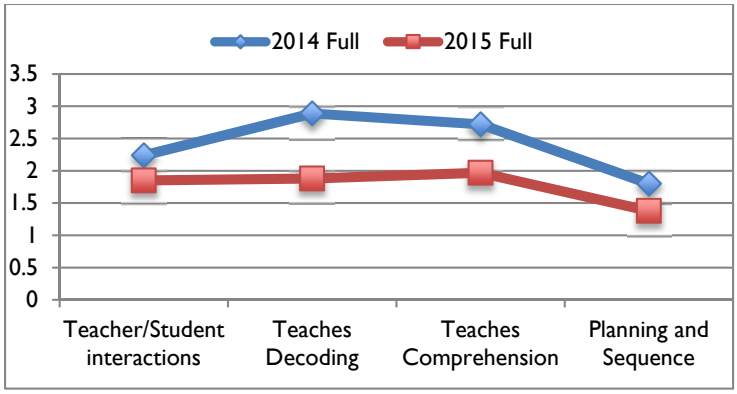
Figure 0-4. Distribution of Familiar Words Correctly read by minute per treatment group



Correct words per minute (cwpm)—both isolated words and words in a connected text—are the key skills to acquire in order to read with comprehension. This is where the efforts should be put if the goals is getting children to learn to read (which implies comprehending what you read). Note that in 2015 only 10.8% of third graders assessed in Full treatment schools, 7.2% of those in Medium and 5.5% of those in Control schools were able to read 20 or more correct words per minute, and even in treatment groups more than half of the children scored zero or read only one word per minute. Even though the differences between the groups are statistically significant, the real educational significance is doubtful when so many students score zero, even third graders that were part of the program in 2014.

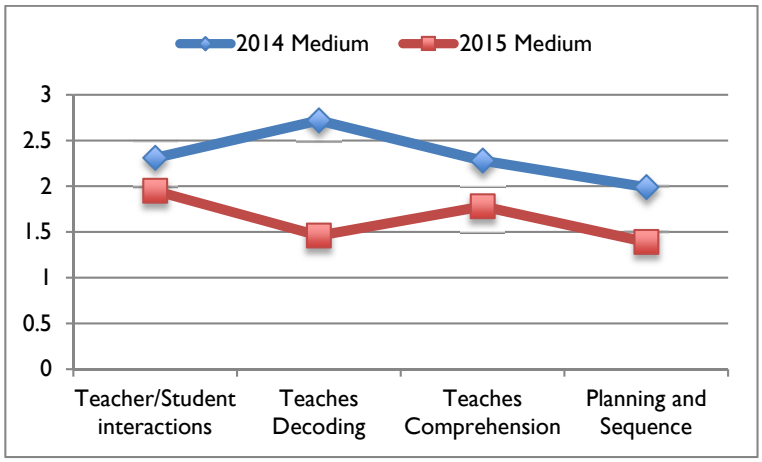
As noted, the treatment group advantages versus the Control group were not as great in 2015. About 70% of the teachers who had been trained in ApaL practices were retained at the same school teaching either Grade 2 or 3. Of these, they did not, for the most part, carry these practices on in 2015. Figures 0-5 and 0-6 contrast teacher observed instructional performance in 2014 and 2015 along four instructional behaviors measured by the SMA.[6]

Figure 0-5. Teacher observed practice by year and treatment (Full)



[6] *Data provided to the IE by ApaL M&E with descriptive analysis. The IE conducted further analyses to reach a composite score that allow for multiple comparisons.*

Figure 0-6 Teacher observed practice by year and treatment (Medium)



Training provided to Full school directors combined with the greater stability of school directors in their position contributed to the larger impacts observed in Full schools. Student reading material possession and classroom teaching and learning aid (TLA) materials dropped from 50% to 60% of their 2014 levels. Nonetheless, enough early grade reading gains remained throughout the two-year period so that, combined with the reduction of unit costs, the cost-effectiveness of the ApaL interventions improved relative to that reported for the 2014-only school year.

Over the two-year period, Full schools show stronger cost-effectiveness results relative to the Medium treatment, reconfirming that the best model of intervention is the Full treatment model. In 2014, the two interventions were examined from a cost-effectiveness perspective: dividing the unit cost per student by the gains observed on each EGRA subtest. To extend the cost-effectiveness analysis to cover the 2015 year required spreading the 2014 costs over two school years, recalculating the beneficiaries (enrollment totals) and developing new unit costs. These unit costs were significantly reduced from $11.54 for Full treatment in 2014 to $6.04 in the combined 2014-2015 period. Medium treatment per unit costs declined from $9.13 to $4.75. This combined with the gains observed reaffirms the cost-effectiveness of the ApaL program.

Summary Recommendations

Many of the basic recommendations made in the Midline 1 and 2 reports are still valid. Specifically, (a) increasing time spent learning basic literacy skills, (b) teaching sounds of all letters of the alphabet, (c) providing students with books that focus on decoding and word-building skills, training teachers to use these strategies in the classroom, (e) putting into place strategies to encourage students and their families to increase the number of words read, and (f) especially providing school directors with the training needed to improve their schools. There are challenges to be faced when attempting to modify behaviors and procedures that have been in place for a long time but ApaL is making an effort to implement the recommendations.

Recommendations 1, 2, 3, 4, and 10 are primarily for USAID to consider in conjunction with its general project design and management activities. The other recommendations are primarily for USAID and MINEDH dialogue and consideration.

1. Make sustainability part of the implementation plan of an intervention. Sustainability should not be an afterthought to be addressed at the end of an intervention. A specific description of the measures that will be taken to make the intervention sustainable should be required as an important element of the design of a project. ApaL has worked with district officers and school directors to make

sure that trained teachers remained in in their schools and classes, to ensure that skilled trainers are located in each ZIP/district, and that teaching and learning materials continue to be used.

2. Strengthen project Monitoring and Evaluation (M&E). A well-developed Monitoring and Evaluation (M&E) component with specific and measurable indicators should be required from implementers as part of the project design, independent of the external evaluation of project results and impact. In addition to the monitoring of project activities and other inputs, greater focus on the achievement of outcomes represented by well-defined and agreed upon indicators will enhance the implementation of a project and provide information to correct its course when necessary.[7]

3. Adopt a non-linear implementation model to provide support more or less intensely as required in different situations. Rather than follow an implementation-and-immediate-scale-up model, interventions such as ApaL may require a more extended period of support to allow for internalizing and routinization of activities. Withdrawing the support according to a fixed schedule (e.g., the end of one school year) without considering the level of routinization achieved may significantly decrease both the effects and the sustainability of an intervention.

4. Assess the level of readiness for the innovation prior to the implementation of the intervention. A range of instruments has been developed and used to identify specific concerns of potential adopters and stakeholders. These instruments provide insights into issues that must be addressed to ensure widespread acceptance, adoption and sustainability of an intervention.

5. Implement the program where the effects are greater. Findings suggest that the program functions better in some settings than in others. For example, on an absolute basis, children in rural schools consistently show lower scores than do children in urban schools—in fact, on average, students in Full treatment rural schools performed worse than students in urban Control schools. This notwithstanding, our findings indicate that, in 2015, ApaL's effects in rural schools are larger 2015 the effects observed in urban settings. That is, although ApaL does not eliminate the relative deficit in learning for rural children, it significantly reduces it. The findings suggest that the program, as designed and implemented, may be more appropriate for rural schools, and that in the future, those schools should become ApaL's main target.[8]

6. Identify the reasons for girls' consistent under-performance and include in all projects strategies to close the gender gap. Both Full and Medium ApaL interventions reduced the difference in performance between boys and girls, when compared to Control schools in the same grade. The program benefited both boys and girls equally but has not eliminated the gender gap observed in 2013 and 2014. Findings show that while ApaL narrows the relative gap between boys and girls, the absolute gap widens as students advance to Grade 3 and when more complex skills are assessed. The increased magnitude of sex differences in Grade 3 suggests that greater efforts will be required to address the root causes of under-performance of girls.

7. Work with MINEDH to improve the ways school directors are selected and prepared. School directors are key to educational improvement. This is shown first by the higher scores of students in Full treatment schools, which is probably closely associated with greater attendance by teachers and students both, and confirmed by the finding that scores were higher at schools where school directors received additional training in 2015. We recommend that ApaL work with MINEDH and district leadership to identify staff with leadership potential to become a school director and

---

[7] Indicators for project outcomes need to be specific, measurable, attainable, relevant and time-bound. For example, Improving reading skills or Expanding time available to teach reading do not qualify as SMART indicators. Indicators developed during the design phase can be adjusted, if necessary, based on the findings obtained at Baseline.

[8] ApaL has indicated that, with the exception of treatment schools in Nampula City and Quelimane, all schools in the project 2015 and 2016 scale-up are rural schools.

develop a clear set of criteria for applicants. Second, because many principals learn the skills they need on-the-job, they need continuous in-service opportunities learn how to improve school management.

8. Provide incentives to keep trained teachers at the school teaching the early grades. The management component received by Full treatment schools contributed to a higher rate of retention of trained teachers in those schools. Nonetheless, approximately 30% of the teachers trained in 2014 either left the school or were assigned to other grades. This highlights the need to develop and strengthen incentives to keep trained teachers in the same grades in subsequent years.

9. Use more effectively the data produced by the School Management Assessment (SMA) instruments, checklists, and assessments. This would require the reduction of available data to rate schools in simple categories of school management aspects. The "grades" assigned would show clearly where a school stands in terms of factors that relate to student performance such as student, teacher and staff attendance, days of class offered, or hours of instruction per day. This will reinforce the use of data to identify and monitor school management factors causal to student performance. This type of exercise should also form part of the MINEDH school supervision process.

10. Consider the sustainability of the various inputs provided as part of the program when designing future projects. Some inputs provided by ApaL, such as TLAs, seem to have only a moderate level of sustainability (50% - 60%) while others were even less sustainable. For example, given the amount of teacher training conducted, only decoding activities showed any difference in the classroom a year after implementation. All inputs require on-going assistance in order to continue. We must identify ways to improve the sustainability of the inputs themselves and ensure that necessary funding will be available to maintain them. This needs to be a part of the intervention design.

The body of the report presents more details on the context, the intervention, the RCT methodology. Findings on the results of the intervention are followed by conclusions and recommendations.

1. CONTEXT

Over the past decade, Mozambique has made significant progress in reducing its out of school population. Net enrollment rates increased from 56% in 2000 to 92% in 2010. Yet despite this progress, it has been estimated that over 1 million children are out of school, and most of them live in rural areas and marginalized communities. Mozambique's poorest children are four times more likely to be out of school than children from the higher-income households. Student absenteeism reaches 60% and retention throughout the grades is considered a serious problem and more than a quarter of children are estimated to drop out before completing grade 6.[9]

The Government of Mozambique has taken three important steps toward realizing universal primary school education: 1) Enacting compulsory education requiring all children between 6 and 12 years old to attend primary education; 2) Extending primary school cycle from five to seven years; and 3) Abolishing school fees for all of these seven grades (UNESCO, 2008). These actions have increased net primary enrollment rates by 35 percentage points to 80% in 2005 and rural-urban disparities in enrollment have decreased (UNESCO, 2008).[10]

The March 2015 World Bank Service Delivery Survey (SDI) conducted in seven African countries

---

[9]UN Special Envoy for Global Education (April 2013). Accelerating progress to 2015: Mozambique. Working paper; World Bank at http://data.worldbank.org/indicator/SE.SEC.NENR?order=wbapi_data_value_2012+wbapi_data_value&sort=asc; Spaull, N. and Taylor, S. (2015) "Access to What? Creating a Composite Measure of Educational Quantity and Educational Quality for 11 African Countries," Comparative Education Review 59:133-165

[10] Mongoi, D. et al. (2010). "Endline Report of Early Literacy among pre-school and primary school children in Mozambique." Save the Children.

concludes that Mozambique is well placed to achieve universal primary education.[11] However, the quality of primary education remains a critical challenge. Strong evidence of the need for improved reading instruction in the early grades in Mozambique came from the results of the Trends in International Mathematics and Science Study (TIMSS) assessments conducted in 2003 and 2007. In TIMSS, Mozambique was ranked the lowest of 36 countries assessed, largely because of low reading levels. As explained by teachers, the low performance of their students was not specifically related to the misunderstanding of math and science concepts, but rather a result of their inability to read and understand the test questions.[12]

More evidence emerged from the Aga Khan Foundation study, "Cabo Delgado: Mozambique Baseline Report," in early 2011. Results indicated that there are large percentages of children in each grade that know less than 60% of the letters of the alphabet. A strong correlation between students' letter knowledge and student word reading ability was found, suggesting that increased instructional focus on alphabetic awareness may lead to improved reading outcomes, especially for those children with the lowest current levels of reading ability. The assessment findings from Cabo Delgado provide strong evidence that the instructional approach should be changed to one, which incorporates phonics instruction.[13]

A third study prepared by RTI International (through EdData II) and supported by United States Agency for International Development (USAID) and the Mozambique Ministry of Education and Human Development of Mozambique (MINEDH) reinforces these findings. The study investigated whether students were developing foundational reading skills, and, if not, where efforts might be best directed.

> *"Mozambique is poised to meet the MDG goal of 100 percent primary education enrollment but the achievement will have limited impact if the quality of the education that pupils have access to is lacking."*
> Mozambique Service Delivery Indicators: Education, World Bank, March 2015

The Early Grade Reading Assessment (EGRA) tool was administered in 2011 to a stratified random sample of 735 students in grades 2 and 3. The study revealed that most students in both grades 2 and 3 were not reading fluently. Students in grade 2 read on average 5.8 correct words per minute (cwpm) and 42% could not read a single word. Students in grade 3 read on average 12 cwpm and 27% were unable to read a single word.[14]

Reacting to these findings, the Government of Mozambique requested assistance, and in 2012 USAID funded World Education Inc. (WEI) to collaborate with MINEDH to implement the USAID *Aprender a Ler* (ApaL) project in two provinces (Nampula and Zambézia), aimed at improving reading skills in the early grades of primary school (AID-656-C-12-00001). At the same time, USAID contracted International Business & Technical Consultants, Inc. (IBTCI) under the Evaluation Services IQC to conduct an Impact Evaluation (IE) using an experimental research design to assess the impact of ApaL on student reading competencies. The IE design included a counterfactual—a Control group of 60 schools similar in all identifiable aspects to the 60 schools in each of the two treatment groups—achieved through *a priori* random assignment of *Zonas de Influência Pedagógica* (ZIPs)[15] to two treatment groups and a Control group. One valuable point to note is that because both ApaL and the IE began together,

---

[11] World Bank (2015). Mozambique Service Delivery Indicators: Education

[12] Aggarwala, N.K. (2004). "Evaluation Report: Quality assessment of primary and middle education in mathematics and science." Retrieved from http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/Aggarwala_UNDP_Evaluation_Report.pdf. Accessed 2013 June 15.

[13] Gavin, S. (March, 2011). "Literacy boost: Mozambique baseline report." Retrieved from http://resourcecentre.savethechildren.se/library/literacy-boost-Mozambiquebaseline-report. Accessed 2013 June 20.

[14] Collins, P. and Messaoud-Galusi, S. (2012). Student Performance on the Early Grade Reading Assessment (EGRA) in Mozambique. EdData II report prepared by RTI International for USAID. Retrieved from http://www.eddataglobal.org/documents/index.cfm/4-

[15] Schools in Mozambique are clustered (usually in groups of 5 – 7) around one lead school to constitute a Zona de Influência Pedagógica (ZIP).

it was possible to work collaboratively, along with MINEDH, to develop the universe of schools to be sampled *before* intervention began, rather than after the fact, a situation which happens not infrequently.

WEI began activities in the schools at the beginning of the 2013 school year[16] and continued throughout the year to pilot and develop strategies for full program implementation; by the end of the 2013 school year, schools had received two months of intervention of the reading program. The project was fully implemented during the 2014 school year. The IE entailed four data collection events: (1) Baseline measurement in February-March 2013 ; (2) Midline 1 assessment conducted in September 2013 after approximately two months of partial implementation;[17] (3) Midline 2 measurement in September 2014, after one full year of implementation; and, (4) a measure of the sustainability of ApaL conducted at the end of the 2015 academic year, with sustainability defined as the measure of student performance one year following the cessation of ApaL activities in the treatment schools. Data were collected jointly by ApaL and the IE.

2. THE USAID/*APRENDER A LER* INTERVENTION

In this section we provide a summary of the USAID/Aprender a Ler (ApaL) early grade reading program. The main purpose of ApaL is to improve the quality of reading instruction through enhancement of teacher instructional practices and behaviors, provision of teaching and learning materials (TLAs), and increased availability of supplemental student reading materials. In addition, ApaL seeks to expand the quantity of reading instruction by including a school management component in the Full intervention. This component addresses issues of school directors' leadership and school management skills needed to support reading instruction and expand the time students are exposed to learning. ApaL also provides a training model and materials to build capacity of Lead Trainers to train teachers at the ZIP level. Pedagogical directors (PDs) and Cycle Leaders are also trained to act as resources to teachers on pedagogical matters, to provide classroom coaching in early grade reading, and in the use of TLAs. Training in school management is also provided to school and pedagogical directors to foster systemic and high-quality early reading instruction in Portuguese. In 2014, the target population for the project was approximately 45,000 Grade 2 and Grade 3 students enrolled in 120 schools. In sum, the main targeted result areas of the USAID/ApaL project are:

1. Improved quality of reading instruction for second and third graders in target schools.
2. Increased quantity of reading instruction for second and third graders in target schools.

As designed, ApaL engages beneficiary schools for one year. The Aprender a Ler approach to improve the teaching of reading in the initial classes includes the following components for both Medium and Full treatment groups:

- **300 systematically organized lessons.** One lesson per day, focused on grades 2 and 3. Each lesson is 45 minutes of reading instruction and includes seven steps for the teacher to follow: (1) Review of previous lessons (not just sound and letters, but specifically words); (2) Phonemic awareness: identifying the sounds of letters in words; (3) Phonics, decoding words; (4) Fluency, practicing reading words, sentences and connected text with decodable books and flashcards; (5) Vocabulary and comprehension, practicing listening comprehension and learn new vocabulary using the "read aloud" books; (6) Writing, review of taught letters and words in the lesson; and (7) Homework.

---

[16] *The Mozambican school year runs from January/February to September/October.*
[17] *This partial implementation was essentially a pilot.*

- **Teaching-Learning Aids (TLAs).** Alphabet charts to be permanently posted in the classroom; key word cards[18]with letters and corresponding images; letter charts (*quadro de pregas)* that allows individual letter cards to be posted on the wall to form words from letters; decodable books,[19] and "read aloud" books.
- **Continuous assessment.** Every four weeks a written assessment is conducted, allowing teachers to track progress of individual students and determine which reading tasks students had difficulty answering.
- **Fluency assessment**. Conducted in weeks 9 and 20 of the program when the teacher and the reading coach individually assess all students allowing teachers to quickly determine which students need extra support in developing reading skills.
- **Training Manuals** for Master Trainers and Training of Trainers (TOTs) and supervisors including strategies on how to structure meetings and give constructive feedback (coaching), Rapid Assessment supervision and coaching cycle, interview.
- **Training and coaching.** Delivery of up to 87 hours of training starting in the beginning of 2014. Teacher training sessions were held on Saturdays targeting all 849 Grade 2 and 3 teachers in the intervention schools. Training followed an enhanced cascade system and worked with teachers on the scripted lessons, the use of teachers' guides and the TLAs provided by the project, and the development of teaching-learning materials. To avoid taking teachers away from their classes training periods were held on Saturdays, initially for the whole day and later reduced to five hours per session.

In addition to the above, the Full treatment schools received a component aimed to improve school management, thought to support the teaching of reading, and to increase the quantity and the quality of hours and days dedicated to the teaching of reading. School directors received focused training in school management routines, class observation, organization of TLAs, and similar.

ApaL selected and adapted strategies through a collaborative process between project staff and MINEDH provincial officers that included organized meetings, weekly participant observation, extensive informal dialogue, opportunities for modeling and demonstration, practice with performance feedback, trial-and-error implementation, and problem-solving. Intervention strategies focused primarily on teacher practices selected due to their demonstrated ability to improve reading skills in early grades and on director school management strategies to support reading instruction.

3. METHODOLOGY OF THE IMPACT EVALUATION/ RCT

Section 3 summarizes the methods utilized to address the fourth evaluation question that the IE needs to answer: *Of the most cost-effective interventions, which fall within the existing technical and financial management capacity of local education institutional personnel?* or the likelihood that processes and initiatives implemented by the ApaL project would continue after the cessation of the project. A listing of the methods, instruments and procedures utilized by the IE and the findings in the first three years, as on file in the Development Experience Clearinghouse, can be found in Annex B.

There have been five stages to this IE: (1) Baseline data collection and analysis at the beginning of the 2013 school year prior to project implementation; (2) Midline 1, near the end of the 2013 school year, after two or three months of implementation; (3) Midline 2, after a full year of implementation, 2014; (4) a study of the cost effectiveness of the intervention that coincided with Midline 2; and, (5) the current

[18] *Key word cards are self-made teaching aids with words on them to practice fluency.*
[19] *Decodable books are small, inexpensive, four- or six-page books with controlled text difficulty that the students are allowed to take home but must bring back. Almost 900,000 of them were distributed to the treatment schools.*

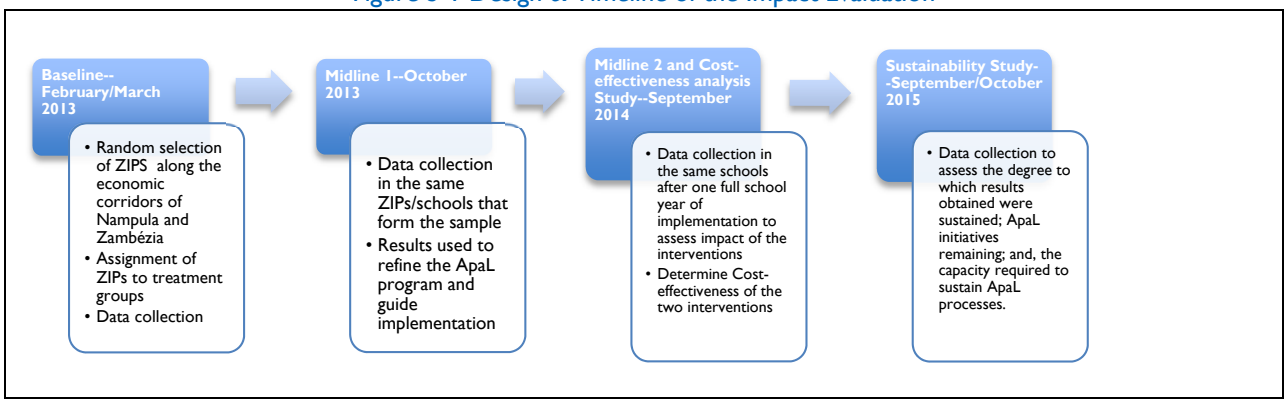
study focused on sustainability. The first four stages corresponded to the pre-implementation and the full implementation phases of the ApaL project and took place under the Evaluation Services task order. The fifth (sustainability), which was part of the design of the IE from its inception and which is the object of this report, examines issues related to the sustainability of the intervention one year after ApaL intervention ceased at 120 schools originally selected as treatment. Originally, the sustainability study was schedule for the end of 2015. Due to delays in the start-up phase of the ApaL project, full implementation only occurred during 2014 and the sustainability study was postponed until 2015, thus requiring a new contract approximately nine months after ApaL intervention ceased at 120 schools originally selected as treatment.[20] Figure 3-1, on the next page, shows this timeline.

This section describes the methodology utilized to conduct the IE of the USAID/Aprender a Ler project. We provide an overview of the research approach, describe the sampling strategy, detail the instruments utilized for data collection, and advance the methods of analyses used and addresses their adequacy to the research question.

The findings presented in this report are based on the analyses performed on systematically collected data at a random sample of 180 schools (60 control and 120 intervention schools) participating in the ApaL project from January 2013 to the end of the 2015 school year. In districts selected by USAID/ApaL along the economic corridors of the Nampula and Zambézia provinces, three groups of ZIPs and their cluster of schools were randomly assigned to either Full or Medium treatment groups or to a no treatment or Control group.

Baseline data were collected from mid-February to mid-March 2013, prior to the start of the intervention, in order to inform the implementer prior to the implementation of the project and to confirm the non-significant difference between groups on every measure selected—Early Grade Reading Assessment (EGRA) and the School Management Assessment (SMA). In October 2013 a second data collection event took place at the same 180 schools after approximately two months of intervention. Results obtained at that point were essential to inform and refine ApaL implementation plan. In September 2014 data were again collected in the same schools to capture the impact of one full school year of exposure to treatment and compare results to those obtained by students in the Control schools that did not have the benefit of the project. The IE also sought to determine whether the School Management component—from which only the Full treatment schools benefitted—had resulted in added benefit to students, e.g., whether students in Full schools scored higher on the EGRA than their counterparts in Medium treatment schools.

Figure 3-1 Design & Timeline of the Impact Evaluation



The primary outcome of interest of the study is the student level of reading competency at Grades 2 and 3. To assess student reading skills the Early Grade Reading Assessment (EGRA) tool was

[20] *The control schools are scheduled to benefit from ApaL in the 2016 expansion.*

administered by trained and supervised assessors to ten randomly selected Grade 2 and 10 randomly selected Grade 3 students in randomly selected classrooms of the sampled schools. The IE focused on student reading outcomes first and then, in order to address the multiple domains of interest in this study, on a number of variables that could explain the results obtained. These intermediate variables included student characteristics, teacher instructional performance during the reading instruction period, availability and utilization of teaching-learning aids (TLAs), and, observable school management practices in Full treatment schools.

In line with the 2011 USAID Evaluation Policy, the Education Strategy, and associated USAID documents and general "best practice" in measuring causal impacts, the IE utilizes a randomized controlled trial (RCT) methodology with a counterfactual—e.g., a control group similar to the treatment groups—to allow estimation of the impact of the project's interventions and measure what would have occurred without the intervention. The IE model treats early grade reading outcomes as a function of the Medium and Full interventions and the development hypothesis is tested under three scenarios: with the Medium treatment sample, with the Full treatment sample, and with the Control group sample. The IE compares reading scores in schools that have received the Full and the Medium treatment to those that did not receive any intervention. The results obtained at the Control schools represent the level of reading skills to be expected without the benefit of the project.

The use of RCT methodology is the most effective way to measure the impact of a project or program for three main reasons. First, it allows for direct attribution of the Aprender a Ler intervention to improve outcomes because the RCT model controls for all other possible determinants of the outcomes. Second, the random sampling component of this RCT eliminates the effects of potential unobservable differences between treatment and control groups on the outcomes. Third, an RCT is a rigorous evaluation method to obtain accurate and valid results to inform plans to scale up the most effective and cost-effective interventions**.** As described in the Inception Report, particular care was taken to minimize the potential of contamination across the groups under study.

The quantitative data collected through the EGRA administration answers questions such as *who* was involved (second and third grade students), *where* (180 schools in two provinces), and *how much* (difference in EGRA scores obtained by students in different groups or from year to year). Additional questions such as *what* happened in the schools and *how* the program was implemented require the description of the intervention and the collection of qualitative data. In the case of ApaL, collecting these data involved talking to and observing teachers and the school environment (e.g., when instruction begins, delays in the start of the school day, school management routines utilized by the school directors, etc.). The results of the SMA instrument provided by ApaL have been integrated into this report, as has the information obtained by interviewing 96 school and pedagogical directors at Full and Medium treatment schools.

3.1 Impact Evaluation Question for 2015

The overarching question of the IE is: To what extent have USAID/Aprender a Ler treatment interventions improved early grade reading outcomes for students in second and third grades in the target schools in Nampula and Zambézia Provinces?

The specific questions to be addressed by the IE are:

Question 1: To what extent does the "reading instruction support" treatment intervention cause early grade reading outcomes to improve for students in grades two and three in target schools whose teachers have received training, coaching and support?

Question 2: To what extent does the treatment intervention of additional "school management" training, coaching and support to school directors cause a significant and additional improvement in early grade reading outcomes when coupled with "reading instruction support" in target schools?

Question 3: Cost-effectiveness - To what extent are the "Medium" and "Full" treatment interventions cost-effective? Specifically, what are the most significant reading outcome effects and unit costs per student, per teacher, per school director, per school of the key treatment interventions?

Question 4: Management Sustainability – Of the most cost-effective interventions, which fall within the existing technical and financial management capacity of local education institutional personnel? What capacity-building activities would be required to ensure sufficient MINEDH technical and financial management capacity to implement the interventions?

The first three questions were addressed under the IQC task order with respect to findings at the conclusion of ApaL intervention. The continuation of the IE through this contract measures the lasting impact of ApaL – its sustainability—as indicated by student learning as measured at the end of the 2015 school year, a year after the schools have ceased to benefit from U.S. government support. The evaluation question to be addressed by the IE in 2015 is stated as: *Of the most cost-effective interventions, which fall within the existing technical and financial management capacity of local education institutional personnel?* This question, which could only be answered in 2015 after ApaL had ceased its direct involvement with the schools in the sample, required first a comparison between and among treatment groups in 2015 and second a comparison between results obtained in 2014 and in 2015. The third aspect, refers to the capacity-building activities that would be required for MINEDH to continue sustain an intervention such as ApaL and is addressed in section 5.9 of this report.

3.2 Data Collection Instruments

Based on the data requirements for the IE, five instruments were adapted or developed to collect the necessary data. These instruments are attached in the Annexes. Based on the data requirements for the IE, five instruments were adapted or developed to collect the necessary data. These instruments are attached in the Annexes.

1. EGRA Instrument—adapted by WEI and by the IE team from the instrument used in Cabo Delgado in 2011. The EGRA was administered to 3,475 randomly selected students (Annex C);
2. Student Interview—developed by the IE team and administered by WEI trained enumerators to each student assessed in randomly selected classes (Annex C);
3. Teacher Interview— developed by the IE team and administered by WEI trained enumerators administered to the teachers whose class was randomly selected for assessment (Annex D);
4. School Management Assessment (SMA) package adapted by WEI from the instrument used in Cabo Delgado in 2011 by the project implemented by the Aga Khan Foundation and administered by WEI trained enumerators includes the structured interview for school directors, the Classroom Observation Instrument, and the Classroom Inventory (Annex D); and,
5. Semi-Structured Interview Protocol—developed by the IE team and administered by the Deputy Team Leader and the IE supervisors to school directors and pedagogical directors (Annex F).

3.2.1 Early Grade Reading Assessment (EGRA) Instrument

The ability to read and understand a simple text is one of the most fundamental skills a child can learn. Yet, measuring early reading can be challenging since most tests are administered in higher grades, such as grades 4 or 6. Because these tests are aimed at higher-level skills, they are not likely to capture the specific fundamental or emerging skills that students need to become fluent readers.[21] Early assessment

[21] *Emergent reading skills are "skills, knowledge, and attitudes that are developmental precursors to conventional forms of reading and writing. These skills are the basic building blocks for how students learn to read and write." (Connor et al, 2006, p. 665).*

of the pre-reading and foundational skills required for fluency allows the implementation of measures to correct deficiencies where they exist.[22]

The EGRA tool offers an opportunity to determine whether students in the early grades are developing the fundamental reading skills, and, if not, where efforts might be best directed. The EGRA has been adapted and used in over 50 countries and can capture more subtle impacts from specific teaching approaches than pencil-and-paper tests, as it incorporates subtasks that measure pre-reading skills.[23]

The assessments are administered orally and individually, when needed using the students' native language to ensure that they understand the instructions for each task. In Mozambique, the ApaL enumerators were instructed to use the local language of the student, when necessary to explain the task. However, given that the language of instruction at schools is Portuguese, the test itself was conducted in Portuguese. Administering the EGRA Instrument for grade 2 and grade 3 took between15 and 25 minutes per child.

Brief Explanation of EGRA Sub-tasks. The IE used the EGRA to determine children's competency in six sub-tasks, as explained below.

1. *Oral comprehension* measures ability to understand basic Portuguese oral vocabulary. The first part of this subtask includes 8 prompts that required students to perform an action (e.g., "*show me your arm*"). A second part, with a maximum score of 6, requests that students follow instructions given orally (e.g., "*place the pencil on the paper"*). The maximum score is 14.

***2. Concepts about print* (CAP)** measures children's emergent reading skills by asking them to demonstrate how they read a book—recognition of the front and back covers, direction in which to read, identifying the title of the story, location of page numbers, etc.[24] The maximum score was 10.

3. Letter recognition assessed ability to provide the names of the letters of the alphabet naturally and without hesitation. This is a timed test that assesses automaticity and fluency of letter recognition and measured in letter names correct per minute. Students were shown a chart containing 10 rows of 10 random letters (in uppercase and lowercase) and asked to name as many letters as they could [25] within one minute yielding a score of correct letters read per minute (clpm).

4. Familiar word reading assessed students' skill at reading high-frequency words. Recognizing familiar words is critical for developing reading fluency. In this timed subtask, children were asked to sound out as many words (in a list of 30) as they could within one minute, yielding a score of correct words per minute (cwpm).[26]

5. Oral passage reading assessed students' fluency in reading a passage of grade-level aloud and their ability to understand what they read. There are two parts to this subtask:

a. <u>*Oral reading fluency*</u>: As described above, the ability to read passages fluently is considered a necessary component of reading comprehension. In this subtask, students were given a second 120-word story and were asked to read aloud in one minute. The oral reading fluency score was the

---

[22] Abadzi, Helen. (2009). "Instructional Time Loss in Developing Countries: Concepts, Measurement, and Implications." World Bank Research Observer. 24 (2): 267-290.

[23] The Early Grade Reading Assessment (EGRA) is a 15-minute test originally developed by the Research Triangle Institute (RTI) administered orally to students in the early grades of primary school. As pointed out by RTI, the EGRA evaluates students' foundational reading skills, including pre-reading skills like phonemic awareness and listening comprehension, which have been shown to predict later reading abilities. Research Triangle Institute (RTI), www.rti.org

[24] The assessor used a book in order to determine the students' facility in handling printed material.

[25] The WEI Reading Specialist at the time made a decision to present letters in either block or cursive formats and lower and upper case format (one type on each side of a large plasticized card) as familiarity with the two formats was found to vary during field-testing of the instrument.

[26] To facilitate recognition, a large plasticized chart of 30 words of 1-3 syllables was presented to the student.

number of correct words read by minute (cwpm).

b. *Reading comprehension*: After the students finished the first passage, or the minute ended, the passage was removed. Students were orally asked four questions that required them to recall basic facts from the passage. The reading comprehension score was the number of correct answers with a maximum possible score of 4. When students were not able to read a minimum of 15 words of the first story, they were presented with a second story. The maximum total score for this subtask was 4.

Students selected to take the EGRA also responded to a brief, orally administered interview before they started the EGRA sub-tasks. The purpose of the interview was to gather information about the home and school contexts that might explain students' reading performance. For example, students were asked about the language they most speak at home with their family and with their friends.

3.2.2 School Management Assessment (SMA) Instrument

The original SMA used at Baseline was adapted from the instrument used in Cabo Delgado in 2011 by the project implemented by the Aga Khan Foundation. The instrument underwent extensive revisions in 2013 for the Midline 1 and again in 2014 for the Midline 2 administrations. The SMA focuses on school management practice and was designed to collect data (1) on indicators related to quantity of instruction (e.g. teacher and student attendance, start time of school shift, school director (SD) and pedagogical director (PD) attendance and (2) classroom teaching and learning processes, including instructional content, student-teacher interactions and availability and use of Teaching and Learning Aids (TLAs). The purpose of the SMA is to produce a multifaceted and comprehensive picture of school management routines and of the school-learning environment for both the IE and ApaL's internal performance monitoring requirements.

3.2.3 Interview Protocol for interviews with school directors

A semi-structured interview protocol was developed to gather the perception of school directors regarding the sustainability of the ApaL intervention and to explore the human, technical, and financial resources existing at the schools as well as the factors that facilitate or limit the continuation of the innovation without support from ApaL.

3.3 Instrument Administration

Each instrument required a different procedure for administration. The training of enumerators focused on preparing the enumerators to collect reliable data[27] and the supervisors to support the effort, advise the enumerators, clarify doubts and review all completed instruments to identify missing data or incorrect entries.

Training was conducted by ApaL senior staff with the participation of the IE Deputy Team Leader and the IE supervisors. Supervisor training was conducted on August 18-19 in Nampula and August 22-23 in Zambézia. Enumerator training was conducted in August 25-29. Over half of the enumerators trained had participated in both the training and in the prior data collection events (Baseline and Midlines). Data collection started on September 14 and was finalized by October 7. Besides participating in training, the IE supervisors made the rounds of the schools visited, observing the work of the different enumerator teams, calling attention to incorrect procedures, if any, and conducting interviews with school directors.

3.3.1 EGRA and Student Interview Administration

The EGRA administration started by randomly selecting among the grade 2 and 3 classes in the school—when there was more than one class—one second grade and one third grade class to participate in the

[27] *Validity of the instruments as well as their ease of administration had already been established prior to the training.*

assessment. After selecting the class, the enumerator spoke to the teacher and explained the purpose of the visit. Next, children were organized in rows and the enumerator would randomly select the ten students to be assessed. When there were ten or fewer students, all were selected for participation. The enumerator would then take the child to a quiet place to administer the EGRA. The detailed process to select the ten students to be assessed was the same utilized previously and is described in detail in the Baseline and Midline reports. Although data on student sex were collected, sex of the student played no role in the process of selecting classes or students for participation; i.e., selection of ten students per class to be assessed was done entirely randomly, based on the students who were actually present in the classroom on the day of assessment.

3.3.2 Teacher Interview

One of the enumerators administered face-to-face interview with the teacher whose class had been selected for EGRA administration. The interview included 33 items, which included questions covering teaching experience, pre- and in-service training, the use of local language to facilitate teaching, the use of class management tools, among others.

3.3.3 School Director Interview

This interview was conducted face-to-face with the school director, or when he/she was not present at the school, with the pedagogical director.[28] Some items required school directors to present proof of the answers they gave. For example, when asked whether they recorded teacher or student attendance or tardiness, they were asked to show the logs or forms used for this purpose.

3.3.4 Portuguese Class Observation

The focus on the observation was on the instructional behaviors exhibited by the teacher during one full class period. The enumerator used a structured observation protocol that listed reading instruction behaviors promoted by ApaL in the teacher training sessions. Instrument administration required that the enumerator arrive at the scheduled time for the class to start, record the time the class started and stay until the scheduled time for the class to end—about 45 minutes.

3.3.5 Classroom Inventory

The enumerator who conducted the classroom observation also filled out the classroom inventory. The instrument captured information to describe the classroom environment—seating patterns, materials posted on the walls, materials available to students, etc.

3.3.6 Interview with School/Pedagogical Directors

A semi-structured interview protocol was developed by the IE team and administered under the IE Deputy Team Leader's guidance by the IE supervisors. The interview protocol focused on the directors' assessment of activities that had been internalized to the point of being sustainable after ApaL ceased its involvement with the school.

3.4 Field Data Collection Procedures

In 2015, the same data collection procedures that were utilized at Baseline, Midline 1 and Midline 2 were used to ensure continuity. ApaL-selected and trained enumerators collected the EGRA and the SMA data at the same schools originally selected as a sample. In each school the EGRA was administered to ten randomly selected students in randomly selected second and third grade classes—one Grade 2 and one Grade 3 class per school.[29] In addition, the enumerators conducted a brief interview with each of

---

[28] Pedagogical directors are often also assistant directors. Their primary responsibility is to assist teachers with methodological issues and to fill in for school directors when they are absent. In many cases they are the deputy directors.

[29] When there was only one class in Grade 2 or Grade 3, then this class was automatically selected.

the children assessed and with the teacher of the selected class. Enumerators were selected and trained by ApaL while the IE trained its own supervisors to assist in ensuring the quality of the data collected. The IE supervisors were responsible for conducting in-depth interviews with school directors at the treatment schools.

3.5 Data Analysis

Different data required different methods of analysis as explained below.

3.5.1 Analysis of the EGRA reading scores

The IE model compares results obtained by students in the schools that were randomly assigned to treatments (Full and Medium) to those of a Control group. The 2015 results were analyzed by the application of the ANOVA model, and all pairs of groups were compared with the Tukey post-hoc statistic to provide an estimation of the significance between the group means. Unequal variances were assumed. Two-tailed Student t-Tests of significance of differences between means were employed for evaluating two-state independent variables such as 2014/2015, male/female, urban/rural, etc.

3.5.2 Analysis of the Class Observation Data

A preliminary analysis was provided to the IE team by the implementer and the IE team conducted further analyses. The steps followed by the IE are described here and the results are included in the Findings section.

The Class Observation instrument groups its 49 items into five sections:
1. Teacher-student interaction (12 items)
2. Teaching decoding (9 items)
3. Teaching comprehension (10 items)
4. Classroom management (10 items)
5. Teaching planning and sequence (8 items)

After examining the frequencies of positive responses to the individual items, it was found that many items in each category were scored very highly. Furthermore, multiple items had been included per section to describe a set of observable teacher behaviors and some items showed weak inter-item correlation or had little discriminatory power.[30] To solve these problems, an index or a composite score per section was created to summarize teacher performance within each section. This was done by first determining the relative "difficulty" of a positive behavior across all 319 observations.

Thus, a simple behavior demonstrated by most teachers received less weight in the composite score than more difficult positive behaviors observed among fewer teachers.[31] These weights were then applied to each item response to each teacher in each category. This allowed us both to compare means overall and to conduct inter-group comparisons in accordance with the IE model. The results obtained express the proportion or percentage of teachers who answered the item correctly weighted by the difficulty of the item. The weights derived in 2014 were used to determine the indices for 2015. The ANOVA and Tukey post-hoc contrasts were used to compare groups.

3.5.3 Analysis of the School/Pedagogical Director Interview Data

Given the sustainability focus, it became necessary to change the questions in order to determine whether processes and strategies implemented by ApaL remained in place once the project ceased. Directors were asked to name the strategies or the procedures that continued to be used at the school

---

[30] What is usually done is to calculate Chronbach's Alpha and delete items that correlate least with the other items in the test or in the block of items until an Alpha of an acceptable standard (usually 0.7 or above) is obtained.

[31] Item difficulty can range from 0.0 (none of the teachers answered the item correctly) to 1.0 (all of the teachers answered the item correctly).

even after the cessation of ApaL. They were asked to take into consideration the resources available to them and respond to questions such as *"How likely are you to use each strategy again?"* and to identify and compare the number of teachers who taught early graders in 2014 and the number that were actually doing so in 2015.

The interviews with 96 school/pedagogical directors were recorded with the permission of the interviewees and then transcribed to facilitate analysis. To analyze the data, we took the following steps:

1. Read the transcribed interviews and identify recurrent themes or the idea categories that emerged from the data;
2. Note patterns in the data by examining the content of each response in order to categorize verbal data for the purpose of classification, summarization and tabulation;
3. Code the data by attaching labels to the lines of text in order to group and compare similar or related pieces of information and then compile similar blocks of text from different sources into a single file; and,
4. Search for answers to the research question.

3.6 Limitations of the Study

As with most studies there are limitations that impact the generalizability of the findings. Some of these limitations are mentioned below:

- The selection of ZIPs, and consequently schools, along the economic corridors of the two provinces, as specified by USAID. It is reasonable to assume that schools along an economic corridor have some characteristics that are different from schools at less economically developed areas.
- The selection of a pool of ZIPs that was accessible to ApaL staff. Schools where travel conditions, etc. make it technically feasible to implement an intervention form a sub-set of the total universe of ZIPs. In fact, flooding that took place in in Nampula province prior to collection of the baseline data necessitated swapping some schools originally intended to be part of the sample for other, comparable schools in the same district.
- Only the teachers, students, SDs and PDs who were present on the day of the visit were assessed and/or interviewed. Although schools that were not in session when an initial visit was made received revisits, sometimes they were not in session on the revisits, either. Those who were absent may differ in important ways from those who were present. (However, the fact that students and teachers and administrators were absent is itself important to note.)
- With the exception of the EGRA and the Class Observation and Classroom Inventory, measures of implementation relied partly on staff self-report, rather than direct observation.
- The routinization of school directors' behaviors or their school management strategies, for example, were observed in 2015 only once, at the end of the school year.
- The field has yet to establish widely accepted time intervals after which an intervention can be called "sustainable" [32](Scheirer (2005), and it is likely that appropriate periods may vary from one intervention setting to another. The assessment timeframe in the current study corresponded to the academic calendar. The sustainability assessment, like the Midline 1 and Midline 2 assessments, occurred near the end of the school year.

[32] *Scheirer MA. Is sustainability possible? A review and commentary on empirical studies on program sustainability. American Journal of Evaluation. 2005; 26:320–347*

4. CHARACTERISTICS OF THE POPULATION

Section 4 describes the populations assessed and/or interviewed.

4.1 Description of Students Assessed by the EGRA and interviewed

Although in general, these characteristics did not turn out to be directly associated with learning outcomes, they nevertheless provide a useful picture of the context learners and teachers face daily, as well as confirmation that the control and intervention schools were not significantly different. As done in previous data collection events in 2013 and 2014, ten randomly selected Grade 2 (1,675) and ten randomly selected Grade 3 students (1,668) in randomly selected classrooms were interviewed and assessed using the EGRA tool.[33]

Table 4-1. Population and sample count of schools where data were collected in 2015

Province	Districts included	ZIPs	Number of schools	Grade 2 EGRA	Grade 3 EGRA
Nampula	Monapo	5	33	304	300
	Nampula Cidade	9	31	306	303
	Murrupula	4	21	198	187
	Rapale	1	7	63	54
Subtotal	4	19	92	871	844
Zambézia	Mocuba	7	38	360	373
	Nicoadala	7	41	394	401
	Quelimane	1	5	50	50
Subtotal	3	15	84	804	824
TOTAL	**7**	**34**	**176**	**1,675**	**1,668**

N=3,343

The students assessed and interviewed are roughly 5.5% of the more than 60,000 students enrolled in Grades 2 and 3 at these schools. Tables 4-2 and 4-3 present the breakouts of Grade 2, Grade 3 and total enrollments in the IE sample schools by student sex and treatment group. Percentages by sex within each breakout are calculated as well. A slightly greater number of girls enrolled versus boys has been consistent in these early grades since Baseline.

Table 4-2. School enrollment by grade and by sex

SCHOOL ENROLLMENT BY GRADE AND SEX								
Grade 2	**Full**		**Medium**		**Control**		**Total**	
	N	%	N	%	N	%	N	%
Girls	6,531	50.2%	6,429	51.4%	3,096	47.9%	16,056	50.2%
Boys	6,481	49.8%	6,082	48.6%	3,370	52.1%	15,933	49.8%
Total	13,012	100.0%	12,511	100.0%	6,466	100.0%	31,989	100.0%
Grade 3	**Full**		**Medium**		**Control**		**Total**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Girls	5,981	51.0%	5,940	52.3%	2,666	48.7%	14,587	51.1%
Boys	5,745	49.0%	5,425	47.7%	2,806	51.3%	13,976	48.9%
Total	11,726	100.0%	11,365	100.0%	5,472	100.0%	28,563	100.0%

The enrollment of girls in Full and Medium surpasses that of boys in both grades, while, in Control schools, boys surpass girls by about the same margins. The apparent decline in third grade enrollment

[33] When there was only one Grade 2 or Grade 3, then that classroom was automatically selected. Procedures for the selection of students are detailed in the Inception Report for work done under the Task Order.

can be due to many factors, such as the "bloating" of second grade enrollments due to repetition, as well as dropout.

Table 4-3. EGRA participants by grade and by sex

EGRA PARTICIPANTS BY GRADE AND SEX								
Grade 2	Full		Medium		Control		Total	
	N	%	N	%	N	%	N	%
Girls	288	51.0%	275	51.3%	309	53.8%	872	52.1%
Boys	277	49.0%	261	48.7%	265	46.2%	803	47.9%
Total	565	100.0%	536	100.0%	574	100.0%	1,675	100.0%
Grade 3	Full		Medium		Control		Total	
	N	%	N	%	N	%	N	%
Girls	287	50.9%	276	50.8%	297	52.9%	860	51.6%
Boys	277	49.1%	267	49.2%	264	47.1%	808	48.4%
Total	564	100.0%	543	100.0%	561	100.0%	1,668	100.0%

No procedure was put into place to ensure equal numbers of boys and girls, nor was it considered desirable to do so since the assessment was designed to measure the performance of students actually present. Since ten students were randomly selected in each of the randomly selected classes in the school, the composition of the sample reflects the general composition of the population—slightly higher number of girls in Grades 2 and 3 in treatment groups, slightly lower in the Control group.

Small differences between the enrollment and EGRA data are to be expected: enrollment data are from the beginning of the school year, while, as noted, EGRA data represents the random selection of students in the randomly selected second or third grade class at each school on the *day of the school visit*. Differences between the sexes or within-year dropout and/or attendance when the visit took place would not be surprising due to small variations between the two methods. The classroom observation and classroom inventory data from ApaL, reported on in the Findings section of this report, included classroom-specific enrollment and attendance data necessary to refine our conclusions in this regard and triangulate results with the data patterns we are seeing.

Table 4-4. Self-reported Age of students taking the EGRA

Self-reported age Grade 2	Grade 2		Grade 3		Self-reported age Grade 3
	Nampula	Zambézia	Nampula	Zambézia	
7 years old or younger	19.7%	24.8%	31.6%	16.6%	8 years old or younger
8 years old	23.0%	25.7%	17.5%	23.3%	9 years old
9 years old	14.0%	18.3%	21.0%	30.0%	10 years old
10 years old	20.9%	19.4%	12.5%	15.7%	11 years old
11 years or older	22.4%	11.7%	17.8%	14.6%	12 years or older
Average age Girls	9.1	8.8	10.4	9.9	Average age Girls
Average age Boys	9.4	8.8	10.1	10.0	Average age Boys

N=2,566

Age of student can be a factor related to performance. Note that age is self-reported since the school does not always have the information. Nampula Grade 2 students tend to be older than their counterparts in Zambézia, with nearly twice as many students in the 11 years or older category and somewhat fewer in the 7 years or younger category. In Grade 3, despite having similar mean ages, Nampula differs due to a greater number of younger students. Differential repetition and drop-out rates between the two provinces may be in play here.

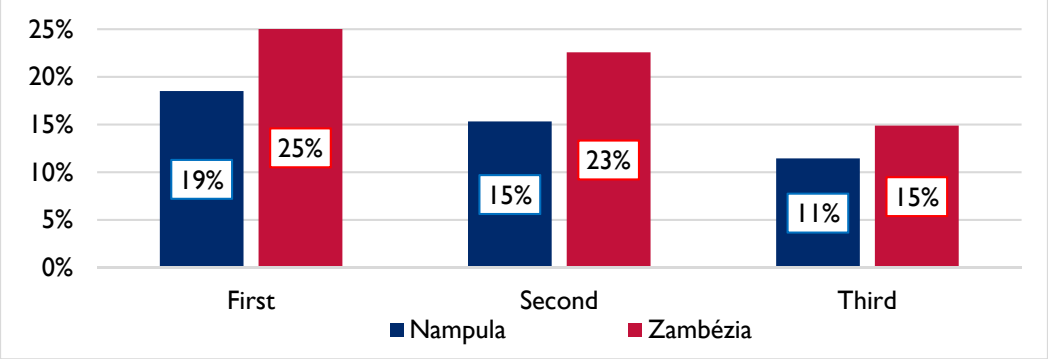
Table 4-5. Age of Students by Treatment Group

Student Age	Contrast with

Treatment	Valid N	Mean	Std. Error	Treatment	Mean Dif.	Std. Error	Tukey Sig
Full	298	10.00	0.098	Medium	-0.60	0.148	**0.000**
				Control	-0.46	0.137	**0.002**
Medium	294	10.60	0.112	Control	0.13	0.138	0.603 NS
Control	404	10.45	0.089	Overall ANOVA	F	Significance	
Total	996	10.36	0.057		9.13	**0.000**	

As noted before, mean scores are useful for evaluating overall differences between groups, although they do not tell us how the actual distribution of ages affects the differences in means. Nonetheless, the preceding table shows that the mean age was significantly associated with the IE group, and that Full treatments students were younger, on average, than either their Medium or Control counterparts, which were not statistically different between themselves.

Figure 4-1 Self-reported repetition by grade repeated, of 2015 second and third grade students



N= 3,330 for First Grade, 3,259 for Second Grade and 1,544 for Third Grade

Grade 2 students were asked if they were repeating grade 2 as well as whether they had repeated grade 1 in prior years. Third graders were asked about currently repeating Grade 3 as well as whether they had repeated Grades 1 and/or 2 earlier.

Even though the information displayed above is self-reported, repetition rates are high, especially in grade 1 and particularly in Zambézia. Holding students back has been proven to be costlier to the system than providing them with expanded opportunity to learn (more hours at school, tutoring, etc.). At the same time, repetition rates are often associated with drop-out rates. Students who repeat are more likely to drop out of school and pushing less mature or less able students out of school in the early grades only adds to the educational problems faced by a country.

The family situation of students interviewed is displayed in Table 4-6 and does not differ from what was observed in 2013 and 2014. Note that at each data collection event twenty different students (ten in each grade) were selected from randomly selected classrooms and results are fairly constant across the approximately 15,000 students assessed from Baseline to 2015. The slight variation in percentages is expected given that at each of the four data collection occasions a different sample of students was randomly drawn. In any case, the treatment and Control groups do not vary significantly from one another.

Table 4-6. Family situation as reported by students who took the EGRA

Interview Questions	Full	Medium	Control	Total
Does not live with mother	15.1%	14.3%	14.3%	14.6%
Does not live with father	26.1%	23.4%	24.9%	24.9%
Orphan of mother	2.1%	2.4%	1.9%	2.1%
Orphan of father	9.6%	6.8%	8.2%	8.2%

Orphan of both father and mother	2.1%	1.1%	2.6%	2.0%
Other reasons for not living with father, mother or both (separation, divorce, work in another area)	15.5%	15.8%	13.8%	15.0%

N=3,343

Of those responding with 'Other reasons,' 40% indicated that it was due to parents working or living outside of the community, 39% said it was because they lived with relatives, most often grandparents, and 14% stated that their parents were divorced or separated. Obviously, some of those who stated that they lived with relatives did so because of the parents living or working outside of their community, or because their parents were separated or divorced. These data were not designed to conduct a detailed study of the living conditions and causes of one or both of the parents of the child not being present in the household, but, given the relatively large percentage of children who indicated reasons other than the death or one or other of the parents, we have included a breakdown of the verbal responses given and recorded in the child's own words. The three treatment groups were quite similar in terms of response patterns. Zambézia reported higher "Other" responses for living or working outside of the community (49%) than Nampula (33%), while living with relatives (45%) and divorce or separation (15%) were higher in Nampula than in Zambézia (29% and 11%, respectively).

Information on student absenteeism was collected in grade 2 and 3 classes where the EGRA was administered. The number of students enrolled in the selected classes was compared to a count of the students present on the day of the visit to determine the absentee rate.

Table 4-7. Overall absentee rates for boys and girls in both provinces

Boys			**Girls**		
Enrolled	56	Absentee Rate	**Enrolled**	55	Absentee Rate
Present	22	61%	**Present**	24	56%

Native language and language spoken outside of the school are important factors that can facilitate or delay reading skills acquisition. The two tables that follow describe the out-of-school language situation reported by the students who took the EGRA.

Table 4-8. Student reported Portuguese use with parent

Parents	Full	Medium	Control
Almost always	40.7%	27.0%	27.1%
Occasionally	28.8%	23.7%	27.4%
Almost never	16.1%	24.6%	21.2%
Never	14.4%	24.7%	24.4%

N = 3,008; Chi-square= 94.95, df=6, p=0.000

Table 4-9. Student reported Portuguese use with siblings and friends

Siblings/Friends	Full	Medium	Control
Almost always	42.2%	28.1%	28.5%
Occasionally	28.8%	28.2%	29.3%
Almost never	15.1%	21.3%	20.2%
Never	13.4%	22.4%	22.0%

N = 3,343; Chi square=91.01, df=6, p=0.000

Note the unexpectedly high percentage of students reporting using Portuguese *almost always* and *occasionally* with parents and siblings or friends. The scores obtained by the students on the Oral Comprehension EGRA subtest confirm this information. Assessors did not report any cases of a child who could not take the EGRA for lack of understanding Portuguese.

4.2 Grade 2 and Grade 3 Teachers Interviewed

A structured interview protocol was administered to Grade 2 and Grade 3 teachers of classes where students took the EGRA. The tables that follow describe the characteristics of the 269 teachers interviewed. Note that the information reported refers only to those teachers whose classrooms were randomly selected to participate in the EGRA assessment—the IE sample.

The proportion of males and females ranged from 63.2% female in Medium to 69.7% female in Full treatment schools with non-significant differences. This is not surprising. The predominance of females is generally found in most primary schools. Full and Medium schools had fewer younger (<= 25 years old) than did the Control schools, and Full schools had a greater number of older (>= 41 years old) teachers than either Medium or Control schools.

Table 4-10. Age of teachers interviewed

Teacher Age	Full	Medium	Control
N reporting	99	86	84
25 or fewer years old	11.1%	12.6%	28.6%
26-30 years old	28.3%	40.2%	19.0%
31-40 years old	39.4%	34.5%	44.0%
41 or more years old	21.1%	12.6%	8.3%
Chi-square sig.	p = 0.001		

N=269

Reported years of experience as a teacher followed the same pattern as with teacher age. More experienced teachers (from six to 11 years of experience or more) are found in Full and Medium schools and make up for about one-half of the teacher contingent at those schools. Control schools have 60% of their teachers with limited experience (five or fewer years).

Table 4-11. Teacher Years of Experience

Teacher Years of Experience	Full	Medium	Control
N reporting	99	86	84
1 or fewer years	5.1%	4.6%	15.5%
2-5 years	41.4%	44.8%	44.0%
6-10 years	31.3%	41.4%	23.8%
11 or more years	22.2%	9.2%	16.7%
Chi-square sig.	p = 0.009		

As indicated in Table 4-12, Macua is the predominant language of the teachers interviewed followed by Chuabo. Native language of the teacher is not significantly different across treatment groups. A little over ten percent report Portuguese as their native language.

Table 4-12. Teacher native language

Teacher Native Language	Full	Medium	Control
N reporting	99	87	84
Portuguese	8.1%	12.6%	13.1%
Macua	44.4%	35.6%	39.3%
Chuabo	24.2%	33.3%	28.6%
Lomwe	9.1%	10.3%	10.7%
Other	14.1%	8.0%	8.3%
Chi-square sig.	p = 0.665 NS		

N=270

Since many of the students assessed report speaking a language other than Portuguese outside of the school, teachers were asked whether they used local languages to facilitate the teaching of Portuguese. Table 4-13 shows that using a local language is common practice across the three groups with no significant differences.

Table 4-13. Teacher use of local language to facilitate teaching of Portuguese

Uses Local Language to Facilitate Teaching of Portuguese?	Full	Medium	Control
N reporting	99	87	84
Yes	85.9%	89.7%	83.3%
No	14.1%	10.3%	16.7%
Chi-square sig.	p = 0.480 NS		

N=270

5. FINDINGS

In this section we present and discuss findings resulting from the analyses of the quantitative (EGRA) data and of the interviews with students who took the EGRA and their teachers. Class observations and interviews conducted with school directors are utilized to provide an answer to Research Question 4—the extent to which reading scores obtained on the EGRA in 2014 were sustained. The findings directly address the sustainability of the processes and strategies implemented by ApaL and the barriers and facilitators to the sustainability of the intervention.

We start by presenting EGRA scores obtained by second and third graders in 2015 by treatment group. Note that if no significant differences had been found between treatment and Control groups no further comparisons would have been necessary. This would imply that no traces of the intervention remained once ApaL ceased its intervention in the sample schools. The significance of the differences found between treatment and Control groups directed us to additional comparisons. Findings that refer to 2015 are presented and discussed at various levels—students, teachers, instructional behaviors observed in the classroom, classroom environment and materials.

Next we compare results obtained in 2015 to those obtained after one year of full implementation (Midline 2, September 2014). Both comparisons show the extent to which gains obtained in 2014 were sustained one full year after the cessation of the ApaL program in the schools originally selected as a sample. Sustained gains could imply that local education personnel (teachers, school and pedagogical directors) continued, to a certain extent, to use ApaL educational processes related to the gains observed in 2014 and that some technical or financial resources are available at the school for that

purpose. Information obtained by conducting further analyses on the data provided by the ApaL M&E—teacher instructional behavior and classroom environment—are integrated into the report. Information obtained by conducting in-depth interviews with 96 school and pedagogical directors are also reported to enrich and expand the findings.

Table 5-1 compares scores obtained by second and third graders in three different occasions: October 2013 after two months of the implementation of ApaL activities (Midline 1); September 2014, after one full year of intervention (Midline 2); and, September 2015 one year after the cessation of the intervention in the sample schools (Endline). The information displayed on Table 5-1 shows the evolution of scores from one year to the next in the different treatment groups.

Table 5-1. Reading Scores at Midline I, 2, and Endline

EGRA Subtests and Maximum Scores	Midline 1- 2013			Midline 2 – 2014			Endline - 2015		
	Full	Med.	Cont.	Full	Med.	Cont.	Full	Med.	Cont.
Grade 2									
Letter Recognition (100)	9.5	7.4	4.7	19.9	17.2	5.4	10.7	7.2	5.9
Familiar Words (30)	1.9	1.0	0.8	3.3	2.6	1.1	2.4	1.5	1.3
Reading Fluency (120)	2.2	1.0	1.0	5.2	4.2	1.7	3.4	2.4	2.0
Reading Comprehension (4)	.07	.02	.02	.21	.12	.03	.25	.22	.14
Grade 3									
Letter Recognition (100)	16.6	15.3	12.0	29.6	27.8	12.3	20.1	19.2	12.6
Familiar Words (30)	3.9	3.3	2.8	8.0	6.0	3.2	5.7	4.8	3.3
Reading Fluency (120)	5.3	4.4	4.3	14.6	12.0	5.2	8.8	7.4	5.0
Reading Comprehension (4)	.16	.12	.12	.53	.43	.15	.50	.36	.30

ANOVA between groups on Final, Second Grade: On all subtests: p=0.000 on all except Reading Fluency (0.001)
ANOVA between groups on Final, Third Grade: On all subtests: p=0.000

As expected, reading scores obtained by second and third graders are lower one year after the program ceased than they were while ApaL was in full implementation mode at the school. However, when compared to Midline 1 – 2013, especially for third graders and especially for the Full treatment schools some differences remain. As often mentioned in previous reports, the scores obtained by students in the Control group show what would have happened without the ApaL project—all scores would be similar to those observed in the Control group.

Another point to be made is the magnitude of the effects observed in the two treatment groups with students in the Full treatment group consistently outperforming their counterparts in the Medium treatment group, even one year after the cessation of ApaL at the school. This confirms the results of the 2014 cost-effectiveness study, which documents the higher gains for the costs incurred by implementing the Full versus the Medium treatment.

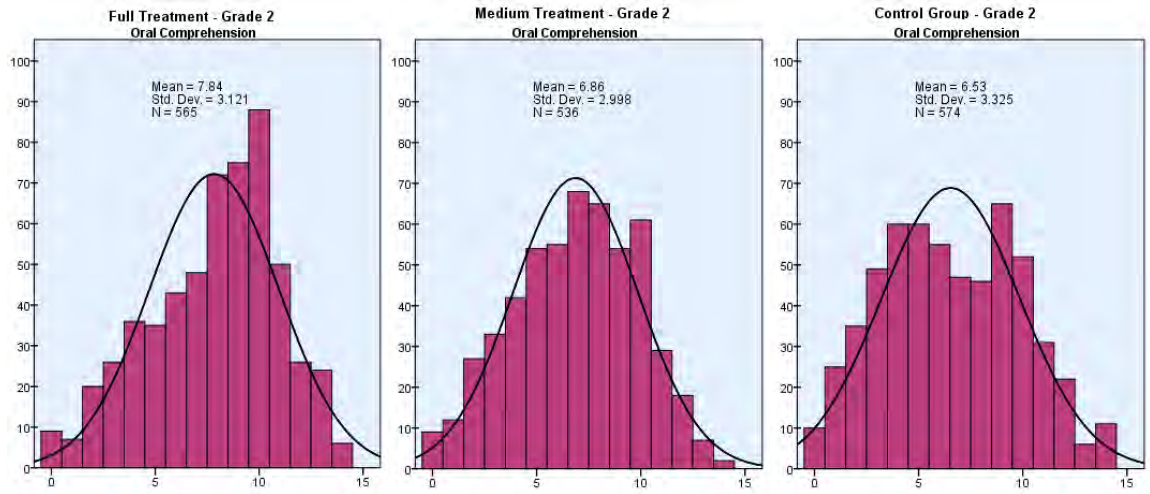
5.1 2015 EGRA results by treatment group and grade

As always, the IE model compares results obtained by students in the schools that were randomly assigned to treatment groups (Full and Medium) or a Control group. The 2015 results were analyzed by the application of the ANOVA model, and all pairs of groups were compared with the Tukey post-hoc statistic to provide an estimation of the significance between the group means. Unequal variances were assumed. A probability of <= 0.05 was used to determine statistical significance.

In cases where a great number of students score zero or close to zero on a number of subtests—as is the case here—the mean is not always the best statistic to describe a situation. The histograms presented here accurately depict how groups of students—in Full, Medium, and Control schools—perform on the EGRA by showing the frequency with which particular scores occur. The histograms also include those students who were not able to understand the questions posed to them and scored

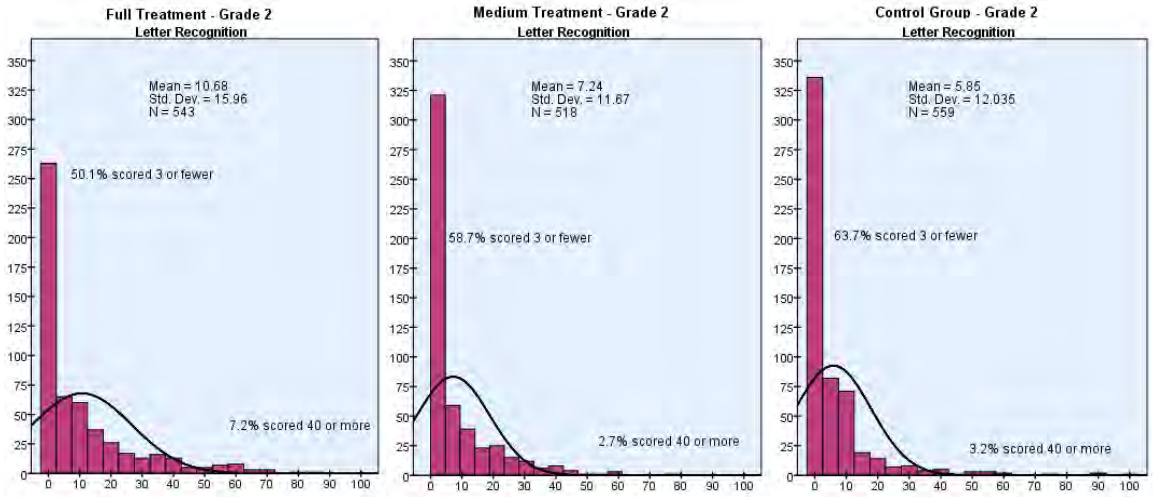
zero. The graphs below show the distribution of scores for two EGRA subtests—Oral Comprehension and Letter Recognition.[34]

Figure 5-1 Distribution of Oral Comprehension scores for Grade 2 students by treatment group



In a histogram the concept of area to represent frequency of EGRA scores is used on the Oral Comprehension subtest. The first bar shows us the number of students who scored zero, probably due to their limited ability to understand spoken Portuguese. Figure 5-1 shows that even in Grade 2 students understanding of Portuguese is good (highest score possible 14; means = 7.84, 6.86, and 6.53 for Full, Medium and Control groups respectively). Letter Recognition is an essential skill for students to master, the first decisive step, if we expect them to be able to learn words in Grade 3. The second set of histograms corresponds to the distribution of scores for the subtest Letter Recognition. The means reported on Figure 5-2 (10.68, 7.24, and 5.85, maximum score possible was 100) for the Full, Medium and Control groups respectively) hide the fact that over 50% of all Grade 2 students are only able to read 3 or fewer letters per minute, and only 7.2%, 2.7% and 3.2%, respectively, could read 40 or more letters per minute.

Figure 5-2 Distribution of Letter Recognition scores for Grade 2 students by treatment group



[34] *The histograms for all subtests for the two grades can be found in Annex G*

Table 5-2 below details what was presented on the graphs above. In 2015 the differences between groups in scores obtained on all subtests by second graders (except Reading Comprehension) are highly significant. Full surpasses both Medium and Control schools (p=0.000) on all subtests (exception: Reading Comprehension in the Medium group). Medium scores no differently from Control schools, with the exception, again, of Reading Comprehension.

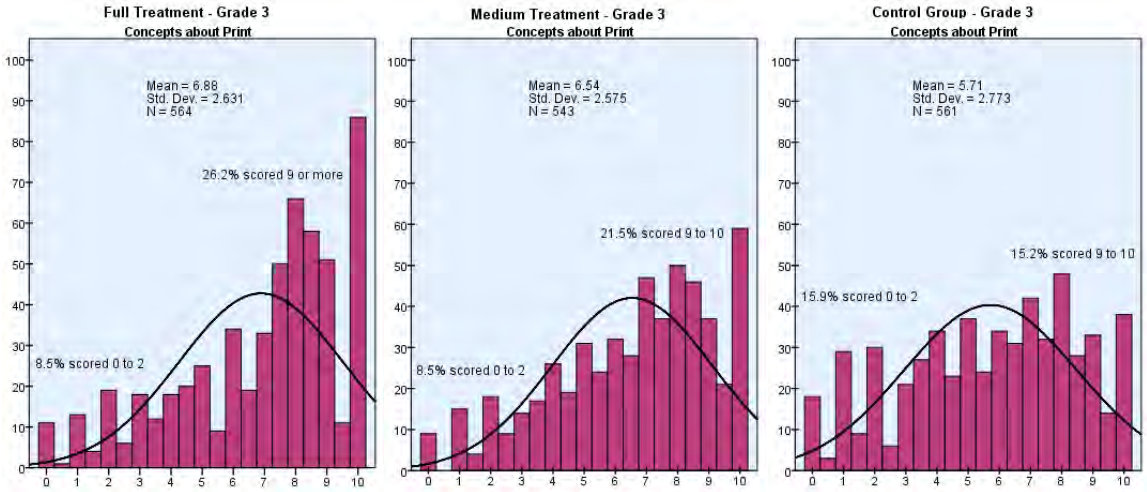
Table 5-2. EGRA Scores obtained by treatment and Control groups in 2015

Second Grade Mean Scores and Contrast Significance	Treatment Group			ANOVA Sig. (p =)	Contrast Full with		Contrast Medium with
	Full	Medium	Control		Medium	Control	Control
Oral Comprehension	7.8	6.9	6.5	**0.000**	**0.000**	**0.000**	0.192 NS
Concepts about Print	5.5	4.5	4.2	**0.000**	**0.000**	**0.000**	0.178 NS
Letter Recognition	10.7	7.2	5.9	**0.000**	**0.000**	**0.000**	0.205 NS
Familiar Words	2.4	1.5	1.3	**0.000**	**0.000**	**0.000**	0.782 NS
Reading Fluency	3.4	2.4	2.0	**0.001**	**0.023**	**0.001**	0.578 NS
Text Comprehension	0.25	0.22	0.14	**0.000**	0.551 NS	**0.000**	**0.014**

N=3,433

The histograms that are presented below show the distribution of scores obtained by third graders on two EGRA subtests: Concepts about Print (CAP) and Familiar Words (cwpm) correctly read per minute. The maximum score possible is 10 for CAP and 30 for cwpm. Given the distribution of scores shown below for CAP, a good proportion of students have acquired some degree of mastery regarding the concepts of what printed materials are and what reading entails—they know where they have to start reading, how to follow a line, they recognize the cover and title of a book. Full and Medium schools surpass Control schools in this regard. However, there are a number of students - especially in the Control group—who have not mastered these concepts as shown by the percentage in each group who scored zero to 2. These contrasts show us that the change in the *distribution* of scores—and not just the *mean*—is helpful in seeing where the ApaL intervention affects results. Is the intervention improving the skills of those who already have some mastery? Or is it reducing the number of students with little or no mastery? Or, as is the case here, some of one and some of the other? While each class is different, this information can help to sensitize teachers to the students in their classes that may be in greater need of attention.

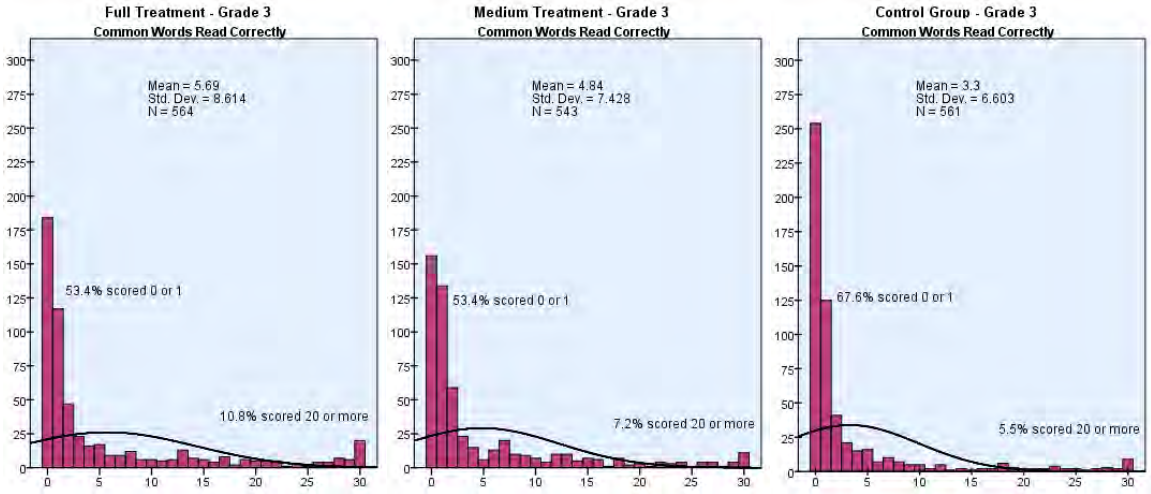
Figure 5-3 Distribution of Concept about Print scores for Grade 3 students by treatment group



The second set of graphs refers to the ability to read correctly, in one minute, words from a list of 30 familiar words. Note that at the end of Grade 3 only 10.8% in Full, 7.2% in Medium, and 5.5% in Control were able to read 20 or more words and that even in treatment groups more than half of third graders can only read zero or one word in one minute. Furthermore, despite having attended school for a

minimum of three years, over 150 children in the treatment groups could not read a single word while in the Control group more than 250 students scored zero.

Figure 5-4 Distribution of Words Correctly Read per minute for Grade 3 students by treatment group



The information depicted above is presented in more detail on Table 5-3. In Grade 3, on all EGRA subtests in both grades, significant differences were associated with the three groups—the Full intervention group out-performed the Control schools on all subtasks in both grades. Medium schools generally did not perform better than Control schools in Grade 2, but did show better results relative to the Control schools in Third Grade on four of the six EGRA subtasks. At the same time, Grade 3 Medium schools were much more similar to the Full schools.

Table 5-3. EGRA Scores obtained by Grade 3 students in treatment and Control groups in 2015

Third Grade Mean Scores and Contrast Significance	Treatment Group			ANOVA Sig. (p =)	Contrast Full with		Contrast Medium with
	Full	Medium	Control		Medium	Control	Control
Oral Comprehension	8.5	8.1	7.7	**0.000**	**0.049**	**0.000**	0.137 NS
Concepts about Print	6.8	6.4	5.6	**0.000**	0.069 NS	**0.000**	**0.000**
Letter Recognition	20.1	19.2	12.6	**0.000**	0.723 NS	**0.000**	**0.000**
Familiar Words	5.7	4.8	3.3	**0.000**	0.150 NS	**0.000**	**0.002**
Reading Fluency	8.8	7.4	5.0	**0.000**	0.192 NS	**0.000**	**0.000**
Text Comprehension	0.50	0.36	0.30	**0.000**	**0.000**	**0.000**	0.299 NS

In spite of the statistical significance reported, scores for pre-reading and actual reading EGRA subtests remain extremely low in both grades. Note that most of the Grade 3 treatment group students had been exposed to the ApaL project in 2014, plus any residual additional impact gained during 2015. In the case of Full and Medium schools, most Grade 2 students were only exposed to the residual effects of changes in teaching practices and surviving Teaching and Learning Aids (TLAs) and other materials.

The relationships observed between all the pre-reading and reading skills measured by the EGRA for Grade 3 for Full and Medium treatment combined for Grade 3 are displayed in Table 5-4. The information presented allows us to conclude that:

1. The pre-reading skills (Oral Comprehension and Concepts about Print) are generally not strong predictors of reading skills;

2. Letter recognition is a strong predictor of Familiar Word Reading (69%), of Text Reading Fluency (59%) and contributes somewhat (30%) to Reading Comprehension;

3. Familiar Word reading is an excellent predictor of Text Reading Fluency (70%) and strongly associated with Reading Comprehension (43%);

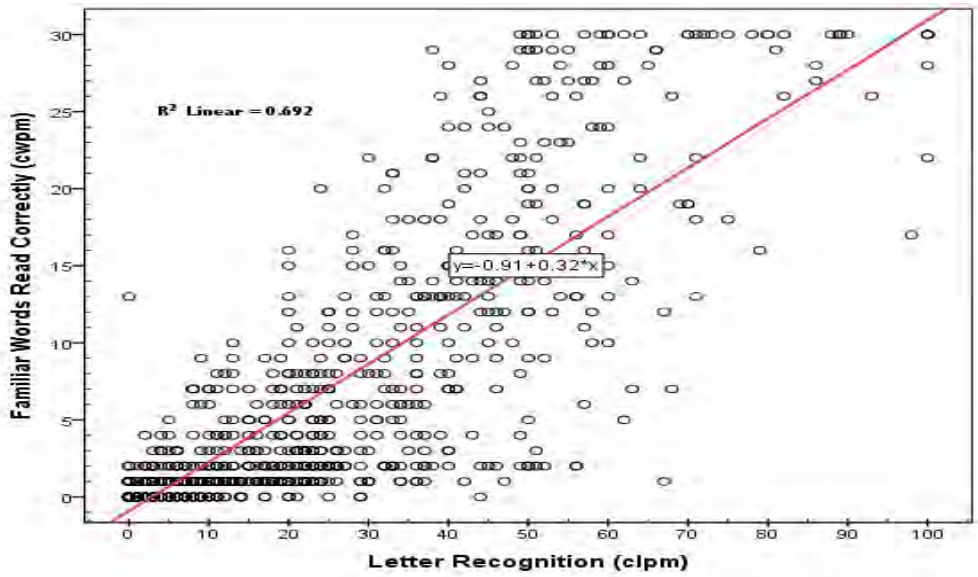
4. Text Reading Fluency predicts fully 49% of Reading Comprehension scores.

Table 5-4. Correlations between Subtests and Percentage of Variance Explained

Grade 3	Concepts about Print	Letter Recognition	Familiar Word Reading	Text Reading Fluency	Reading Compre-hension
Oral Comprehension	**r = 0.58** (R^2 = 34%)	**r = 0.38** (R^2 = 44%)	**r = 0.38** (R^2 = 14%)	**r = 0.34** (R^2 = 11%)	**r = 0.45** (R^2 = 20%)
Concepts about Print		**r = 0.54** (R^2 = 29%)	**r = 0.49** (R^2 = 24%)	**r = 0.43** (R^2 = 18%)	**r = 0.45** (R^2 = 20%)
Letter Recognition			**r = 0.83** (R^2 = 69%)	**r = 0.77** (R^2 = 59%)	**r = 0.55** (R^2 = 30%)
Familiar Word Reading				**r = 0.89** (R^2 = 70%)	**r = 0.65** (R^2 = 43%)
Text Reading Fluency					**r = 0.70** (R^2 = 49%)

The importance of learning letters and words in order to achieve the fluency necessary to read with comprehension—the ultimate goal of reading—cannot be overstated and must be at the core of any reading program. Figure 5-5 confirms the strong relationship between Letter Recognition and Familiar Word Reading as well as the dispersion of values at the higher end.

Figure 5-5 Relationship between Letter Recognition and Familiar Words Reading – Grade 3



As shown, students who are able to recognize correctly 40-45 or more letters per minute have higher scores on Familiar Word Reading. The findings presented above are summarized below.

Summary

- *One year following the end of ApaL intervention, on all EGRA subtests in both grades, significant differences were associated with the three group**s***—the Full intervention group out-performed the Control schools on all subtasks in both grades. Medium schools generally did not perform better than Control schools in Grade 2, but did show better results relative to the Control schools in Grade 3 on four of the six EGRA subtasks. At the same time, Grade 3 Medium schools were much more similar to the Full schools.
- *Scores for advanced pre-reading and actual reading skills EGRA subtests remain shockingly low in both grades.* Remember that most of the Grade 3 students had been exposed to the ApaL project in 2014, plus any residual additional impact gained during 2015. In the case of Full and Medium

schools, most Grade 2 students were only exposed to the residual effects of changes in teaching practices and surviving Teaching and Learning Aids (TLAs) and other materials.

5.1.1 Comparing 2014 with 2015 EGRA scores, by treatment group and grade

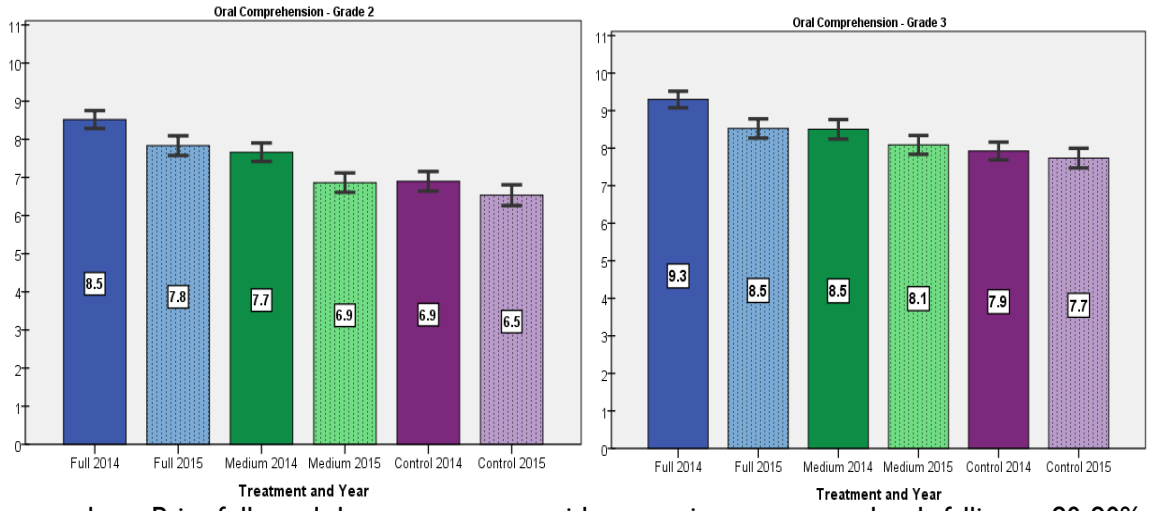
The tables that follow show results of the comparisons made between scores obtained in 2014 and in 2015 for both grades on all EGRA subtests. This allows us to estimate the degree to which project effects observed in 2014, due to a mix of inputs focused on directors, teachers, the classroom environment and TLAs and other materials, still can be detected a year after these inputs had ceased. We are interested to see how well the Full and Medium interventions sustained advantages in scores for each of the EGRA subtests. Some gains achieved in 2014 may be easier to sustain than others. Later in this report, we look in detail at the sustainability of the inputs themselves. At the end of 2015 the Oral Comprehension scores obtained by students in both treatment groups were about 90% of their 2014 levels, a drop that is statistically significant. As expected, Control group schools did not change between years.

Table 5-5. Oral Comprehension Scores for Grades 2 and 3 by Year and by Treatment Group

Oral Comprehension	Grade 2				Grade 3			
	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	8.5	7.8	92%	**0.000**	9.3	8.5	92%	**0.000**
Medium	7.7	6.9	90%	**0.000**	8.5	5.4	64%	**0.025**
Control	6.9	6.5	95%	0.057 NS	7.9	7.7	98%	0.291 NS

Figure 5-6 provides a visual representation of the differences observed between 2014 and 2015 by treatment group for both grades. The line bars show the 95% confidence interval of the estimated mean—when they overlap, difference are significant. Note that the differences are significant for both grades for the Full and Medium treatment groups, while for the Control group inter-year comparisons are not.

Figure 5-6 Oral Comprehension scores for Grades 2 and 3 by treatment group and year



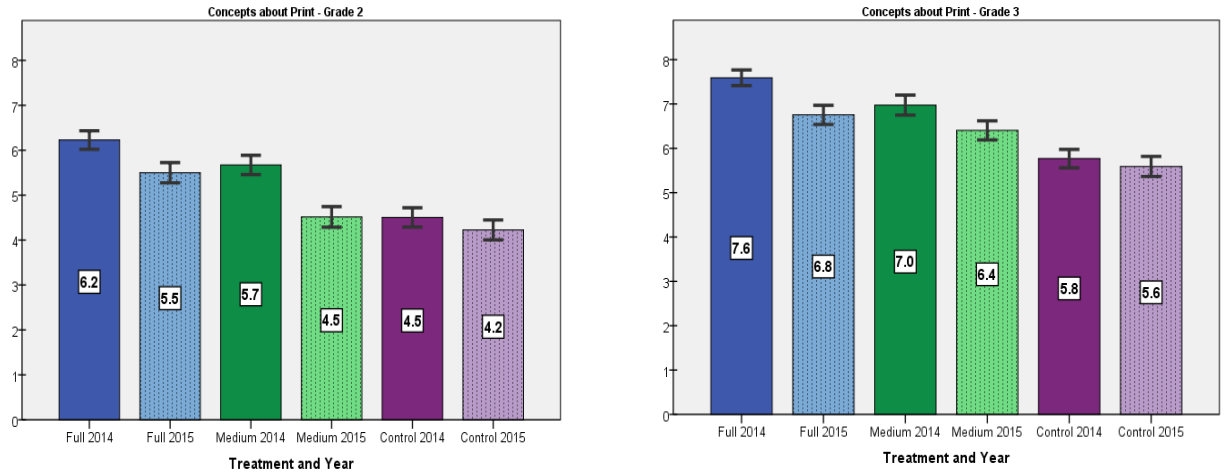
Concepts about Print followed the same pattern, with scores in treatment schools falling to 80-90% of 2014 levels.

Table 5-6. Concepts about Print Scores for Grades 2 and 3 by Year and by Treatment Group

Concepts about Print	Grade 2				Grade 3			
	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	6.2	5.5	88%	**0.000**	7.6	6.8	89%	**0.000**
Medium	5.7	4.5	80%	**0.000**	7.0	6.4	92%	**0.000**
Control	4.5	4.2	94%	0.078 NS	5.8	5.6	97%	0.259 NS

Scores on the EGRA subtest Concepts about Print, an emerging, pre-reading skill, are generally around 50%—maximum score is 10—for all groups, and the gains achieved in the treatment groups were sustained by about 80-90% in 2015, although this drop is significant in both cases. Control sample schools remained at their 2014 levels. Figure 5-7 visually compares scores obtained in 2014 and 2015 by students in the three groups.

Figure 5-7 Concepts about Print scores for Grades 2 and 3 by treatment group and year



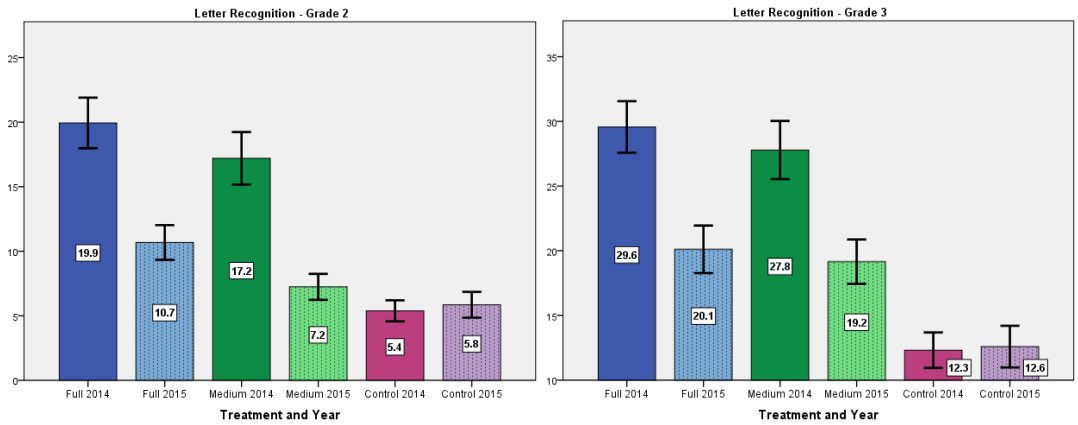
Letter Recognition scores were much lower in 2015 for both treatment groups. In Grade 2, Full schools were 54% of the 2014 level, Medium schools 42%, while Control schools remained essentially the same as in 2014. In third grade, the two treatment groups dropped to 67-69% of 2014 and Control showed no change. The maximum score on this subtest was 100. Note that in 2014 second graders in treatment schools were part of the program while second graders in 2015 are new students, never exposed to the ApaL intervention unless ApaL trained teachers remained at the school teaching Grade 2, or TLAs and other classroom materials remained. The lower scores obtained on this subtest may imply that either teachers went back to their old practices or were new to the former treatment school and had not received any training provided by ApaL. Even despite this drop-off, the difference in Letter Recognition scores between treatment groups and control is significant in both technical terms and in common understanding terms and is likely to have particular pedagogical significance given announced plans to transition to mother-tongue instruction. Table 5-7 displays information on Letter Recognition scores per grade, year, and treatment group.

Table 5-7. Letter Recognition Scores for Grades 2 and 3 by Year and by Treatment Group

	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	19.9	10.7	54%	**0.000**	29.6	20.1	68%	**0.000**
Medium	17.2	7.2	42%	**0.000**	27.8	19.2	69%	**0.000**
Control	5.4	5.9	109%	**0.000**	12.3	12.6	102%	**0.000**

Figure 5-8 visually compares scores obtained in 2014 and 2015 by students in the three groups.

Figure 5-8 Letter Recognition scores for Grades 2 and 3 by treatment group and year

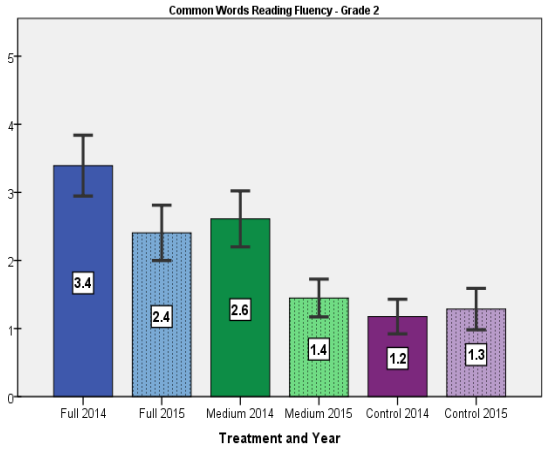


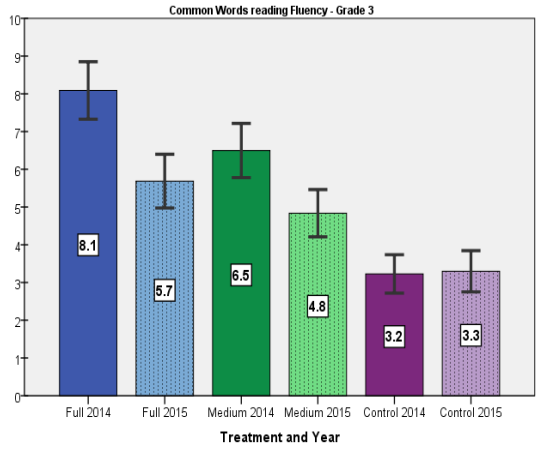
Scores on the Familiar Words read correctly subtest fell to 71% of the 2014 level in Full and 55% in Medium schools in Grade 2. In Grade 3, the decline was 68-69% in both treatment groups. Control schools did not change, but continue to lag behind Full schools. Medium school in both grades drifted down to Control school levels of performance. Table 5-8 details this pattern graphically.

Table 5-8. Scores on familiar words read correctly

Familiar Words Read Correctly (wcpm)	Grade 2				Grade 3			
	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	3.4	2.4	71%	**0.001**	8.1	5.7	70%	**0.000**
Medium	2.6	1.5	55%	**0.000**	6.5	4.8	74%	**0.000**
Control	1.2	1.3	109%	0.586 NS	3.2	3.3	102%	0.854 NS

Figure 5-9 Familiar Words correctly read scores for Grades 2 and 3 by treatment group and year



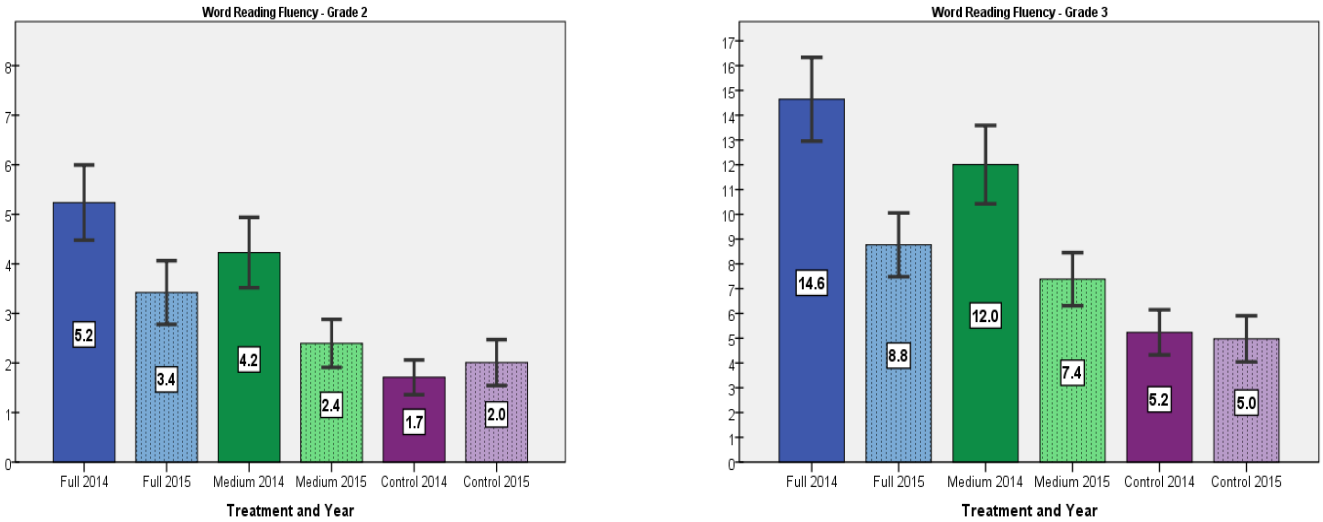


In 2015, scores obtained on the Text Reading Fluency by students in Full and Medium treatment schools averaged around 60% of 2014 levels in both grades, a significant decline between years. Control schools, again, did not change. In second grade, Medium school performance in 2015 declined to a level close to that of the Control schools. In third grade, Full and Medium schools were similar, and continue to out-pace the Control schools by a substantial margin. Figure 5-9 and Figures 5-10 highlight these patterns.

Table 5-9. Scores on Text Reading Fluency

Text Reading Fluency (wcpm)	Grade 2				Grade 3			
	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	5.2	3.4	65%	**0.000**	14.6	8.8	60%	**0.000**
Medium	4.2	2.4	57%	**0.000**	12.0	7.4	61%	**0.000**
Control	1.7	2.0	118%	0.314 NS	5.2	5.0	95%	0.692 NS

Figure 5-10 Fluency scores for Grades 2 and 3 by Treatment Group and Year



Reading Comprehension scores continue to be extremely low, as reading fluency remains a serious limitation for almost all students. In 2015, Full treatment schools were generally the same as in 2014, although Medium schools in Grade 2 improved to match the Full schools.

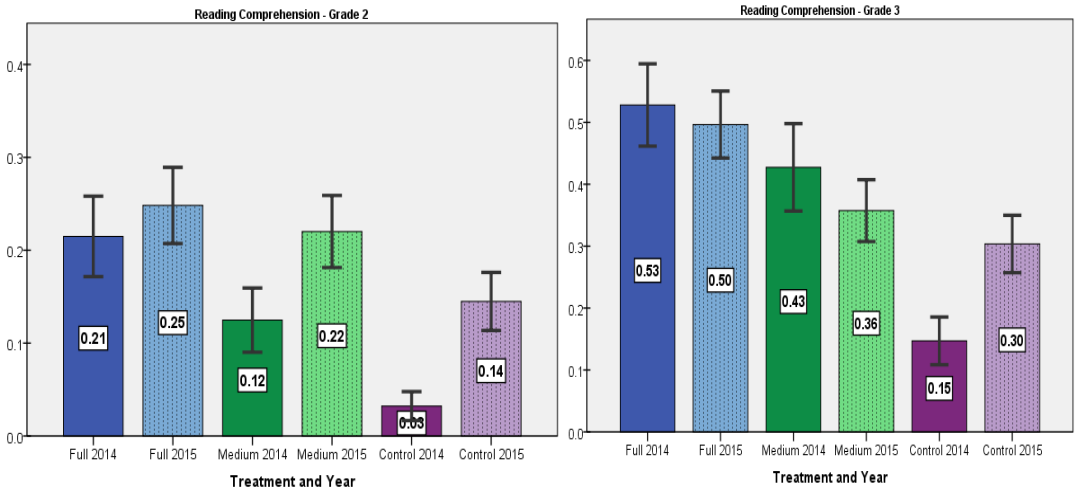
Control scores still lagged behind the treatment schools, but in both grades showed significant gains between 2014 and 2015. Finally, scores on the subtest Reading Comprehension are displayed on Table 5-10 and Figure 5-11.

Table 5-10. Scores on Reading Comprehension

Text Reading Comprehension	Grade 2				Grade 3			
	2014	2015	2015 as a % of 2014	Sig. (p =)	2014	2015	2015 as a % of 2014	Sig. (p =)
Full	0.21	0.25	119%	0.273 NS	0.53	0.50	94%	0.469 NS
Medium	0.12	0.22	183%	**0.000**	0.43	0.36	84%	0.113 NS
Control	0.03	0.14	467%	**0.000**	0.15	0.30	200%	**0.000**

Note: NS indicates that the differences between groups are not significant at the p less than or equal to 0.05 level

Figure 5-11. Reading Comprehension scores for Grades 2 and 3 by Treatment Group and Year



The findings presented above are summarized below.

Summary

- *Oral Comprehension in 2015 in treatment schools was about 90% of 2014 level.* Control schools did not change between years. (This finding for Control schools is to be expected, of course, and goes to confirm the validity of the overall sample.)
- *Concepts about Print followed the same pattern*, with scores in treatment schools falling to 80-90% of 2014 levels.
- *Letter Recognition was much lower in 2015 for both treatment groups.* In Grade 2, Full schools were 54% of the 2014 level, Medium schools were at 42%, while Control schools remained essentially the same as in 2014. In third grade, the two treatment groups dropped to 67-69% of 2014 and Control showed no change. Despite the drop-off, the difference between Letter Recognition in Treatment Schools vis-à-vis Control schools is significant in both technical terms and in practical terms, especially given announced plans to transition to instruction in First Language.
- *Familiar Word Reading fell to 71% of the 2014 level in Full and 55% in Medium treatment schools.* Control schools did not change, but continue to lag behind Full schools. Medium school in both grades drifted down to Control school levels of performance.
- *Text Reading Fluency of 2015 Full and Medium schools averaged around 60% of 2014 levels.* Control schools did not change. In second grade, Medium school performance in 2015 declined to a level close to that of the Control schools.
- *Text Reading Comprehension scores continue to be extremely low, as reading fluency remains a serious limitation for almost all students*. In 2015, treatment schools were generally the same as in 2014, although Medium schools in second grade improved to match the Full schools. Control scores still lagged behind the treatment schools, but in both grades showed significant gains between 2014 and 2015.

5.2 Differences in performance of boys and girls

The effects of sex on student performance were examined in relationship to performance on the six EGRA subtests in 2015. Table 5-11 displays the results.

Table 5-11. Differences in Performance of Girls and Boys by Treatment Group and Grade

EGRA Subtest	Student Sex	SECOND GRADE			THIRD GRADE		
		Means by Treatment Group			Means by Treatment Group		
		Full	Medium	Control	Full	Medium	Control
Oral Comprehension	Girls	7.6	6.8	6.5	8.3	7.8	7.3
	Boys	8.1	6.9	6.6	8.8	8.4	8.2
	Sig. t-test	0.053 NS*	0.674 NS	0.645 NS	0.055 NS	**0.037**	**0.000**
	Girls/Boys	93.8%	98.4%	98.2%	94.3%	93.7%	88.6%
Concepts about Print	Girls	5.2	4.4	4.1	6.5	6.3	5.1
	Boys	5.8	4.7	4.4	7.0	6.7	6.2
	Sig. t-test	**0.003**	0.196 NS	0.230 NS	**0.049**	0.197 NS	**0.000**
	Girls/Boys	88.2%	93.5%	93.8%	93.8%	94.1%	81.5%
Letter Recognition (clpm)	Girls	10.0	6.1	4.5	17.6	16.7	9.9
	Boys	11.4	8.4	7.4	22.7	21.7	15.7
	Sig. t-test	0.305 NS	**0.030**	**0.005**	**0.007**	**0.004**	**0.000**
	Girls/Boys	87.7%	73.3%	60.8%	77.7%	76.8%	62.8%
Familiar Words Read Correctly (cwpm)	Girls	2.2	1.1	1.0	4.8	4.4	2.5
	Boys	2.7	1.8	1.6	6.6	5.3	4.2
	Sig. t-test	0.241 NS	**0.015**	0.066 NS	**0.011**	0.128 NS	**0.003**
	Girls/Boys	81.9%	61.7%	63.8%	71.9%	81.8%	59.7%
Text Reading Fluency (wpm)	Girls	3.0	2.0	1.8	7.4	6.3	3.8
	Boys	3.8	2.8	2.3	10.3	8.5	6.3
	Sig. t-test	0.214 NS	0.137 NS	0.340 NS	**0.028**	**0.043**	**0.010**
	Girls/Boys	78.6%	73.6%	80.0%	71.7%	74.0%	60.7%
Reading Comprehension	Girls	0.21	0.20	0.15	0.44	0.29	0.23
	Boys	0.29	0.24	0.14	0.55	0.43	0.39
	Sig. t-test	0.073 NS	0.392 NS	0.957 NS	0.053 NS*	**0.004**	**0.001**
	Girls/Boys	72.4%	83.3%	107.1%	80.0%	67.4%	59.0%

Note: NS signifies that significance level is "borderline" as we use less than or equal to p=05 to denote significance.*

In no case in either grade do girls EGRA performance surpass that of boys. Girls typically achieve results most similar to boys on Oral Comprehension and Concepts about Print. Second graders in Full treatment schools show significant difference between girls and boys on only one out of six subtests (Concepts about Print) as well as the generally higher scores that we have seen in the preceding tables. Second graders in Medium schools had significant sex differences on two out of six subtests (Letter Recognition and Familiar Words), while in Control schools only Letter Recognition differed significantly by student sex. While the common pattern is for girls to lag behind their male counterparts in Grade 2 by 20% or 30%, there is great variability between individuals, not just between the sexes. This makes comparisons less likely to yield statistically significant differences.

By the end of Grade 3, the picture has changed considerably. In Control schools, the results of all six of the EGRA subtests significantly favor boys (by up to 40%), and the performance gap has increased even more so with the overall lower performance of Control schools. Medium schools now have significant differences in 4/6 and almost 5/6 subtests, moving closer to Control schools in terms of sex differences and, in some cases, overall performance. Full schools show significant differences favoring boys on Concepts about Print, Letter Recognition, Familiar Words, Reading Fluency and borderline significant differences on Reading Comprehension. Main findings are summarized below.

Summary

- *Although gender differences in performance on the EGRA subtests continue to be widespread, girls as well as boys benefit from the learning gains that the Full and Medium interventions have provided.* The increased magnitude of sex differences in Grade 3 suggests that greater efforts will be required to address the root causes of under-performance of girls. The available data and other sources of information do not allow us to pursue possible causes for the differences at this point.

5.3 Difference in Performance for Students in Urban and Rural Schools

On every single comparison of results from the EGRA, urban school students strongly out-performed their counterparts in rural schools. They have higher EGRA scores for Oral Comprehension, Concepts about Print, Letter Recognition, Familiar Words, Text Fluency and Reading Comprehension. Tables 5-12 and 5-13 contrast scores obtained by students in rural and urban second and third grade classrooms.

Table 5-12. Comparison of 2015 Grade 2 EGRA Mean Scores by Urban/Rural Locality

EGRA Subtest – Grade 2	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Oral Comprehension	8.6	6.6	11.62	728.30	**0.000**
Concepts about Print	5.7	4.4	8.60	766.62	**0.000**
Letter Recognition (lpm)	11.7	6.5	6.20	543.19	**0.000**
Common Words Read Correctly (wpm)	2.6	1.4	4.50	611.06	**0.000**
Text Reading Fluency (wpm)	3.7	2.2	3.68	679.72	**0.000**
Text Reading Comprehension	0.31	0.17	5.14	648.18	**0.000**

In grade 2, Urban schools out-perform their Rural counterparts by large margins on every EGRA subtest, sometimes (Letter Recognition) more than doubling Rural performance. The same pattern holds in grade 3, where if anything, the urban-rural gap increases, as shown in the following table:

Table 5-13. Comparison of 2015 Grade 3 EGRA Mean Scores by Urban/Rural Locality

EGRA Subtest - Grade 3	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Oral Comprehension	9.7	7.6	13.83	826.67	**0.000**
Concepts about Print	7.4	5.8	11.83	826.88	**0.000**
Letter Recognition (lpm)	25.2	14.4	8.43	627.34	**0.000**
Common Words Read Correctly (wpm)	7.2	3.7	7.05	588.31	**0.000**
Text Reading Fluency (wpm)	10.8	5.7	5.93	596.23	**0.000**
Text Reading Comprehension	0.58	0.32	7.30	652.52	**0.000**

Given these large differences, we also looked at several other factors perhaps related to the urban-rural effects observed.

Students in the Urban classrooms are younger than their counterparts in Rural classrooms (which implies a lower proportion of over-age students and/or of students who are repeating a grade), have more class days offered to them and attend more days of class in a given month.

Table 5-14. Intermediate variables that may affect EGRA scores by Urban/Rural Locality

Other Variables	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Student Age	8.6	9.2	-5.25	782.02	**0.000**
Student Days Attendance in July	11.9	8.3	6.20	673.74	**0.000**
Class Days Offered in July	20.9	11.8	22.67	1,166.74	**0.000**

As would be expected, Portuguese is less often used with parents, siblings and friends among Rural students than among Urban students. Table 5-15 presents the data:

Table 5-15. Use of Portuguese with Family and Friends

Portuguese Use with Parents	Urban	Rural	Urban as a % of Rural
Almost Always	57.8%	20.2%	286.1%
Occasionally	14.8%	27.8%	53.2%
Almost Never	9.9%	27.2%	36.4%
Never	17.4%	24.7%	70.4%

N=1,530; Chi-square=200.92, df=3, p=0.000			
Portuguese Use with Siblings/Friends	**Urban**	**Rural**	**Urban as a % of Rural**
Almost Always	54.9%	21.2%	259.0%
Occasionally	16.9%	31.3%	54.0%
Almost Never	10.1%	25.0%	40.4%
Never	18.1%	22.4%	80.8%
N=1,675; Chi-square=183.35, df=3, p=0.000			

This shows us that there are large systematic differences between the two areas in terms of days of class attended and language used outside of the classroom. The large and highly significant performance gaps between urban and rural schools raises an important question: Are the ApaL interventions equally effective in both areas? To answer this, the three treatment groups were compared, separately by grade, in each area. The resulting tables are quite large and numerous, and therefore are contained in Annex H. Table 5-16 presents the highlights of those analyses.

Table 5-16. Comparison of Treatment Group Effects within Urban/Rural Areas – Grade 2

Grade 2	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Oral Comprehension	Urban	8.7	8.9	7.7	0.789 NS	0.098 NS	**0.037**
	Rural	7.4	6.0	6.4	**0.000**	**0.000**	0.194 NS
Concepts about Print	Urban	5.9	5.6	5.2	0.558 NS	0.096 NS	0.428 NS
	Rural	5.3	4.0	4.1	**0.000**	**0.000**	0.949 NS
Letter Recognition	Urban	12.5	10.7	11.4	0.510 NS	0.852 NS	0.953 NS
	Rural	9.5	5.7	5.1	**0.000**	**0.000**	0.772 NS
Familiar Word Reading	Urban	2.6	2.2	3.2	0.770 NS	0.622 NS	0.341 NS
	Rural	2.3	1.1	1.0	**0.000**	**0.000**	0.917 NS
Text Word Reading Fluency	Urban	3.3	3.7	4.4	0.849 NS	0.548 NS	0.816 NS
	Rural	3.5	1.8	1.7	**0.001**	**0.000**	0.917 NS
Reading Comprehension	Urban	0.26	0.39	0.26	**0.045**	0.998 NS	0.167 NS
	Rural	0.24	0.15	0.13	**0.008**	**0.000**	0.772 NS

In Grade 2, virtually no differences are found between treatment and Control groups in the Urban area. In Rural schools, on the other hand, the Full treatment is consistently and substantially higher than either the Medium or Control groups. Medium treatment schools are indistinguishable from Control schools on all EGRA subtests. Table 5-17, below, examines the case in Grade 3.

Table 5-17. Comparison of Treatment Group Effects within Urban/Rural Areas – Grade 3

Grade 3	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Oral Comprehension	Urban	9.8	10.0	8.8	0.742 NS	**0.014**	**0.003**
	Rural	7.8	7.3	7.6	**0.000**	0.614 NS	0.282 NS
Concepts about Print	Urban	7.6	7.5	6.6	0.931 NS	**0.004**	**0.014**
	Rural	6.3	5.9	5.5	0.202 NS	**0.000**	**0.017**
Letter Recognition	Urban	26.5	23.8	24.7	0.546 NS	0.857 NS	0.963 NS
	Rural	16.4	17.2	10.8	0.860 NS	**0.000**	**0.000**

Familiar Word Reading	Urban	8.0	6.5	6.4	0.337 NS	0.997 NS	0.519 NS
	Rural	4.4	4.1	2.9	0.849 NS	**0.002**	**0.013**
Text Word Reading Fluency	Urban	12.2	9.6	9.7	0.480 NS	0.519 NS	1.000 NS
	Rural	6.8	6.5	4.3	0.908 NS	**0.006**	**0.013**
Reading Comprehension	Urban	0.63	0.55	0.51	0.450 NS	0.413 NS	0.940 NS
	Rural	0.42	0.28	0.27	**0.002**	**0.001**	0.992 NS

In Urban schools, Full and Medium treatment groups outperform Control schools only on Oral Comprehension and Concepts about Print by a small margin of 12% - 14%. In Rural schools, both Full and Medium groups show substantially stronger performance on the Concepts about Print, Letter Recognition, Familiar Words, Text Reading Fluency subtests, and Full outperforms both Medium and Control on Reading Comprehension. Full treatment effects tend to be larger than those of Medium schools.

ApaL showed no significant impact in Urban schools a year after cessation of the direct interventions. In Rural schools, however, gains were significant and substantial, especially for the Full treatment model. Attendance is generally higher in Grade 3 than in Grade 2, and in Urban versus Rural schools. What is of particular note is that, whereas no differences in days of attendance are noted between groups in Urban schools, Full schools in Rural areas have approximately 50% more reported days of attendance than either Medium or Control schools, approaching Urban school levels. This represents an important effect of the ApaL project in rural area schools. Table 5-18: Student Days Attended in July 2015, shows one possible cause for the difference in these performance results.

Table 5-18. Student Days Attended in July 2015 by Grade, Area and Treatment Group

Grade	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Grade 2	Urban	12.4	11.4	11.4	0.618 NS	0.758 NS	0.999 NS
	Rural	10.9	6.9	7.8	**0.000**	**0.000**	0.529 NS
Grade 3	Urban	15.1	13.3	14.1	0.217 NS	0.642 NS	0.733 NS
	Rural	13.5	9.0	9.1	**0.000**	**0.000**	0.990 NS

There is one area in which Rural schools surpassed Urban schools. As indicated in Annex H, Rural schools generally retained reading books and other TLAs from the end of the ApaL intervention (2014) through the end of the subsequent (2015) academic year at rates higher than did Urban schools.

Key findings are summarized below.

Summary

- *On every single comparison of results from the EGRA and the Student Coversheet, Urban schools strongly out-performed their Rural counterparts.* This includes student age, days attending and class days offered in July and Portuguese use with parents and siblings/friends, as well as on the EGRA subtests themselves.
- *An extensive analysis conducted shows that a year after project cessation, ApaL Full and Medium interventions had little impact in Urban schools in either grade relative to Control schools, but a large impact in Rural schools.* This is not a new phenomenon; since Baseline, Rural schools have consistently underperformed Urban schools. ApaL interventions have reduced the performance gap relative to Urban schools. Both at Midline 2 and, even more importantly, one year the project cessation, Rural

Full treatment and, to a lesser extent, Medium treatment schools significantly outperform Rural Control schools.

5.4 Unplanned additional training and training activities provided to some school directors, pedagogical directors and teachers during 2015

Contrary to the IE design, which called for no additional inputs from ApaL in the schools making up the sample, ApaL recruited 11 school directors from Full treatment schools of the IE sample for 12 days of additional training and 12 days delivering training to school directors in expansion schools. In addition, 38 pedagogical directors and teachers received 11 days of additional training and carried out 17 days of activities in expansion schools as Reading Coaches.

The teachers and PDs were not drawn from schools with appreciably different performance in 2014. The school directors came from Full treatment schools whose 2014 Grade 2 students performed somewhat better only on Oral Comprehension; in Grade 3 students performed better on the Oral Comprehension, Concepts about Print, and Reading Comprehension EGRA subtests. We therefore conclude that, for the most part, these school directors were not selected based on the performance of students in their schools.

EGRA scores showed no significant differences as a result of teachers/PDs receiving additional training and participating as teaching coaches in 2015 in the expansion schools. The significant amount of time spent in schools other than their own may have reduced the effects of additional training in their own schools. In any case, consideration this particular approach to address ongoing maintenance of ApaL acquired skills does not find support in the findings.

On the other hand, EGRA scores in the schools where directors received additional training 2015 are significantly different in Grade 2 in Oral Comprehension and CAP. In Grade 3 significant differences were found in Oral Comprehension, CAP, Letter Recognition, Familiar Words Reading, Fluency and reading Comprehension. These results allow us to conclude that the amount of training that a SD receives does affect student performance in the year following cessation of other project interventions and these effects are most noticeable in Grade 3. The table that follows compares student performance by whether the SD received (or not) additional training in 2015.

Table 5-19. EGRA Scores by whether the SD received additional training in 2015

Oral Comprehension		SD trained	SD not trained	t	df	Sig. (2-tailed, unequal variances)
Second Grade	2014	9.5	8.3	4.18	155.76	**0.000**
	2015	9.5	7.5	6.57	163.12	**0.000**
Third Grade	2014	10.6	9.0	6.88	155.18	**0.000**
	2015	10.1	8.2	6.50	167.48	**0.000**
Concepts about Print		**SD trained**	**SD not trained**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Second Grade	2014	6.6	6.2	1.85	159.17	0.067 NS
	2015	6.1	5.4	2.77	163.14	**0.006**
Third Grade	2014	8.3	7.4	4.96	217.52	**0.000**
	2015	7.8	6.5	5.44	196.51	**0.000**
Letter Recognition (clpm)		**SD trained**	**SD not trained**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Second Grade	2014	19.6	20.0	-0.14	129.07	0.890 NS
	2015	11.8	10.4	0.76	138.23	0.452 NS
Third Grade	2014	33.9	28.7	1.88	133.15	0.062 NS
	2015	26.4	18.7	2.92	134.02	**0.004**

Familiar Words Read Correctly (cwpm)		SD trained	SD not trained	t	df	Sig. (2-tailed, unequal variances)
Second Grade	2014	3.5	3.4	0.29	150.79	0.773 NS
	2015	3.1	2.3	1.35	128.40	0.181 NS
Third Grade	2014	9.1	7.9	1.19	140.34	0.235 NS
	2015	8.9	5.0	3.62	128.94	**0.000**
Text Reading Fluency (wpm)		**SD trained**	**SD not trained**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Second Grade	2014	5.4	5.2	0.18	152.72	0.858 NS
	2015	4.0	3.3	0.75	145.43	0.456 NS
Third Grade	2014	20.9	20.5	1.54	132.64	0.125 NS
	2015	13.4	7.8	3.18	139.22	**0.002**
Text Reading Comprehension		**SD trained**	**SD not trained**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Second Grade	2014	0.24	0.21	0.49	125.24	0.626 NS
	2015	0.29	0.24	0.67	146.69	]0.385 NS
Third Grade	2014	0.72	0.49	2.53	133.80	**0.012**
	2015	0.73	0.45	3.90	141.36	**0.000**

5.5 USAID Indicators

The tables that follow correspond to four variables that IBTCI provided to be included in the USAID 2015 performance report. They correspond to the following indicators:

1. Average score for third graders on the 30 Familiar Words subtest (timed, one minute).
2. Number and percentage of third grade students getting fewer than 20 correct common words; number and percentage getting 20 or more correct.
3. Number and percentage of third grade students getting fewer than 45 connected text words (timed, one minute); number and percentage getting 45 or more correct.
4. Number and percentage of third grade students getting 1, 2, and 3 or 4 comprehension answers correct corresponding to the story. As before, 3 and 4 were recoded to combine their results, as so few students got four correct comprehension questions asked or answered correctly.

Each of these items is broken out by province, by sex, and by treatment/control group. Totals are given as well to facilitate presentation and understanding of the broken out results. Also included are the 2014 results in a side-by-side manner for all of the variables and their breakouts.

It must be understood that these results correspond to the 2015 application of the EGRA in the original 180 schools of the IE sample, where the project was not active in 2015 (i.e., the sustainability follow-on). These results do not include schools corresponding to the expansion phase of the project.[35] Similarly, the results IBTCI presents on the tables show N's solely relating to the sample of students actually assessed by the EGRA application. These percentages can be projected out to the actual enrollment (or some other basis) in each of the two grades of the 180 schools. Table 5-20 displays the average score of third graders on familiar words correctly read per minute.

[35]The EGRA was administered in the expansion schools by World Education/ApaL, and WEI should be able to provide the corresponding data on the numbers of students enrolled as well as the sample results directly.

Table 5-20. Average Score of Grade 3 Students on Familiar Words Correctly Read per Minute

Province	2015	2014
	Mean score	Mean score
Nampula	5.34	6.46
Zambézia	3.85	5.37
Third graders average scores on familiar words per minute read correctly by sex		
Female	3.86	4.96
Male	5.40	6.85
Third graders average score on familiar words per minute correctly read by treatment group		
Full	5.69	8.09
Medium	4.94	6.50
Control	3.30	3.23

Table 5-21. Grade 3 Students Correctly Reading 20 or more Words per Minute by Sex

Sex	Yes	Yes %	No	No %
Female	54	6.3%	806	93.7%
Male	77	8.5%	731	90.5%
TOTAL	**131**	**7.9%**	**1,537**	**92.1%**

Table 5-22. Grade 3 Students Correctly Reading 20 or more Words per Minute by Treatment Group

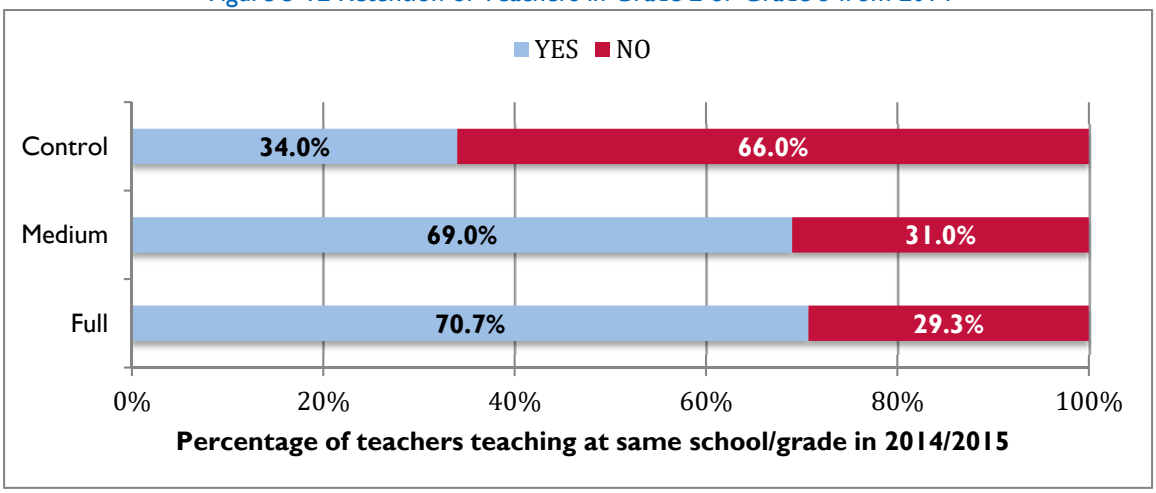
Treatment group	Yes	Yes %	No	No %
Full	61	10.8%	503	89.2%
Medium	39	7.2%	504	92.8%
Control	31	5.5%	530	94.5%
TOTAL	**131**	**7.9%**	**1,537**	**92.1**

5.6 Teachers

Sustainability of interventions such as ApaL is, in large part, dependent on trained individuals remaining in the positions for which they have been prepared for sufficient time. In 2014, teachers of Grades 2 and 3 acquired skills that focused on the improvement of reading in the early grades. The Midline 2 IE report clearly confirms the links between improved teacher performance and student EGRA scores. Teachers also learned to develop TLAs specifically related to those two grades. When those teachers leave the school or no longer teach the same grades, much of the effort put into preparing better teachers is lost.

One unanticipated outcome of the ApaL project was the improved rate of retention of ApaL-trained teachers in treatment schools compared to the Control schools. The 270 Grade 2 and 3 teachers whose classrooms were randomly selected to participate in the EGRA assessment in 2015 were asked whether they had taught either Grade 2 or Grade 3 at the same school in 2014. In the case of the Full and Medium schools, this would mean that they had, or had not, been exposed to training in 2014. Figure 5-12 shows the percentage of teachers who answered affirmatively.

Figure 5-12 Retention of Teachers in Grade 2 or Grade 3 from 2014



N=270

A question that followed was whether in 2015 the teacher had continued teaching the same grade as she/he had done in 2014. Table 5-23 displays the information by grade and by treatment group.

Table 5-23. Retention at the school, teaching the same grade by grade and by treatment group

Teacher Taught Second or Third Grade at this School in 2014?		Full	Medium	Control
Second Grade	Total N	49	44	42
	% Retained	71.4%	61.4%	26.2%
	Chi-square sig.	p = 0.000		
Third Grade	Total N	50	43	42
	% Retained	70.0%	76.7%	61.9%
	Chi-square sig.	p = 0.330 NS		

N=270

In 2015 Full and Medium treatment schools showed much higher retention of teachers in Grade 2 who had taught Grade 2 or Grade 3 in the same school (and therefore received ApaL training and other inputs) during 2014. More than 71% of Full school teachers were retained; at Medium schools the retention was 61%. In comparison, only 26% of current Control school Grade 2 or Grade 3 teachers remained at the schools teaching the same grades. These findings are confirmed by the information provided by school directors during the interviews conducted in 96 Full and Medium treatment schools. Seventy percent of directors of Full schools reported that all the teachers trained by ApaL in 2014 remained at the school teaching the same grades. Medium schools show a lower rate of retention with 60% of the trained teachers remaining at school. We believe that this combination of quantitative and qualitative data, including the great difference in retention between Control schools and Treatment schools suggests a causation between retention and participation in ApaL training as distinct from a correlation, although it is outside the Scope of Work of the IE to identify which aspects of ApaL, if any, have likely led to this increase in retention.

In an ideal world, teachers would remain in the same school teaching the same grade, but reality paints a different picture. Teachers will continue to move from one school to the other for many reasons: they get bored teaching one grade and when there is an opportunity to teach a higher grade, they take it; they move to a school that is closer to their homes or more prestigious; or they have disagreements with school administration and look for posts where they feel more comfortable. At the same time, part of the sustainability of the intervention is dependent on trained individuals remaining in the same position for sufficient time to benefit the school, the students and their colleagues. Thus, the design of an intervention need to take this into consideration and explore incentives or other mechanisms to

improve the likelihood that trained teachers will remain in the school teaching the same grades. An in-depth examination of the reasons behind Full schools retaining 71% of their trained teachers—while Control schools seem to have lost over half of theirs—would provide some useful insights for the implementer.

5.6.1 Classroom Observation (SMA/ApaL)

Second and Third Grade teachers of the classrooms selected for the administration of the EGRA (one for each grade per school) were observed during the Portuguese class on the day of the visit. In 2014, 319 teachers were observed—116 in Full schools, 107 in Medium schools, and 96 in Control schools. In 2015, of the 282 observations conducted 98 were in Full schools, and 82 each in Medium and Control schools.

The observer used a yes/no checklist of 49 teacher practices and instructional behaviors (see Annex D), which were grouped into five major categories:

- Teacher-student interactions
- Teaching decoding
- Teaching comprehension
- Classroom management[36]
- Planning and sequencing of teaching activities.

The items in each category were found to vary considerably in terms of "difficulty" and showed significant correlation among them. An index was therefore created by weighting each item response by the proportion of teachers observed to not demonstrate use of the behavior or practice, thus giving greater weight to more discriminating items. These were summed to provide an index for each of the major categories. The instrument and items were identical to those employed in 2014. The weights applied in 2015 were also the same as those used in 2014 to permit cross-year comparisons.

We consider the classroom observation indices as intermediate variables—the behaviors and practices observed represent the direct outcome of the considerable resources applied through ApaL teacher training and coaching interventions, but also are hypothesized to be causal to EGRA-measured student outcomes in conjunction with other project inputs and teacher-teacher variation in practices employed.

As throughout this report, the first question is whether, in 2015, differences exist between Full, Medium and Control group teachers one year after cessation of interventions in treatment schools. Table 5-24 summarizes the 2015 situation.

Teaches Decoding demonstrated significant differences between the groups, due to very low scores among Control teachers: Full teachers scored three and a half times higher than Control and Medium teachers scored two and three quarters times their Control counterparts. On the Planning and Sequence category, the three groups were different although the pairwise contrasts were not significant. Planning and Sequence was the same in both treatment groups indicating no Full-Medium differential, and Control teachers scored about 27% lower than the treatment groups.

[36] *Some of the items in the Classroom Management section were phrased in the negative (e.g., "teacher did not leave classroom to make/receive a cell phone call"), which means that lower scores on this index are "better". This technical flaw appears to have caused confusion among the observers and as a result the data obtained were not reliable. Thus, we have chosen not to report results for this section.*

Table 5-24. Classroom Practices and Behaviors Observed in 2015 by Treatment Group

2015 Classroom Observation Indices (N=262)	Second and Third Grades			ANOVA Sig. (p =)	Contrast Full with		Contrast Medium with
	Full	Medium	Control		Medium	Control	Control
Teacher-Student Interactions	1.85	1.95	1.67	0.101 NS	0.701 NS	0.337 NS	0.087 NS
Teaches Decoding	1.88	1.46	0.53	**0.000**	**0.036**	**0.000**	**0.000**
Teaches Comprehension	1.97	1.78	1.65	0.349 NS	0.671 NS	0.324 NS	0.842 NS
Planning and Sequence	1.38	1.39	1.02	**0.048**	0.999 NS	0.075 NS	0.086 NS
N =	98	82	82				

Note: Contrasts were performed using the Tukey statistic for Post-Hoc pair-wise comparisons.

These results contrast to those obtained in 2014[37], where treatment groups scored significantly higher on all of the indices. This was also true in pairwise contrasts of Full and Medium teachers with Control teachers. Full teachers outperformed Medium school teachers only on the Comprehension category, suggesting that for the most part, Full and Medium schools performed alike. The tables that follow show the 2014 results.

Table 5-25. Classroom Practices and Behaviors Observed in 2014 by Treatment Group

2014 Classroom Observation Indices (N=319)	Second and Third Grades			ANOVA Sig. (p =)	Contrast Full with		Contrast Medium with
	Full	Medium	Control		Medium	Control	Control
Teacher-Student Interactions	2.25	2.31	1.69	**0.000**	0.788 NS	**0.000**	**0.000**
Teaches Decoding	2.89	2.72	0.41	**0.000**	0.287 NS	**0.000**	**0.000**
Teaches Comprehension	2.72	2.28	1.30	**0.000**	**0.043**	**0.000**	**0.000**
Planning and Sequence	1.80	1.99	0.87	**0.000**	0.329 NS	**0.000**	**0.000**
N =	116	107	96				

From the foregoing, it is clear that the substantial treatment school gains in teacher instructional behavior achieved in 2014 were generally not sustained a year later; for the most part, teachers had not internalized the classroom behaviors ApaL had endeavored to inculcate. The following table tests whether the 2014-2015 differences are significant within each of the three groups, by each of the four practices and behaviors categories. It also shows the percentage of 2014 scores, which were observed in 2015, thus providing a proxy for the relative sustainability of results. The tables that follow compare combined Grade 2 and Grade 3 class observation index scores by year and treatment group for each of the instructional behaviors observed.

Table 5-26. 2015-2014 Comparisons of Classroom Practices and behaviors by Treatment Group

Teacher-Student Interactions	2014	2015	2015 as a % of 2014	t	df	Sig. (2-tailed, unequal variances)
Full	2.25	1.85	82%	3.98	170.43	**0.000**
Medium	2.31	1.95	85%	3.00	146.87	**0.003**
Control	1.69	1.67	99%	0.14	154.21	0.888 NS

[37] *Fewer classroom observations were conducted in 2015 than in 2014, but based on analysis of the distribution of the observations across the samples, these do not appear to have resulted from bias in selection on any key variables.*

Teaches Decoding	2014	2015	2015 as a % of 2014	t	df	Sig. (2-tailed, unequal variances)
Full	2.89	1.88	65%	6.73	160.97	**0.000**
Medium	2.72	1.46	54%	7.45	144.21	**0.000**
Control	0.41	0.53	131%	-1.21	160.15	0.227 NS
Teaches Comprehension	**2014**	**2015**	**2015 as a % of 2014**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Full	2.72	1.97	72%	3.71	204.70	**0.000**
Medium	2.28	1.78	78%	2.27	165.00	**0.024**
Control	1.30	1.65	127%	-1.77	150.85	0.079 NS
Planning and Sequence	**2014**	**2015**	**2015 as a % of 2014**	**t**	**df**	**Sig. (2-tailed, unequal variances)**
Full	1.80	1.38	77%	2.83	207.45	**0.005**
Medium	1.99	1.39	70%	3.55	160.44	**0.001**
Control	1.71	2.01	118%	-1.13	146.98	0.262 NS

As expected, Control teachers did not have different results in 2015. Both Full and Medium school teachers scored significantly lower in 2015 relative to 2014 scores on all categories. Teacher-Student Interactions retained the most of 2014 scores (82% in Full, 85% in Medium), while Decoding Activities lost the most (65% retained in Full, 54% in Medium). The others range from 70% to 78%. We conclude that the patterns seen in 2015 are due to the substantial loss of 2014 gains, that is to say, the relatively weak sustainability of teacher behavioral and teaching practices gains obtained through the treatment interventions, in the absence of continuing direct support through those interventions. It is also generally clear that the Full treatment schools fared no better in this regard than did Medium schools.

Teacher sex was not related to any of the four category index scores in any of the three groups in 2014 or in 2015. The classroom grade (Grade 2 or Grade 3) being observed was also not associated with scores on any of the categories in 2015. In 2014, small but significant differences were found favoring Grade 3 in Full schools on Teaching Comprehension, and favoring Grade 2 over Grade 3 on Teacher-Student Interactions in Control schools. We conclude that these factors did not affect the degree of sustainability on teacher practices and behaviors in 2015. The main findings of this section are summarized below.

Summary

- *The turnover of personnel seems to have an impact on teacher instructional behaviors and practices* even if mentoring support, and perhaps other ApaL inputs, were available and absorbed by those teachers who received it.
- *About 70% of Grade 2 treatment school teachers were retained; only 33% of Control school Grade 2 teachers were.* In Grade 3, all groups had about 71% retention in the same grades. Maintaining trained teachers in their positions is an important element contributing to the sustainability of this intermediate variable.
- Turnover is a fact of life in Mozambique, as well as in many other countries, and must be taken into account when designing interventions.

5.7 The Classroom Environment

Aside from the teacher, the classroom environment in which students are to learn consists of many different factors, some of which the ApaL treatment interventions attempt to address. These cover a range of characteristics including days of class offered in a given month, teacher and student attendance rates, Portuguese class duration observed, TLAs and other materials availability, and more. Data obtained from students, teachers, direct observation and enrollment data for Grades 2 and 3 at the

school and from the SDEJT are blended to provide a picture of the classroom environment a year after ApaL interventions ceased. When appropriate these data are compared to data from 2014 to help assess the sustainability of improvements gained in 2014.

The number of class days offered is well below the official calendar. The student interview recorded data from the classroom logbook for days of classes offered and attended in July.

Table 5-27. July 2015 Class Days Offered and Attended in July 2015 by Grade

July 2015 Classroom Days	Grade	Full	Medium	Control	ANOVA Sig. (p =)	Contrast Full with		Contrast Medium with
						Medium	Control	Control
Offered	Grade 2	17.2	13.8	13.6	**0.000**	**0.000**	**0.000**	0.918 NS
Attended		11.6	8.4	8.3	**0.000**	**0.000**	**0.000**	0.977 NS
Offered	Grade 3	19.1	14.3	14.1	**0.000**	**0.000**	**0.000**	0.903 NS
Attended		14.3	10.2	9.9	**0.000**	**0.000**	**0.000**	0.881 NS

Consistently, Full treatment schools offered more classroom days than either Medium or Control did, in both grades, up to five more days in the month. Medium and Control schools were indistinguishable. Attendance days followed suit, with Grade 2 students attending 3.3 additional days in the month and Grade 3 students attended 4.4 days more. These data give us an approximate attendance rate for students selected for the EGRA in 2015.

Table 5-28. Attendance rate of students assessed

Attendance Rate	Full	Medium	Control
Grade 2	67.4%	60.9%	61.0%
Grade 3	74.9%	71.3%	70.2%

We looked at self-reported data from the teacher interview for days missed for the month of July. The results are displayed on Table 5-29, on the next page. No differences were found between groups, and with an average of well less than one day reported missed, teachers are either under-reporting their absences or a considerable number of days were lost to school closing (if school is closed, a teacher might not consider it missed) or other similar factors.

Table 5-29. Teacher-reported days missed in July 2015

2015 Teacher-Reported Days Missed in July	Full		Medium		Control		Total	
	N	%	N	%	N	%	N	%
None	53	53.5%	44	51.2%	45	53.5%	142	52.8%
1 or 2	29	29.3%	26	30.2%	21	25.0%	76	28.3%
3 or more	17	17.2%	16	18.6%	18	21.4%	51	19.0%
Total	99	100.0%	86	100.0%	84	100.0%	269	100.0%

Chi-square = 0.978, df = 4, p = 0.913 NS

Portuguese class duration was obtained from the start and stop time of the observed class. This was compared to the official duration, and the variance determined in minutes. The following table shows that in 2015 no differences were associated with treatment or control group.

Table 5-30. Portuguese Class Duration in 2015

2015 Observed Portuguese Class Duration	Full		Medium		Control		Total	
	N	%	N	%	N	%	N	%
Short 15 minutes or more	8	8.2%	6	7.4%	7	8.5%	21	8.0%
Short 5 to 14 minutes	31	31.6%	23	28.4%	20	24.4%	74	28.4%
Short 1 to 4 minutes	21	21.4%	16	19.8%	20	24.4%	57	21.8%
Official duration or more	38	38.8%	36	44.4%	35	42.7%	109	41.8%
Total	98	100.0%	81	100.0%	82	100.0%	261	100.0%

Chi-square = 1.672, df = 6, p = 0947 NS

2014 showed generally better times for all groups, with more classes meeting or surpassing official duration times, and fewer classes in the categories short five to 14 minutes and 15 minutes or more. These differences, however, were not significant between groups as shown on Table 5-31.

Table 5-31. Observed Portuguese Class Duration in 2014

2014 Observed Portuguese Class Duration	Full		Medium		Control		Total	
	N	%	N	%	N	%	N	%
Short 15 minutes or more	5	4.3%	8	7.5%	1	1.0%	14	4.4%
Short 5 to 14 minutes	23	19.8%	22	20.6%	18	18.8%	63	19.7%
Short 1 to 4 minutes	28	24.1%	23	21.5%	19	19.8%	70	21.9%
Official duration or more	60	51.7%	54	50.5%	58	60.4%	172	53.9%
Total	116	100.0%	107	100.0%	96	100.0%	319	100.0%

Chi-square = 6.420, df = 6, p = 0.378 NS

The Classroom Inventory collected information from the teacher on enrollment and attendance on the day of the visit in September-October, as well as direct counts of students possessing different materials. These data were broken out by girls and boys. The following table shows the comparison of these variables between treatment and Control groups by sex in 2015.

Table 5-32. Attendance, reading materials and supplies shown by students on the day of the visit by sex and treatment group

2015	Sex	Full %	Medium %	Control %	ANOVA Sig.	Contrast Full with		Medium with
						Medium	Control	Control
Present / Registered	Girls	47.8	40.7	44.7	0.098 NS	0.079 NS	0.605 NS	0.480 NS
	Boys	48.6	41.0	39.3	**0.017**	0.076 NS	**0.022**	0.889 NS
Portuguese Book	Girls	77.8	80.7	78.1	0.741 NS	0.053 NS	0.998 NS	0.807 NS
	Boys	76.4	82.4	77.2	0343 NS	0.354 NS	0.983 NS	0.486 NS
Reading Book(s)	Girls	45.6	34.4	9.0	**0.000**	0.170 NS	**0.000**	**0.000**
	Boys	45.9	37.6	5.1	**0.000**	0.363 NS	**0.000**	**0.000**
Notebook	Girls	97.7	89.3	93.2	0.529 NS	**0.001**	0.107 NS	0.227 NS
	Boys	92.8	92.1	94.8	0.529 NS	0.946 NS	0.689 NS	0.521 NS
Pencil	Girls	94.3	91.0	91.6	0.222 NS	0.246 NS	0.379 NS	0.962 NS
	Boys	91.3	93.3	92.2	0.655 NS	0.627 NS	0..906 NS	0.883 NS

Note: Sig. is significance of differences between groups determined by ANOVA with unequal variances assumed.

Student attendance varies significantly by treatment group only for boys, where Full schools do better than Control schools. Attendance of girls relative to registration also tends to favor the Full schools, although this did not achieve statistical significance.

The other variables in the table represent the percentage of students present who were able to show possession of each material. MINEDH has been committed to providing the basic textbook to all students in these grades in all of Mozambique, and indeed about 80% of students had them on the day of the visit. Notebooks and pencils were shown by 92-95% of all students, regardless of sex or treatment group.

ApaL directly provided reading booklets (*"livrinhos de leitura"*) to students in Grades 2 and 3 in 2014. In 2015, 45% of Full school students, and 35-38% of Medium school students had one or more of these booklets—while only 5-9% of Control students did. Thus, the highly significant overall differences between groups derives from Full or Medium versus Control advantages. The difference between Full and Medium schools was not significant. For reference, the following table presents the same results for 2014.

Table 5-33. Attendance, reading materials and supplies shown by students on the day of the visit by sex and treatment group (2014)

2014	Sex	Full %	Medium %	Control %	ANOVA Sig.	Contrast Full with		Medium with
						Medium	Control	Control
Present / Registered	Girls	51.4	45.0	45.3	0.060 NS	0.089 NS	0.125 NS	0.995 NS
	Boys	47.7	42.6	38.0	**0.007**	0.194 NS	**0.005**	0.309 NS
Portuguese Book	Girls	77.0	79.0	85.1	**0.048**	0.821 NS	**0.044**	0.821 NS
	Boys	76.3	79.4	82.0	0.283 NS	0.647 NS	0.255 NS	0.761 NS
Reading Book(s)	Girls	77.3	78.2	4.9	**0.000**	0.979 NS	**0.000**	**0.000**
	Boys	77.0	79.0	5.0	**0.000**	0.896 NS	**0.000**	**0.000**
Notebook	Girls	94.6	93.0	93.6	0.745 NS	0.780 NS	0.797 NS	1.000 NS
	Boys	94.9	92.7	92.1	0.434 NS	0.585 NS	0.970 NS	0.970 NS
Pencil	Girls	94.6	92.4	91.3	0.397 NS	0.624 NS	0.389 NS	0.915 NS
	Boys	92.4	91.7	90.6	0.792 NS	0.960 NS	0.775 NS	0.912 NS

Note: Sig. is significance of differences between groups determined by ANOVA with unequal variances assumed.

In 2014, Boys attendance slightly lags behind Girls, although attendance of boys was improved in both Full and Medium schools relative to that of students in the Control schools. The possession of reading booklets in Full and Medium schools was 77-79%, equaling that of the MINEDH textbook. Only about 5% of Control school students could show supplementary reading materials. As in 2015, no differences were found between the high levels of possession of other materials between the treatment or Control groups. We then compared possession of supplementary reading materials between 2015 and 2014, both to confirm the significance of the differences and to estimate the "survivorship" of these materials.

Table 5-34. Possession of supplementary materials between 2015 and 2014

Reading Book Possession: 2015 - 2014	Full		Medium		Control	
	Dif.	Sig.	Dif.	Sig.	Dif.	Sig.
Girls	59.0%	**0.000**	44.0%	**0.000**	183.5%	0.267 NS
Boys	59.7%	**0.000**	47.6%	**0.000**	103.0%	0.962 NS
Total	58.7%	**0.000**	45.2%	**0.000**	151.3%	0.444 NS

Note: Dif. Is the 2015 possession rate as a percentage of that of 2014 Sig. is the significance level determined by two-tailed t Test with unequal variances assumed

Full schools retained about 59% of the rates of 2014 student with possession of supplementary reading materials; Medium schools averaged 45%—these declines are quite significant in both cases. Control schools showed some fluctuation between 2014 and 2015, but these changes were not significant due to the small percentage of students having these materials in either year. No association with student sex was found in any of the groups.

During 2014, the ApaL project promoted enrichment of the classroom environment to enhance acquisition of reading skills. The Classroom Inventory observed the presence of these and other materials in both 2015 and 2014.

Table 5-35. Materials observed in the classroom

Classroom was Observed to Have…	2015				2014			
	Full	Medium	Control	Sig.	Full	Medium	Control	Sig.
Alphabet chart	51.0%	41.5%	6.1%	**0.000**	82.8%	81.3%	3.1%	**0.000**
Material to create words from letters	53.3%	32.9%	7.3%	**0.000**	68.1%	73.8%	3.1%	**0.000**
Movable letters or words	39.8%	26.8%	3.7%	**0.000**	73.3%	74.8%	3.1%	**0.000**
Posters with letters or words	26.5%	22.0%	3.7%	**0.000**	55.2%	60.7%	1.0%	**0.000**
Permanent blackboard	92.9%	89.0%	91.5%	0.661 NS	90.5%	91.6%	87.5%	0.607 NS
Student-made materials displayed	7.1%	3.7%	0.0%	0.141 NS	5.2%	6.5%	1.0%	**0.045**
Teacher-made materials displayed	32.7%	20.7%	4.9%	**0.000**	26.7%	33.6%	1.0%	**0.000**
Girls, boys seated together	73.5%	80.5%	78.0%	0.620 NS	85.3%	86.0%	83.3%	0.861 NS
Students seated rows/groups	89.8%	93.9%	89.0%	0.502 NS	92.2%	94.4%	87.5%	0.200 NS

Note: Sig. is significance of differences between groups determined by ANOVA with unequal variances assumed.

No project impact was noted for the presence of a permanent blackboard in either year, in line with expectations since no blackboards were provided and blackboards are nearly universal. Very substantial effects in 2014 between the intervention groups and the Control group are noted for the presence of an alphabet chart, materials to create words from letters, letter or words items movable by the teacher, posters on the walls with letters or words and teacher-made materials on display. Though lower in 2015, the same pattern exists for these variables. No effect was noted in either year regarding seating arrangements in the classroom, as these are highly practiced in most schools. The display of student-created materials was universally low in all schools, regardless of year or treatment type.

We examined the degree to which the observed gains in intervention schools in 2014 were sustained in the 2015 classrooms. The following table looks at the percentage of the 2014 score that was observed in 2015.

Table 5-36. Materials sustained in 2015

Classroom Characteristics: 2015 as a Percentage of 2014	Full		Medium		Control	
	Dif.	Sig.	Dif.	Sig.	Dif.	Sig.
Alphabet chart	61.6%	**0.000**	51.0%	**0.000**	196.8%	0.355 NS
Material to create words from letters	78.3%	**0.000**	44.6%	**0.000**	235.5%	0.220 NS

Movable letters or words	54.3%	**0.000**	35.8%	**0.000**	119.4%	0.846 NS
Posters with letters or words	48.0%	**0.000**	36.2%	**0.000**	370.0%	0.264 NS
Permanent blackboard	102.7%	0.537 NS	97.2%	0.561 NS	104.6%	0.390 NS
Student-made materials displayed	136.5%	0.555 NS	56.9%	0.366 NS	0.0%	0.320 NS
Teacher-made materials displayed	122.5%	0.348 NS	61.6%	**0.046**	490.0%	0.144 NS
Girls, boys seated together	86.2%	**0.034**	93.6%	0.323 NS	93.6%	0.378 NS
Students seated rows/groups	97.4%	0.537 NS	99.5%	0.888 NS	101.7%	0.754 NS

Note: Dif. Is the 2015 possession rate as a percentage of that of 2014 Sig. is the significance level determined by two-tailed t Test with unequal variances assumed

The sustainability of 2014 materials availability in Full schools ranged from 48% to 78% for the presence of an alphabet chart, materials to create words from letters, letter or words items movable by the teacher, and posters on the walls with letters or words. Medium schools followed the same pattern, with sustainability percentages ranging from 36% to 51%; teacher-made materials displayed, at 57%, was barely significant. Control schools showed no change on any of the nine items between 2014 and 2015.

5.8 School Directors/Pedagogical Directors and Their Perspectives on Sustaining ApaL Interventions

In this section we use information obtained by conducting 96 in-depth interviews with school directors (SD) and pedagogical directors (PD) to attempt to answer some *why* and *how* questions. First, it was important to document whether any changes had taken place in the schools between 2014, when ApaL was directly supporting the schools, and 2015, when the intervention had ceased. Had the strategies promoted by ApaL in 2014 continued at the same level? Were school management routines that proved to be effective (according to school directors interviewed in 2014) maintained? Were TLAs—considered by directors interviewed in 2014 the most important ApaL contribution—used at the same level? Second, the interviews focused on whether technical and financial resources were available at the school to continue ApaL initiatives and which were or were not within the technical, managerial, and financial resources available to the school director.

The fieldwork involved collecting information at 52 Full (54%) and 44 Medium (46%) treatment schools where ApaL had been implemented in 2014. Sixty-four school directors and 32 pedagogical directors were interviewed between September 14 and October 7, 2015. The procedure was to arrive at the school and interview the school director and when the interviewee agreed, the interview was recorded. When the SD was absent the interview was conducted with the pedagogical director (PD). Table 5-37 summarizes the sources of information.

Table 5-37. Summary of interviews conducted

Provinces/Districts	SDs/PDs Interviewed	School Directors	Assigned to school in 2015	Additional days of ApaL training in 2015
Nampula				
Monapo	20	14	1 PD	1
Murrupula	10	5	1 SD	1
Nampula Cidade	24	13	1 SD	2
Sub-Total	54	32	3	4
Zambézia				
Mocuba	23	19	2 SDs	1
Nicoadala	11	8	-	2
Quelimane	8	5	-	-
Sub-total	42	32	2	3
TOTAL	96	64	5	8

Out of the 96 SDs and PDs interviewed, one SD and two PDs in Nampula and two SDs and one PD in

Zambézia had not been part of ApaL implementation in 2014—that shows a very low turn-over rate of 6.2% when compared to teachers. Two SDs from Full treatment schools, one for Medium treatment schools, and three PDs from Medium treatment schools reported being new to the school. That reinforces the importance of providing training to school and pedagogical directors since retention of trained individuals is viewed as a key factor for the sustainability of an intervention. We do not have information on turnover rates for SDs and PDs in Control schools.

In Nampula 41% of the interviews were conducted with PDs, showing that on the day of the visit almost one half of the school directors were not at the school. In Zambézia 24% of the interviewees were PDs; this means that one-third of the SDs were not present on the day of the visit. One key determinant of teacher absence seems to be the high absence among school directors. A survey recently conducted by World Bank on Support Delivery Indicators (SDI) in seven African countries (Mozambique included) [38] showed that when a school director was present at the school the average teacher absence rate was 34 percent, while when a school director was absent the teacher average absence rate was 64 percent. This implies that in schools where the school director was absent, teachers were almost twice as likely to be absent, suggesting that leadership and accountability matters in the performance of teachers.[39]

In line with the above and one explanation for better results reported for students in Full treatment schools, it is useful to note that a larger percentage of Full treatment schools (73%) were present on the day of the visit than school directors of Medium treatment schools (54%).

As mentioned earlier in this report, eight school and pedagogical directors participated in additional training and coaching activities provided by ApaL in 2015. This had not been part of the original design of the IE—no activity related to ApaL should have taken place in 2015 at the schools which were part of the IE sample. These individuals came from schools that in 2014 had significant differences in scores for only Oral Comprehension in Grade 2 (8% higher) and Grade 3 (6%) Grade. We conclude that the additional training and the coaching practice provided to the pedagogical directors and to the teachers in 2015 reflect neither selection based on school performance nor resulted in significant changes in outcomes.

In contrast, at the end of the 2015 school year, Grade 2 students of additionally trained school directors out-performed their counterparts on Oral Comprehension (27% higher) and Concepts about Print (14%). Grade 3 students out-performed their counterparts on all six EGRA subtests: Oral Comprehension (23% higher), Concepts about Print (19%), Letter Recognition (41%), Familiar Words Read Correctly (71%), Reading Fluency (72%) and Reading Comprehension (62%). We conclude that, while little gain is observed in Grade 2 results, Grade 3 students responded positively and consistently to the additional training provided to school directors.

5.8.1 ApaL strategies sustained one year after the end of the intervention

The interviews give us a representative snapshot of the sustainability of the processes implemented by ApaL in 2014 and provide information on various dimensions of sustainability allowing us to assess the availability at the school of key inputs, such as decodable books, alphabet charts, TLAs, etc. and school directors' perceived training needs as well as the resources available at the school to sustain processes and procedures implemented by ApaL in 2014.This information is helpful in the identification of facilitators and barriers to sustainability in similar settings. The chart below indicates, the respondents' perception of the components of the intervention that remain in the school one year after the cessation of ApaL activities.

---

[38] World Bank (2015). Mozambique Service Delivery Indicators: Education

[39] Interviewers were instructed to conduct interviews with the school director and only in the absence of the SD, conduct the interview with the PD.

*Which activities, processes implemented by ApaL continue to be used at this school in 2015? **	*%*
School management routines—bell at the beginning of the school day and at the beginning/end of class periods; monitoring of absences and tardiness of teachers and students; class observation; communication with staff and feedback; the five management routines	*56*
Production, organization and use of TLAs in the reading class	*16*
Focus on reading; intensive use of decodable books	*13*
Methodology of teaching reading—lesson plans; five steps to teaching reading; teaching the sounds of letters before showing the letters, etc.	*13*

** Percentages are given considering the 128 references made by 96 interviewees. In Zambézia, 2% of the respondents made reference to strategies to reduce gender gap.*

5.8.2 Respondents' view of the training they require to sustain the intervention

All respondents—from Full or from Medium treatment schools—consider that they need additional training if they are to continue implementation of the ApaL strategies. The need for school management training is shared by 70% of respondents regardless of the treatment group to which the school was assigned. Specifically the training speaks to the establishment and maintenance of school management routines, the use of forms to monitor punctuality and absenteeism of teachers and students, feedback given to teachers, and financial management. Sixteen percent of respondents request additional training in the methodology of teaching reading to early graders so that they can better assist their teachers. The remaining 14% would like to receive training on how to organize and conduct training sessions to improve early grade reading at their schools.

Also, as reported by ApaL, in some areas teachers prefer to attend courses at an *Instituto de Formação Profissional* (IFP),[40] where there are tangible benefits, to participating in ApaL training: having certificates recognized by MINEDH leads to salary raises. This may compromise the sustainability of training provided by ApaL.

5.8.3 Trained teachers who remain in school teaching the same grades

A key question when considering barriers to sustainability is related to the number of trained teachers who remain in the school teaching the same grades from one year to the other. The school or pedagogical directors interviewed were asked to inform the number of teachers trained by ApaL in 2014 who were still at the school teaching the same grades. Seventy-eight percent of interviewees at the 52 Full treatment schools indicated that all Grade 2 and Grade 3 teachers trained by ApaL in 2014 continued at the same school teaching the same grades. The remaining 18% schools indicated that one or more ApaL-trained teachers had left the school, either as a result of transferring to other schools or for illness-related reasons. In some cases these teachers had moved to other grades.

Forty-eight percent of those interviewed at the 44 Medium treatment schools indicate that all Grade 2 and Grade 3 teachers trained in 2014 were still at the school at the end of 2015 teaching the same grades. In four schools in Mocuba it was reported that all of Grade 2 and Grade 3 teachers that had been trained by ApaL in 2014 were no longer at the school.[41] This is important information for ApaL as it continues to emphasize teacher training.

5.8.4 Training sustained in 2015

Training was a main activity of ApaL during the 2014 school year and included not only teachers but also resource teachers (coaches) who would be the facilitators once the project ceased at that school. Therefore, one key question on the interview protocol was whether any training focused on teaching

[40] *Teacher Training College*
[41] *EPI Montero, EPI Munhiba, EPI Mucoia, and EPI Hacanis2, all in Mocuba.*

reading to Grade 2 and Grade 3 students had occurred at the school in 2015. In Nampula 17 (23%) respondents reported that training had taken place at the school in 2015—12 were at Full treatment and 5 at Medium treatment schools. The remaining respondents (64%) stated that no training had taken place at the school. In Zambézia, where 42 interviews were conducted, 69% of the respondents reported that no training had taken place at the school in 2015. Of the 13 (31%) who reported training, 10 were at Full treatment and 3 at Medium treatment schools. In both provinces training seems to occur more often at Full treatment schools and at the schools that are head of ZIPs. Most often the training is conducted by the ZIP coordinators or the coaches.

Table 5-38. Training received in 2015

Training received in 2015 on teaching reading to early grades	Full	Medium
Training was provided by the ZIP or conducted at the school in form of internal training led either by the SD, the PD or the coach. Internal training included group discussions, exchange of experiences between and among teachers, and class observation and monitoring.	22 (42%) of 52 SDs and PDs interviewed	11 (25%) of 44 SDs and PDs interviewed
No training was conducted at the school in 2015 the area of reading. No training was received from the ZIP.	29 (56%)	33 (75%)

N=96

5.8.5 Support received to continue activities promoted by ApaL

One-third of the interviewees report that the school has not received any support from the ZIP or from the District to improve reading skills in the early grades. The remaining two-thirds mention the role of the ZIP in the distribution of MINEDH materials such as books, chalk, paper, cardboard, etc. Some explicitly refer to training provided and/or meetings organized by the ZIP to exchange experiences, make recommendations, and clarify teachers doubts and still others mention visits made by the district and the ZIP to address issues such as teacher and student absenteeism.

5.8.6 The Use and Replacement of Teaching and Learning Aids (TLAs)

TLAs available to teachers and to the school have a limited life and it was expected that some would have to be replaced. While materials such as cardboard and chalk may be expendables, though subject to verification from ApaL, we believe that the reading booklets (*"livrinhos de leitura"*) and materials such as alphabet charts, letter cards, and the like should probably last for about three years.) In 29 (30%) of the schools there was no need to develop, replace or add teaching and learning materials in 2015. Materials received from ApaL in 2014 were available and still in good condition. One group of nine schools in Zambézia (Mocuba) needed to develop, replace or add materials, but this was not done because those interviewed at this group of schools reported not having received the ADE[42] funds and resources from MINEDH that would allow them to add, replace, or develop materials. The remaining 58 schools replaced or added materials specifically alphabet charts, cards with letters and words, word games, and clay objects. They also made copies of forms to control attendance and punctuality and of pictures without words. Interviewees indicated that resources provided by MINEDH through ADE coupled with teachers' ability to develop TLAs as a result of having participated in ApaL, allowed the production of TLAs to continue at least in some schools.

Summary

- At the schools where interviews were conducted, the turn-over rate for school/pedagogical directors was low—5%. Five interviewees—three in Nampula and two in Zambézia—indicated that they had been assigned to the school in the 2015 school year.

[42] Schools receive from MINEDH resources to buy necessary materials such as paper, chalk, cardboard, etc. through the Auxilio Direto à Escola (ADE) or Direct Assistance to Schools. The amount sent to each school is based on the number of students enrolled on March 3rd of each year.

- At Full treatment schools 27% of the SDs were absent on the day of the visit compared to the 46% absentee rate of SDs at Medium treatment schools.
- At the end of the 2015 school year, Grade 2 students of additionally trained school directors out-performed their counterparts on Oral Comprehension (27% higher) and Concepts about Print (14%). Grade 3 students out-performed their counterparts on all six EGRA subtests: Concepts about Print (23% higher), Concepts about Print (19%), Letter Recognition (41%), Common Words Read Correctly (71%), Reading Fluency (72%) and Reading Comprehension (62%).
- The school or pedagogical directors interviewed were asked to inform the number of teachers trained by ApaL in 2014 who remained at the school teaching the same grades. In Nampula, out of the 54 interviews, 32 (59%) indicated that all Grade 2 and Grade 3 teachers trained by ApaL in 2014 continued at the school teaching the same grades. In Zambézia, out of the 42 interviewees, 23 (55%) indicated that all Grade 2 and Grade 3 teachers trained in 2014 were still at the school at the end of 2015 teaching the same grades.
- In Nampula 63% of the 54 interviewees stated that no training focused on teaching reading to early grades had taken place at the school during 2015. In Zambézia, where 42 interviews were conducted, 64% reported that no training had taken place at the school in 2015. It was reported by interviewees that bringing teachers from adjoining schools to attend training conducted at the ZIP head school remains a challenge.
- Most commonly, interviewees cited the unavailability of resources for their lack of implementation of strategies and process promoted by ApaL in 2014. Strategies most often continued as reported by the interviewees include (1) the bell at the start of the day and at the start and end of each class period; (2) the control of teachers and student attendance; (3) the communication mechanisms implemented by ApaL such as feedback; (4) the organization and continuation of the use of TLAs.
- One-third of the interviewees indicate that the school has received no support either from the ZIP or from the District. The remaining two thirds mention the ZIP's role in the distribution of MINEDH materials such as books, chalk, paper, cardboard, etc. They also refer to training provided and/or meetings organized by the ZIP to exchange experiences, make recommendations, and clarify teachers' doubts. Seven interviewees mention visits made by the district and/or the ZIP to address issues such as teacher and student absenteeism.

TLAs have a limited shelf life. In 2015 many schools needed to replace materials specifically the "*quadro de pregas*" and the cards with letters. Other materials that needed replacement include the alphabet charts, images and forms that needed to be copied, and letter games. Forty-seven percent of those interviewed state that they still have materials brought to the school by ApaL and that they take good care of these materials so that they will last. Forty-one percent consider that the materials received from ADE (direct support to school, a MINEDH fund) allows them to continue the effort and when necessary build new materials with local resources. Individuals in both groups refer to the training received by ApaL as key as it facilitates the building of new materials when needed. In seven schools the director stated that the school has no resources to sustain the strategies recommended by ApaL. In nine schools there was the need to replace the materials but this was not done because the school did not receive the material usually distributed by MINEDH—paper, cardboard, etc. that would allow the development of new TLAs.

5.9 Evaluation Question # 4: Management Sustainability

The fourth evaluation question was stated as: **Of the most cost-effective interventions, which fall within the existing technical and financial management capacity of local education institutional personnel? What capacity-building activities would be required to ensure**

sufficient MINEDH technical and financial management capacity to implement the interventions? Based on the data available, mostly the interviews conducted with school directors, this section of the report we discuss and provide some answers to this question. Additional insights may be derived from the WEI analyses of the SMA.

The findings reported in 2013, 2014, and 2015 clearly show the impact of the Full intervention. Students in Full treatment schools score significantly higher than their counterparts in Medium and especially in Control schools, they come to class more often, and have access to more teaching and learning materials. In addition, as show in this report, Full treatment schools retain more of the gains observed in 2014, more of their ApaL trained teachers continue teaching the same grades and they share lower absentee rates for teachers and school directors. The cost-effectiveness analyses conducted in 2014 and in 2015 show that for the same investment, more gains are associated with Full treatment schools than with the Medium alternative.

These findings underscore the importance of the school director and suggest that investing in the selection and the preparation of SDs may be the single most effective strategy to transform and improve schools. For this to happen, some activities required to ensure sufficient MINEDH technical and financial management capacity to implement the Full intervention. Two main courses of action should be considered.

The first, to be conducted at the central and provincial levels, is to assist MINEDH in the development of procedures and of a set of criteria for the selection of school directors. The intention here is to select school directors on the basis of criteria that are transparent and relevant to the work they are expected to perform as school leaders. The second course of action involves the preparation of SDs already on the job or of those that are new to the function. The low turnover of directors—only 5% of the 96 SDs interviewed were new to the school—emphasizes the benefits to be accrued by providing these individuals with the opportunity to enhance their school management skills. This preparation could be conducted by the use of a training manual with accompanying CDs and yearly meetings at the provincial capital led by SDs identified selected by the superior performance of their schools.

Recent research suggests that one of the main factors related to school effectiveness is the leadership exerted by school directors (Leithwood & Son, 2012; Leithwood, et al., 2005; Marks & Printy, 2003). Based on the interviews conducted we can start to draw a profile of the SD and of the areas where capacity-building activities are required to ensure sufficient management capacity to implement interventions such as proposed by ApaL.

- Most SDs interviewed mentioned the importance of the training provided by ApaL to the teachers and how this training contributed to the improvement of reading scores. However, few SDs were able to give examples of specific actions that they had taken at the school level in order to improve the teaching skills of the staff.
- Few directors mentioned observing classes or providing feedback to their teachers based on class observation. The need for a "how to" procedure such as a class observation instrument was cited by the interviewees. This point is especially important because, as we show in the report, observed instructional behavior that led to improved student higher reading scores in 2014 was not sustained in 2015.
- SDs are aware that tardiness and absenteeism of teachers and students contribute to the low level of student achievement but few directors mentioned strategies to reduce these to acceptable levels and when probed could not clearly outline a plan to do so.
- Most SDs interviewed show limited ability to gather and use information (that could and should be available at the school) to make decisions. For example, data on student or teacher

absenteeism could be kept current and shared with parents of absentee children; teacher tardiness and absenteeism could be discussed in staff meetings; student age and repetition rates could indicate the magnitude of the problems the school faces with overage and repetition. Few SDs are aware of the importance of keeping these data up to date and, most importantly, to utilize them to improve the performance of the school.

- When school resources are discussed, all school directors mention the ADES funds. Few directors engage in collaborative work with external agents–other schools, parents, the community–to explore the expansion of resources. When probed, most SDs could not describe how they would involve the community in the school improvement effort.
- Most school directors emphasize the role as administrators and fail to see themselves as instructional leaders. Actually, few directors understand the concept of instructional leadership and what it entails.

The gaps noted above as a result of the interviews conducted show where school management training could make an impact. They also point to the need for a targeted study to investigate how leadership is conceptualized and practiced at the school that could contribute to a deeper knowledge of school management styles and practices that result in some schools being more successful than others.

5.10 Costs and Cost-Effectiveness over the 2014-2015 Period

Cost-effectiveness of the 2014 implementation year was analyzed in the Midline 2 report, where we took cost data by treatment group, and divided it by the number of potential beneficiaries—students enrolled in second and third grade in each of the two intervention treatments—to arrive at unit costs. Effectiveness was determined within the treatment groups and by grade for each of the six EGRA subtests in terms of absolute gains relative to the corresponding Control group. This provided both the cost both for the absolute and relative gain in EGRA scores. Key results discussed in other sections of the Findings chapter of this report are incorporated as part of our analysis of the extent to which ApaL's interventions have been sustained.

To extend this analysis to include the 2015 school year, we expand the total number of potential beneficiaries to include the respective 2015 enrollments (virtually halving the unit costs, as no additional project inputs were provided in 2015) and use the combined 2014 and 2015 gains on EGRA scores (which lowers the project impact on each subtest).[43] Table 5-39 presents the unit costs per student enrolled for both the 2014 and combined 2014-2015 school years:

Table 5-39. Unit costs per student enrolled for 2014 and combined 2014-2015

Costs for Intervention Types	2014	2014		2015	Combined 2014 and 2015	
	Total Costs	Enroll-ment	Cost per Student	Enroll-ment	Enroll-ment	Cost per Student
Medium Treatment	$ 480,997	52,710	$ 9.13	48,614	101,324	$ 4.75
Additional for Full	$ 68,295	28,282	$ 2.41	24,738	53,020	$ 1.29
Total for Full Treatment			$ 11.54			$ 6.04

[43] *In 2014, we reported the cost per school ($ 5,062.19 in Full, $ 3,942.60 in Medium), cost per teacher/classroom ($710.33 in Full, $ 566.55 in Medium), and per student ($ 11.54 in Full, $ 9.13 in Medium). No additional costs were incurred in 2015, so the per school and per teacher/classroom figures remain. The per student costs for the present analysis are adjusted for the combined 2014 and 2015 enrollments in second and third grades and applied to the results obtained through the IE samples in both years.*

This allows us to estimate the cost per observed gain in a "one year with intervention, one year without" model. Table 5-40 presents these data for Grade 2:

Table 5-40. Cost effectiveness Grade 2 "one year with intervention, one year without" model

Grade 2 Combined 2014 and 2015	Mean Scores			Gains over Control		Cost per Gain (US$)		Cost per % Gain (US$)	
	Full	Med.	Cont.	Full	Med.	Full	Med.	Full	Med.
Oral Comprehension	8.2	7.3	6.7	1.5 (22%)	0.6 (8%)	4.03	7.92	0.27	0.59
Concepts about Print	5.9	5.1	4.4	1.5 (34%)	0.7 (17%)	4.03	6.79	0.18	0.28
Letter Recognition	15.3	12.3	5.6	9.7 (173%)	6.7 (120%)	0.62	0.71	0.03	0.04
Familiar Word Fluency	2.9	2.0	1.2	1.7 (137%)	0.8 (66%)	3.55	5.94	0.04	0.07
Text Reading Fluency	4.3	3.3	1.9	2.5 (132%)	1.4 (77%)	2.42	3.39	0.05	0.06
Reading Comprehension	0.23	0.18	0.09	0.14 (156%)	0.09 (100%)	43.14	52.78	0.04	0.05

Notes: All group means different by ANOVA (p=0.000). Tukey significance used for pair-wise post-hoc contrasts. Full vs Medium p=0.000 except Familiar Word Fluency (0.002) and Reading Comprehension (0.005). Full vs Control contrasts p=0.000 on all subtests. Medium versus Control p=0.000 on all subtests.

For the combined 2014-2015 Grade 2 students, we see that for all EGRA subtests, the mean scores are significantly different for all pair-wise contrasts in both Full and Medium treatment groups relative to the Control group. Full schools enjoy a much larger advantage (range of 22% to 173%) over Control schools than do Medium schools (range 8% to 120%) in terms of gains. The cost per unit gain achieved in Full schools is lower than in Medium schools, resulting in a lower cost of the relative gain (percentage improved over Control) in Full versus Medium schools. The data for Grade 3 are presented in Table 5-41.

Table 5-41. Cost effectiveness Grade 3 "one year with intervention, one year without" model

Grade 3 Combined 2014 and 2015	Mean Scores			Gains over Control		Cost per Gain (US$)		Cost per % Gain (US$)	
	Full	Med.	Cont.	Full	Med.	Full	Med.	Full	Med.
Oral Comprehension	8.9	8.3	7.8	1.1 (14%)	0.5 (6%)	9.52	17.63	0.43	0.79
Concepts about Print	7.2	6.7	5.7	1.5 (26%)	1.0 (18%)	6.06	8.82	0.23	0.26
Letter Recognition	24.9	23.4	12.5	12.4 (100%)	11.0 (88%)	1.08	0.96	0.06	0.05
Familiar Word Fluency	6.9	5.7	3.3	3.7 (112%)	2.4 (73%)	2.72	3.21	0.05	0.07
Text Reading Fluency	11.7	9.6	5.1	6.6 (130%)	4.5 (88%)	1.42	1.56	0.05	0.05
Reading Comprehension	0.53	0.43	0.15	0.38 (253%)	0.28 (187%)	35.08	37.79	0.02	0.03

Notes: All groups different by ANOVA (p=0.000). Tukey significance used for pair-wise post-hoc contrasts. Full vs Medium p=0.000 except Letter Recognition (0.266 NS); Familiar Word Fluency (0.001); Reading Fluency (0.003). Full vs Control contrasts p=0.000 on all subtests. Medium versus Control contrasts p=0.000 on all subtests except Oral Comprehension (0.001).

In Grade 3, the mean scores again are significantly different, and all pair-wise contrasts between groups are very significant, with the one exception of Letter Recognition, where Full did not better Medium

schools. The advantage of Full over Control schools ranged from 14% to 253%, showing higher relative gains in the actual reading subtests. Medium versus Control schools showed relative advantages ranging from 6% to 187%, again following the pattern of larger relative gains as the subtest difficulty increases. In terms of costs, Full schools have lower cost/unit and cost/percentage gain figures than do Medium schools. For Oral Comprehension and Concepts about Print, the generally high scores in Grade 3 make further improvement relatively expensive

SUMMARY

- Unit costs per student of the two interventions were significantly reduced by combining the 2014 and 2015 enrollments.
- Especially Full treatment schools, but also Medium, continue to out-perform their Control school counterparts on all EGRA subtests.
- Although results in the intervention schools in 2015 were not as strong as during the 2014 intervention year, enough gains remained so that, combined with the virtually double number of beneficiaries, unit costs were lowered and the cost-effectiveness of the ApaL interventions improved relative to the 2014-only cost effectiveness analysis previously reported on. Full schools continue to show stronger cost-effectiveness results relative to the Medium treatment, confirming that the best model of intervention, as was also reported in the 2014 cost-effectiveness analysis, is the Full model.

6. CONCLUSIONS

Based on the data presented in this report we conclude that participation in the ApaL program caused children to improve their basic reading skills, even a year after project cessation of direct intervention and in particular in letter recognition and familiar words. Only 42% of Full, 31%% of Medium and 26% of Control Grade 3 students could correctly answer any of the four comprehension questions. This poor reading comprehension performance can be directly linked to low reading fluency scores, which, while still low, do show clear signs of project impact. Evidence from EGRA studies worldwide, as well as our own findings, confirms that there is a strong predictive relationship between these early pre- and reading skills and later reading fluency. Only students at the top decile of treatment group Grade 3 (scoring 40 words or more per minute) were able to correctly answer one or more (average 1.5) reading comprehension questions. The second highest decile (reading an average of 22 words per minute) achieved an average of only 0.8 questions correct. Therefore, although reading comprehension strategies are important at all ages, the instructional focus at Grades 1 and 2 must be on the fundamental early skills of phonological awareness and alphabetic principles, coupled with practice in applying these skills to decoding of new words. Regular and ample practice will increase the automatic recognition of words and this will translate into increased fluency, which in turn, will increase children's ability to read with comprehension.

Our findings suggest that children who participate in programs such as ApaL have a better probability of succeeding in school later on by learning the reading skills essential for learning in all subject areas. However, the intensity of inputs still is not sufficient to have raised overall means to benchmark levels. This is partially due to the persistence of a large number of zero scores across the subtests, especially on the Fluency and Reading Comprehension sub-tests, which lowers the means substantially. Means would be higher if the number of those scoring zero decreased. Therefore, efforts should be made to move children towards the center of the score distribution. An examination of the distribution of scores in Grades 2 and 3 shows that, in many cases, the better are getting better, while the non-performers are not being as effectively reached.

The project impact was not uniform for all segments of the target population. The impact of

the program on urban schools one year after the end of direct intervention is practically undetectable. At schools located in rural settings, students profit more from the ApaL program than their counterparts at schools located in urban settings. One possible cause may be the higher rate of student attendance in Full schools in rural areas. It is possible to conclude that ApaL is more effective in rural settings where the same inputs have more impact, although rural schools continue at a disadvantage relative to urban schools.

Boys and girls benefit equally from the ApaL program, on average, and both score higher than their counterparts in Control schools. However, benefitting equally means that the differences noted between the performance of girls and that of boys will continue to exist and possibly widen as children advance to the next grade as indicated by the findings reported in Section 5. The increased magnitude of sex differences in Grade 3 suggests that greater efforts will be required to address the root causes of under-performance of girls.

In addition to reporting outcomes of the early grade reading assessment, the IE also examined factors that seem to be associated with learner achievement and that increase the likelihood that the effects will remain after the intervention ceases. This information can help educators and policy makers prioritize actions.

Although the sustainability of the intervention was generally low, maintenance at Full schools was higher than at the Medium treatment schools. Compared to Medium, SDs at Full schools appear to have made better use of available school resources. Full schools were better able to retain trained staff teaching early grades. In-service training as implemented by ApaL in 2014 was especially challenging to sustain. Directors would describe meetings and informal sessions with teachers and feedback that had taken place at the school but they did not consider this to be *training*. Schools generally do not have resources to conduct training, pay teachers to attend meeting at the ZIPs on Saturdays, or cover the costs of traveling to the ZIP as was provided by ApaL. These factors limit the sustainability of the training component of the ApaL model.

The reading intervention, which includes training school directors as one factor, is responsible for the higher scores observed in Full treatment schools. School directors are in a unique position to influence, motivate, and monitor teachers. Whether it is the content of training, the teaching method, the increased time on task, the reduced absenteeism or a combination of the above cannot be discerned through this study methodology. This, combined with the larger and consistent gains shown by Full over Medium, and especially over Control, leads us to conclude that directors are more likely to maintain the routines acquired during training. The important conclusion is that in schools where the school management component was implemented in 2014 and where school directors received additional training in 2015 children had better results and came closer to the goal of reading with comprehension.

One year of full implementation is not sufficient for teachers to internalize the instructional behaviors promoted by ApaL. Except for "Teaching Decoding," teacher instructional behavior in 2015 has fallen to levels similar to prior implementation levels (2013, Midline 1). The findings from the class observation data indicate that once direct support from the program ended, the majority of teachers fell back to their previous mode of instruction. In comparison, interviews conducted with Full school directors show that over 70% of them retained the initiatives championed by ApaL. This, combined with the larger and consistent gains shown by Full over Medium, and especially over Control, leads us to conclude that directors are more likely to maintain the routines acquired during training.

Receiving the Full treatment had a ripple effect on schools in a number of different areas. These schools are more able to retain their trained Grade 2 and 3 teachers, post lower rates of director, teacher and student absenteeism, and more days of teaching per semester. Nevertheless, 70% of all SDs interviewed, whether from Full or from Medium treatment schools, consider that they need

additional training in school management. Specifically, training that promotes the establishment and maintenance of school routines, the use of forms to monitor punctuality and absenteeism of teachers and students, the improvement of feedback given to teachers, and the strengthened financial management.

Taken together, the findings suggest that investing in the training of school directors may therefore be more effective and necessary than simply providing teachers with additional training. Our findings show that students whose teachers received additional ApaL training in early grade reading in 2015 did not score significantly higher than those whose teachers did not, despite their being selected from schools whose scores in 2014 were no different from the latter. This leads us to conclude that the additional teacher training received in 2015 did not, in itself, benefit Grade 2 and Grade 3 students at their school. On the other hand, additional training provided to school directors seems to make a difference: at Full schools whose directors received additional training in 2015 students scored higher on all subtests.

The sustainability of reading materials (decodable books provided by ApaL) in the possession of students dropped significantly in treatment schools from the levels observed in 2014. Full schools fell from 77% in 2014 to 46% in 2015, a sustainability rate of 59% and in Medium schools, from 77% to 36% for a sustainability rate of 46%. Classroom materials TLAs, materials to create words from letters, moveable letters or words, and posters with letters or words also dropped significantly in 2015: only about 50% to approximately 60% of classroom materials inputs of the ApaL project survived one year after cessation of their provision in 2014.

When examined over a two-year period (2014 and 2015) per-student unit costs are reduced to one half and enough performance gains remain in 2015 so that overall cost-effectiveness of the ApaL project was enhanced. This was especially so for the Full treatment schools. It is not possible to project sustainability of ApaL inputs nor student performance gains beyond 2015, but the relatively rapid decline in teacher practices, teacher retention, student materials and classroom materials suggests that, without on-going maintenance or renewal of these inputs, only a small effect will be noticeable in subsequent non-intervention years.

In sum, intervention sustainability in development settings such as Mozambique continues to represent a significant challenge to project implementers. Despite the enthusiastic endorsement and support of school/pedagogical directors at the end of 2014 and the successes of full implementation in 2014 (Midline 2 report), it was expected that staff would find it challenging to implement the intervention with fidelity without support in 2015. In general, the interviews with school directors indicate that use of the intervention strategies was lower at Medium schools with only the use of TLAs and school routines such as the bell at the beginning of the school day and at the beginning and end of each class periods demonstrating any notable continuation. This may be due to constraints at each school (e.g., teacher and school director absenteeism and turn-over, limited number of days allotted to school and of hours of actual teaching/learning, among others) and the limited resources available to schools.

Improving the sustainability of an intervention will require careful examination of successes achieved in the different segments of the target population. Assessing the readiness for the adoption of an intervention could provide important information regarding the time and effort required to implement initiatives such as those promoted by ApaL. Incorporating and directly targeting context variables such as parents or the community (Glisson, 2002) may be one way to enhance sustainability. Changing the intervention model to ensure sufficient time for the internalization of innovation (rather than a pre-determined period of one school year) may also increase long-term impact.

7. RECOMMENDATIONS

Many of the basic recommendations made in the Midline 1 and 2 reports are still valid. Specifically, (a) increasing time spent learning basic literacy skills, (b) teaching sounds of all letters of the alphabet, (c) providing students with books that focus on decoding and word-building skills, training teachers to use these strategies in the classroom, (e) putting into place strategies to encourage students and their families to increase the number of words read, and (f) especially providing school directors with the training needed to improve their schools. There are challenges to be faced when attempting to modify behaviors and procedures that have been in place for a long time but ApaL is making an effort to implement the recommendations.

1. Make sustainability part of the implementation plan of an intervention. Sustainability should not be an afterthought to be addressed at the end of an intervention. A specific description of the measures that will be taken to make the intervention sustainable should be required as an important element of the design of a project. ApaL has worked with district officers and school directors to make sure that trained teachers remained in in their schools and classes, to ensure that skilled trainers are located in each ZIP/district, and that teaching and learning materials continue to be used.

2. Strengthen project Monitoring and Evaluation (M&E). A well-developed Monitoring and Evaluation (M&E) component with specific and measurable indicators should be required from implementers as part of the project design, independent of the external evaluation of project results and impact. In addition to the monitoring of project activities and other inputs, greater focus on the achievement of outcomes represented by well-defined and agreed upon indicators will enhance the implementation of a project provide information to correct its course when necessary.[44]

3. Adopt a non-linear implementation model to provide support more or less intensely as required in different situations. Rather than follow an implementation-and-immediate-scale-up model, interventions such as ApaL may require a more extended period of support to allow for internalizing and routinization of activities. Withdrawing the support according to a fixed schedule (e.g., the end of one school year) without considering the level of routinization achieved may significantly decrease both the effects and the sustainability of an intervention.

4. Assess the level of readiness for the innovation prior to the implementation of the intervention. A range of instruments has been developed and used to identify specific concerns of potential adopters and stakeholders. These instruments provide insights into issues that must be addressed to ensure widespread acceptance, adoption and sustainability of an intervention. Implementers should be required to assess the barriers to sustainability and the factors that facilitate it prior to starting the implementation of a program.

5. Implement the program where the effects are greater. Findings suggest that the program functions better in some settings than in others. For example, on an absolute basis, children in rural schools consistently show lower scores than do children in urban schools—in fact, on average, students in Full treatment rural schools performed worse than students in urban Control schools. This notwithstanding, our findings indicate that, in 2015, ApaL's effects in rural schools are larger 2015 the effects observed in urban settings. That is, although ApaL does not eliminate the relative deficit in learning for rural children, it significantly reduces it. The findings suggest that the program, as currently

---

[44] *Indicators for project outcomes need to be specific, measurable, attainable, relevant and time-bound. For example, Improving reading skills or Expanding time available to teach reading do not qualify as SMART indicators. Indicators developed during the design phase can be adjusted, if necessary, based on the findings obtained at Baseline.*

designed and implemented, may be more appropriate for rural schools, and that in the future, those schools should become ApaL's main target.[45]

6. Identify the reasons for girls' consistent underperformance and include in all projects strategies to close the gender gap. Both Full and Medium ApaL interventions reduced the difference in performance between boys and girls, when compared to Control schools in the same grade. The program benefited both boys and girls equally but has not eliminated the gender gap observed in 2013 and 2014. Findings show that while ApaL narrows the relative gap between boys and girls, the absolute gap widens as students advance to Grade 3 and when more complex skills are assessed. The increased magnitude of sex differences in Grade 3 suggests that greater efforts will be required to address the root causes of under-performance of girls.

7. Work with MINEDH to improve the ways school directors are selected and prepared. School directors are key to educational improvement. This is shown first by the higher scores of students in Full treatment schools, which is probably closely associated with greater attendance by teachers and students both, and confirmed by the finding that scores were higher at schools where school directors received additional training in 2015. We recommend that ApaL work with MINEDH and district leadership to identify staff with leadership potential to become a school director and develop a clear set of criteria for applicants. Second, because many principals learn the skills they need on-the-job, they need continuous in-service opportunities learn how to improve school management.

8. Provide incentives to keep trained teachers at the school teaching the early grades. The management component received by Full treatment schools contributed to a higher rate of retention of trained teachers in those schools. Nonetheless, approximately 30% of the teachers trained in 2014 either left the school or were assigned to other grades. This highlights the need to develop and strengthen incentives to keep trained teachers in the same grades in subsequent years.

9. Use more effectively the data produced by the School Management Assessment (SMA) instruments, checklists, and assessments. This would require the reduction of available data to rate schools in simple categories of school management aspects. The "grades" assigned would show clearly where a school stands in terms of factors that relate to student performance such as student, teacher and staff attendance, days of class offered, or hours of instruction per day. This will reinforce the use of data to identify and monitor school management factors causal to student performance. This type of exercise should also form part of the MINEDH school supervision process.

10. Consider the sustainability of the various inputs provided as part of the program when designing future projects. Some inputs provided by ApaL, such as TLAs, seem to have only a moderate level of sustainability (50% - 60%) while others were even less sustainable. For example, given the amount of teacher training conducted, only decoding activities showed any difference in the classroom a year after implementation. All inputs require on-going assistance in order to continue. We must identify ways to improve the sustainability of the inputs themselves and ensure that necessary funding will be available to maintain them. This needs to be a part of the intervention design.

---

[45] ApaL has indicated that, with the exception of treatment schools in Nampula City and Quelimane, all schools in the project 2015 and 2016 scale-up are rural schools.

ANNEXES

Annex A. Scope of Work

Annex A.1. Scope of Work for the Evaluation Services Task Order

AID-656-TO-12-00002, EGRA+QIM Impact Evaluation

SECTION C – STATEMENT OF WORK

C.1 BACKGROUND

Following a long colonial period, a 10-year war for independence, and 16 years of civil war, Mozambique is rebuilding its education sector and strengthening its capacity to provide quality educational services. Under the colonial government, the Portuguese provided very little education for most Mozambicans. By 1962, after centuries of Portuguese rule, only 25 percent of the population had any education at all (Candido et al, 1986). The illiteracy rate at independence, in 1975, was estimated to be as high as 93 percent. The civil war, which lasted from 1977 to 1992, resulted in the destruction of 50 percent of school infrastructure (especially primary schools) and saw many teachers kidnapped or killed.

Despite these challenges, access to primary education in Mozambique expanded rapidly after the civil war ended in 1992. The Government of Mozambique (GOM) created a national system of primary school education and, between 2003 and 2007, the number of children in primary school increased from 3.3 million to 5.3 million at an average growth rate of 8 percent per year. Retention of children in school has improved from 30 percent in 2006 to 41 percent in 2010 (Mozambique Fast Track Initiative (FTI) and Catalytic Fund Application, 2010).

Although available budget has not fully met increasing demand for services, the GOM has continuously increased allocations to the education sector, consistently allocating around 20 percent of its budget, including both internal and external sources, to the education sector, with more than half these funds going to primary education. The GOM has also initiated reforms in public financial management, decentralization, and human resource management aimed at improving the provision of basic services. In the education sector, increasing operational budgets have been managed at the school level, and specific funds have been channeled to both provincial and district accounts for supervision activities.

Despite laudable progress in access made since 1992, key challenges still exist and will have to be overcome to help the primary school system teach children the basic skills required for the country's economic and social development. Improving the quality and increasing the quantity of instructional time, crucial for academic success, remains a challenge and an elusive goal across the public primary school system.

The rapid expansion of the primary school system in sheer numbers of children has not been accompanied by a similar increase in the number of modern classrooms. Double and triple shifts of teaching (early morning/midday/afternoon) have become standard among the majority of schools in the system, meaning that students spend a limited amount of time in the classroom and receive a low quantity of instruction in basic skills such as reading. Education is a vital element in the socio-economic and democratic transformation of Mozambique. Progress along the development continuum and full participation in a competitive global economy will require Mozambique to address critical educational quality and learning outcomes issues, especially in light of lackluster performance as has been highlighted in a sequence of assessments (USAID/ Educational Quality Improvement Program 2 (EQUIP 2) Aga Khan 2011; Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) 2007). Moreover, many Mozambicans believe that education is one of the key tools for improving socio-economic conditions, and narrowing the gap between rich and poor.

C.2 STATEMENT OF NEED

The title of the program to be implemented under this contract is the "*Impact Evaluation (IE) for the Early Grade Reading Assessment Plus Quality Instruction and Management (EGRA+QIM) program".* The main purpose of the IE is to use the findings and conclusions to guide USAID, the Ministry of Education (MINED), and the EGRA+QIM program to select the most powerful interventions to be brought to scale in subsequent school years within the EGRA+QIM program's timeframe.

In addition, the IE will provide the rigorous evidence needed to advocate effectively with the government and other donor agencies about the best way to invest scarce resources to improve the quality of education, more specifically reading outcomes, in primary schools.

Given the lack of rigorous external and independent evaluation evidence on EGRA+QIM type interventions in USAID Africa, this IE has the potential to raise awareness of the effectiveness of such interventions. The results of the IE will also be shared globally with other international partners active in early grade reading, such as the World Bank, networks such as the Early Grade Learning Community of Practice, and academic organizations or think tanks that will be able to use the findings from the IE to strengthen the research base on early grade reading.

In full support and in alignment with USAID Forward, the IE will demonstrate the Agency's renewed commitment to learning, systematically monitoring performance, and rigorously evaluating program impact. USAID will link enhanced monitoring and evaluation efforts to its program design, budgeting and strategy work. The IE contract will be one of several mechanisms for implementing the USAID/Mozambique Basic Education Program for the 2012 to 2017 period.

The IE contract will contribute towards achieving the global USAID Education Strategy – Goal 1: "Improved reading skills for 100 million children in primary grades by 2015;" and toward the fourth objective of the USAID/Mozambique Country Assistance Strategy (CAS): "Expand opportunities for quality education and training."

Based on evidence in multiple countries and in accordance with the global USAID Education Strategy and the United States Government (USG) Mozambique CAS, USAID/Mozambique has determined that the most effective way to support education—and ultimately the country's economic and social development—is by focusing resources on improving early grade reading outcomes.

1. TARGET GEOGRAPHICAL REGIONS AND BENEFICIARIES

The EGRA+QIM program will target the provinces of Nampula and Zambezia in central and northern Mozambique. The rationale for selecting these two provinces can be summarized as follows: These two provinces combined contain 42 percent of the school-age population of Mozambique and, according to national statistics, these heavily populated and rural provinces have posted weak education performance results compared to national averages. .

In addition, girls in these two provinces are grossly underserved and disadvantaged. The national illiteracy rate for women is 62.7 percent; in Nampula it is 77.4 percent, and in Zambezia 79 percent (Multiple Indicator Cluster Survey (MICS), 2008). Furthermore, despite clear and

pressing needs, there is relatively little bilateral or multilateral support for education in these provinces, with the exception of UNICEF working in only one district in each province.

The IE Contractor will collaborate closely with the EGRA+QIM implementing partner and operate in the same provinces and with the same target students, teachers, school directors and schools to evaluate the impact of the various interventions introduced.

2. RATIONALE AND NEED FOR IMPACT EVALUATION OF THE EGRA+QIM PROGRAM

USAID and governments in Africa urgently need rigorously generated evidence about the effectiveness of education interventions to help them make informed decisions about the allocation of scarce resources in the struggle to improve early grade reading outcomes. In Mozambique, the government is poised to take pro-active steps towards establishing quality standards in basic education, but lacks the necessary assessment instruments and skills to effectively measure learning outcomes. This IE, which will be the first external and independent impact evaluation of its kind in USAID Africa, will provide a thorough understanding of the benefits of EGRA+QIM type programming as it relates to improving early grade reading outcomes.

The IE will utilize a randomized control trial (RCT) methodology, which is recommended by the new USAID evaluation policy and which is considered the gold standard in measuring causal impacts. A carefully designed RCT can provide exactly the rigorous evidence needed to advocate effectively with the government and donor agencies about the best way to invest scarce resources to improve reading outcomes in primary schools.

Given the lack of rigorous external and independent evaluation evidence on EGRA+QIM type interventions in USAID Africa, this IE has the potential to raise awareness of the importance of such interventions as well as the importance of conducting well thought out impact evaluation with the goal of improving reading outcomes.

3. DEVELOPMENT HYPOTHESIS

The IE will test USAID/Mozambique's EGRA+QIM program development hypothesis that, reading outcomes in grades 2 and 3 will improve if the quality and quantity of reading instruction in those grades is improved through better teacher training and coaching and strengthened school management.

C.3 OBJECTIVES

The main objective of the IE will be to measure the causal effect of EGRA+QIM program treatment interventions on improving early grade reading outcomes in grades 2 and 3 in targeted schools. The evaluation will test two treatment models of early reading interventions, against a control group that will not receive any of the interventions. The first model is the "medium" treatment model, which will include training, coaching and support for teachers; and the second is the "full" treatment model which will include training, coaching and support for teachers, and additional school management training for directors.

Impact of these interventions will be measured at two levels. First, the evaluation will determine which of the two models is the most effective in achieving the planned results to improve early grade reading in the EGRA+QIM program; and secondly, which of the two models is the most cost-effective in producing the reading outcomes achieved, measured through unit costs per student, and technical/management sustainability.

EVALUATION QUESTIONS

The IE will address the following overarching question:

1. To what extent have EGRA+QIM treatment interventions improved early grade reading outcomes for students in target schools.

Specifically the IE will address the following questions related to the impact of the two intervention models:

a. Impact of "Medium" Treatment Intervention - To what extent does the "reading instruction support" treatment intervention cause early grade reading outcomes to improve for students in grades 2 and 3 in target schools whose teachers have received training, coaching and support (direct student beneficiaries)?

b. Impact of "Full" Treatment Intervention – To what extent does the treatment intervention of additional "school management" training, coaching and support to school directors cause a significant and additional improvement in early grade reading outcomes when coupled with "reading instruction support" in target schools?

c. Cost Effectiveness - To what extent are the various "medium" and "full" treatment interventions cost-effective? Specifically, what are the most significant reading outcome effects and unit costs per student, per teacher, per school director, per school of the key treatment interventions?

d. Management Sustainability – Of the most cost effective interventions, which fall within the existing technical and financial management capacity of local education institutional personnel? What capacity building activities would be required to ensure sufficient MINED technical and financial management capacity to implement EGRA+QIM interventions?

C.4 SCOPE OF WORK

1. MAIN ACTIVITIES

A. **Develop Detailed Inception Report** - The IE Contractor will develop a detailed inception report; which will include: a detailed evaluation design; a comprehensive plan to fully address each evaluation question; proposed sampling methods; proposed methods (quantitative and qualitative) for data collection and analysis; and a list and description of proposed data sources to be used. The Inception Report should include a Final Evaluation Work Plan detailing the main activities and objectives to be achieved during the full 30 month IE period. Additionally, the Final Evaluation Work Plan will outline how the Contractor will "hit the ground running." This includes having the appropriate staff positioned and ready to go, a list of activities to be accomplished, responsible personnel, and a timeline for proposed activities through December 31, 2014. The Detailed Inception Report is due 45

days after the effective date of the award.

The IE Contractor will draw upon but is not limited to the following data sources: EGRA assessment data, performance data and progress reports from the EGRA+QIM program, and available data and information from MINED, local education institutions, and civil society. The IE Contractor will employ appropriate qualitative methods to gather information on the local operating environment of schools in target areas to effectively contextualize and interpret data, findings and results.

B. **Design, pilot, and implement RCT to establish baseline and evaluate EGRA+QIM interventions** - The IE Contractor will design, pilot, and provide oversight for an RCT to establish a baseline and evaluate key EGRA+QIM treatment interventions. The IE Contractor will analyze results to determine the most cost-effective interventions.

Using data from an EGRA assessment, the IE Contractor will gather empirical data (ensuring that USAID evaluation policy and impact evaluation standards are rigorously applied) from the RCT to establish a baseline, track progress in reading outcomes, evaluate treatment interventions, and inform the selection of interventions to be brought to scale in subsequent school years.

The IE Contractor will design the RCT to evaluate the non-treatment (control) and the two treatment interventions. It is noted that the design, piloting, and implementation of the RCT should strictly adhere to USAID evaluation policy and acceptable professional education evaluation standards.

C. **Analyze Results of EGRA Tests** - The IE Contractor will analyze EGRA assessment results, taking into account gender equality and rural/urban factors, and mother tongue issues. Data should be disaggregated by sex, school, grade level, and language group.

D. **Disseminate Results of EGRA Tests** –The IE Contractor will broadly disseminate results of the EGRA assessment in an accessible form to key stakeholders (See Attachment J.1: Illustrative Communication and User Engagement Plan). The IE Contractor will also make all assessment instruments, data, analysis, results, reports, and supporting documents available to all key stakeholders, partners and the public.

2. METHODOLOGY

The IE will apply an experimental design using RCT to test the impact of the "medium" and "full" EGRA+QIM models on reading levels among 2nd and 3rd graders in Nampula and Zambezia provinces, Mozambique.

The EGRA+QIM contractor will be responsible for designing and implementing intervention activities of both the "medium" and "full" models, and develop strong, positive and productive relationships with *local education institutions and personnel*, as well as, with key MINED directorates, civil society, and education sector donors. The IE contractor will work closely with the EGRA+QIM implementing contractor in designing and conducting the IE activities. The IE contractor will be responsible for:

1. Working with USAID and the EGRA+QIM in developing a sampling frame for evaluation that will include randomly selecting 60 schools that will participate in the "full" model

days after the effective date of the award.

The IE Contractor will draw upon but is not limited to the following data sources: EGRA assessment data, performance data and progress reports from the EGRA+QIM program, and available data and information from MINED, local education institutions, and civil society. The IE Contractor will employ appropriate qualitative methods to gather information on the local operating environment of schools in target areas to effectively contextualize and interpret data, findings and results.

B. Design, pilot, and implement RCT to establish baseline and evaluate EGRA+QIM interventions - The IE Contractor will design, pilot, and provide oversight for an RCT to establish a baseline and evaluate key EGRA+QIM treatment interventions. The IE Contractor will analyze results to determine the most cost-effective interventions.

Using data from an EGRA assessment, the IE Contractor will gather empirical data (ensuring that USAID evaluation policy and impact evaluation standards are rigorously applied) from the RCT to establish a baseline, track progress in reading outcomes, evaluate treatment interventions, and inform the selection of interventions to be brought to scale in subsequent school years.

The IE Contractor will design the RCT to evaluate the non-treatment (control) and the two treatment interventions. It is noted that the design, piloting, and implementation of the RCT should strictly adhere to USAID evaluation policy and acceptable professional education evaluation standards.

C. Analyze Results of EGRA Tests - The IE Contractor will analyze EGRA assessment results, taking into account gender equality and rural/urban factors, and mother tongue issues. Data should be disaggregated by sex, school, grade level, and language group.

D. Disseminate Results of EGRA Tests –The IE Contractor will broadly disseminate results of the EGRA assessment in an accessible form to key stakeholders (See Attachment J.1: Illustrative Communication and User Engagement Plan). The IE Contractor will also make all assessment instruments, data, analysis, results, reports, and supporting documents available to all key stakeholders, partners and the public.

2. METHODOLOGY

The IE will apply an experimental design using RCT to test the impact of the "medium" and "full" EGRA+QIM models on reading levels among 2nd and 3rd graders in Nampula and Zambezia provinces, Mozambique.

The EGRA+QIM contractor will be responsible for designing and implementing intervention activities of both the "medium" and "full" models, and develop strong, positive and productive relationships with *local education institutions and personnel*, as well as, with key MINED directorates, civil society, and education sector donors. The IE contractor will work closely with the EGRA+QIM implementing contractor in designing and conducting the IE activities. The IE contractor will be responsible for:

1. Working with USAID and the EGRA+QIM in developing a sampling frame for evaluation that will include randomly selecting 60 schools that will participate in the "full" model

intervention, 60 schools in the "medium" intervention, and 60 schools that will serve as control schools with no EGRA+QIM interventions;

2. Designing, piloting, and providing oversight for the RCT that evaluates the impact of the various interventions;

3. Collecting and analyzing all quantitative and qualitative data associated with the evaluation process, including baseline, mid-term, and end line data;

4. Using empirical data generated from the RCT to determine the extent to which EGRA+QIM interventions caused improvements in early grade reading outcomes in target schools.

5. Generating the baseline for all future impact and performance evaluation activities related to the EGRA+QIM program, as well as, inform USAID and MINED's selection of the most cost-effective interventions to bring to scale in subsequent years.

The Contractor will design and set up an impact evaluation that will allow for the comparison of a control group and the two intervention treatments. The control group of 60 schools, randomly selected from the target provinces will not receive any program intervention but whose performance will be measured. The control group represents the proximate counterfactual in the evaluation component of the program—that is, what reading achievement levels would be over the same period of time in the absence of any program intervention.

The first treatment group, the "medium" intervention, will comprise 60 additional, randomly selected schools; the "medium" group of schools will receive an intensive slate of training, coaching and support activities focused on improving reading instruction. The second treatment group, the "full" intervention, will be another set of 60 schools. The "full" treatment group will receive a slate of training, coaching and support activities targeting reading instruction, and an additional slate of activities to improve school management focusing on increasing instructional time devoted to reading.

Qualitative data collection methodologies should also be used to compliment the experimental RCT, to further understand the challenges, obstacles, and motivation experienced by teachers, students, and administrators before, during and after implementation of the EGRA+QIM program.

The IE Contractor will ensure that USAID evaluation policy is adhered to and that rigorous impact evaluation standards are maintained including the responsibility for determining the random assignment of schools into treatment groups, and the validity of findings accounting for school clustering and the largest urban and rural schools in economic corridors in the target provinces.

3. LOCAL COLLABORATION AND COLLABORATION WITH EGRA+QIM CONTRACTOR

The IE Contractor will partner with a local indigenous non-governmental organization (NGO), local university, or a private sector company to recruit, train, and manage locally hired staff to help conduct the IE. Additionally, through the evaluation process (design, pilot, implementation, and analysis) the IE Contractor will collaborate closely with the EGRA+QIM Contractor to

design, pilot, implement and analyze results for a successful RCT. Furthermore, the IE and EGRA+QIM Contractors will coordinate closely so as not to duplicate activity or engage staff positions and resources in a manner that is redundant.

4. GUIDING PRINCIPLES

The following are guiding principles that will help in the successful implementation of this evaluation. The Contractor should review the themes listed below carefully, to ensure they are considered throughout the evaluation.

a. USAID Forward

The IE Contractor will adhere to key reform initiatives described under *USAID Forward* which emphasizes new partnerships, local capacity building, innovation, and rigorous monitoring and evaluation to achieve results. The Contractor should focus on ways to strengthen host country systems and build local technical and managerial capacity to ensure sustainability.

b. USAID Evaluation Policy

The IE Contractor should ensure that the Impact Evaluation follows the USAID Evaluation Policy requirements for rigorous impact evaluations. In addition, the IE should lead to more focused and collaborative education investments aimed at identifying low unit costs and at improving learning outcomes and institutional sustainability in the host country.

c. Data Quality Standards

The IE contractor must ensure that the Impact Evaluation adheres to USAID's requirements for data quality. USAID data quality standards are detailed in Automated Directives System (ADS) 578 and ADS 203, which will be provided to the IE Contractor.

d. Data Analysis

The qualitative and quantitative data that is collected must undergo separate, but complementary analyses. The analysis of *qualitative data* will consist of four components:

1) data reduction;
2) displaying data;
3) drawing conclusions; and
4) verification through data triangulation. Qualitative data should undergo analysis using a coding system to be developed by the team's Statistician/Data Expert.

The IE Contractor will utilize a variety of techniques, including computer-based tools to draw conclusions from the data such as noting patterns, themes, and relations between variables, assessing plausibility, and uncovering intervening variables. The consultant will protect against bias by testing explanations, examining exceptions, and confirming findings. *Quantitative data* from the survey must be reviewed for missing information and when possible corrected. The data must be cleaned and inputted into SPSS, CSPro or similar statistical program to begin analysis.

e. Consultation with Key Stakeholders

The IE Contractor will consult with key education stakeholders throughout the evaluation process and create opportunities for input and information sharing. Primary stakeholders include MINED, local education institutions, civil society, and education cooperating partners donors. Transparent and consistent communication with key stakeholders will be critical for building interest and momentum around the IE findings to ignite higher level policy changes and inform GOM and donor resource allocation decision making, especially as it relates to scaling up early grade reading interventions. USAID/Mozambique will provide a list of key stakeholders with contact information to the IE Contractor prior to the commencement of the IE.

f. Stay Results-Focused

The IE Contractor should remain cognizant at all times and during program planning that the demonstration of concrete results at all levels is important in building critical support for educational reform related to changes in reading assessment, evaluation and improving early grade reading outcomes.

C.5 IMPLEMENTATION AND MANAGEMENT PLAN - IMPACT EVALUATION TIMELINE

1. Coordinated IE and EGRA+QIM Implementation Timeframe

The IE will be conducted in close collaboration with the EGRA and School Management Assessment[1] timeline coinciding with the first two academic years of the EGRA+QIM program, 2013 and 2014, as follows in the table below:

Timeframe	Key Tasks	Impact Evaluator	EGRA+QIM Implementer
July – December 2012	Project Start-up	• Recruit and train locally hired staff • Consult USAID, EGRA+QIM Contractor, and MINED	• Recruit and train locally hired staff • Consult USAID, IE Contractor and MINED • Orient and train local education institution personnel to participate in relevant aspects of EGRA+QIM
Aug 2012	Baseline Survey and IE/RCT Design	• Define "universe" • Conduct baseline survey of demographic and relevant information • Develops overall IE/RCT design and instruments; trains local staff	• Coordinate with IE Contractor as necessary • Develops EGRA instrument and shares with IE Contractor; trains local staff
September 2012	Pilot	• Pilots IE/RCT instruments and makes adjustments as needed for final version • Continues staff training	• Pilots EGRA assessment and shares results with IE Contractor • Prepares final version

[1] Henceforth in this document, the term "EGRA Assessment" refers to a hybrid assessment which includes an *Early Grade Reading Assessment* component coupled with a *School Management Assessment* component.

Timeframe	Key Tasks	Impact Evaluator	EGRA+QIM Implementer
			of instrument • Trains local education institution personnel and local enumerators to conduct EGRA assessments
January 2013	RCT initiated in 180 schools	• Conducts IE/RCT: Accompanies EGRA+QIM staff to conduct EGRA assessments across 180 schools, assuring quality data collection • Uses other qualitative methods to gather relevant information • Analyzes results • Establishes baseline	• In collaboration with IE, conducts EGRA assessments in 180 schools • Trains local education institution personnel to conduct and manage EGRA assessments
January – September 2013	Conduct EGRA+QIM Interventions	• Available for (virtual) consultation	• Conducts teacher and school director training, coaching, and support in initial 120 "medium" and "full" treatment schools
September 2013	RCT completed in 180 schools for initial academic year	• Conducts IE/RCT: Accompanies EGRA+QIM staff to conduct EGRA assessments across 180 schools, assuring quality data collection • Uses other qualitative methods to gather relevant information • Analyzes results	• In collaboration with IE, conducts EGRA assessments in 180 schools • Trains local education institution personnel to conduct and manage EGRA assessments
October 2013	Data Analysis and Final Year 1 IE Report	• Analyzes IE/RCT data • Presents findings, conclusions and recommendations including most cost-effective interventions to bring to scale in subsequent academic year	• Provides EGRA assessment, progress reports and other information to IE Contractor on demand
November 2013	Dissemination of Year 1 Results	• Disseminate results to key stakeholders: USAID, MINED, local education institutions, civil society, cooperating partners (see Attachment J.1 for more	• Participate in dissemination activities

Timeframe	Key Tasks	Impact Evaluator	EGRA+QIM Implementer
		detail)	
September 2014	RCT completed in 180 original schools for second academic year	• Conducts IE/RCT: Accompanies EGRA+QIM staff to conduct EGRA assessments across 180 schools, assuring quality data collection • Uses other qualitative methods to gather relevant information • Analyzes results	• In collaboration with IE, conducts EGRA assessments in 180 schools • Trains local education institution personnel to conduct and manage EGRA assessments. • Conducts EGRA assessments for performance monitoring and evaluation purposes for 480 additional schools external to impact evaluation effort
October 2014	Data Analysis and Final IE Report	• Analyzes IE/RCT data • Presents findings, conclusions and recommendations including most cost-effective interventions to bring to scale in later academic years	• Provides EGRA assessment, progress reports and other information to IE Contractor on demand
November 2014	Dissemination of Final Report and Results	• Disseminate results to key stakeholders: USAID, MINED, local education institutions, civil society, cooperating partners (see Attachment J.1 for more detail)	• Participate in dissemination activities
December 2014	IE Close out	• Completes all pending tasks and closes out all operations related to IE	• Continues with scheduled EGRA+QIM program activity

2. Personnel and Logistics

A. Staffing

In order to successfully conduct the IE, the Contractor will field a highly qualified and highly motivated team that will best accomplish the IE objectives. The Contractor will partner with a local indigenous NGO, Mozambican university, or private sector company to recruit, train, and manage locally hired staff for IE activities.

The Contractor will have a maximum of four (4) key personnel. At least one (1) of the key personnel is a Mozambican citizen or permanent resident of Mozambique.

Key personnel on the evaluation team is comprised of a mixture of international and local experts to ensure that the necessary technical skills for designing and running a rigorous education impact evaluation, as well as the necessary country knowledge and experience, are covered. Key personnel at minimum include an Evaluation Team Leader, and a Statistician/Data Specialist.

The IE Contractor will ensure that additional personnel who are Mozambican citizens or permanent residents possess qualifications to cover the following technical areas:

- Education impact evaluation skills and experience
- Experience in developing country and Mozambique context
- Survey, sampling, and statistical skills
- Early grade reading assessment
- School management assessment
- Financial and cost-benefit analysis
- Scheduling and Logistics
- Superior written and oral communication skills in English and Portuguese are essential.

B. Key Personnel

A maximum of four (4) key personnel was proposed. The following two positions are required:

1. Magdala Raupp **- Evaluation Team Leader**

The Evaluation Team Leader is responsible for overall management of the impact evaluation and provides overall technical leadership support for the IE. She is the primary liaison with USAID/ Mozambique, MINED, EGRA+QIM implementing partner, and all participating local institutions and key stakeholders.

Required qualifications include:

- Advanced degree (Masters/PhD) in evaluation with an emphasis on education evaluation, policy and planning;
- Minimum 15 years' experience and expertise leading, supervising and managing education evaluation teams, including managing impact evaluations in the education sector; at least 10 years of this experience in developing countries.
- Ability to work with various counterparts, implementing partners, and host country government stakeholders;
- Ability to travel to remote and challenging areas to conduct evaluation activities and provide technical expertise;
- Excellent interpersonal skills and team work;
- Superior written and oral communication skills in English and Portuguese;
- Strong computer skills.

2. Bruce Newman **- Statistician/Data Specialist**

The Statistician/Data Specialist is responsible for the overall survey design, including sampling design and the actual conduct of the various survey rounds, including training and oversight of the survey staff, i.e. enumerators, data entry clerks and supervisors. The Statistician/Data Specialist is also responsible for the statistical data analysis programs.

Required qualifications include:

- Advance degree (Masters/PhD) in statistics, Evaluation, Monitoring and Evaluation (M&E) or related field;
- Minimum 15 years' experience in qualitative and quantitative data collection and analysis methods and in designing education evaluations, and at least 7 of these years in a developing country context;
- Minimum 10 years' experience in designing and leading education national surveys, including expert knowledge of state-of-the-art sampling or census methods;
- Minimum 10 years' experience in running statistical analysis programs;
- Ability to work with various counterparts, implementing partners, and host country government stakeholders;
- Ability to travel to remote and challenging areas to conduct data collection and analysis activities;
- Excellent interpersonal skills and team work;
- Superior written and oral communication skills in English and Portuguese;
- Strong computer skills.

3. Luis Reves **- Deputy Team Leader**

4. Assane Pinto **- Data Specialist**

C. Non-Key Personnel

RCT Survey Staff and Research Assistants

In addition to the above key personnel, the IE Contractor is expected to hire and manage the following:

- **Mozambican data entry clerks** to ensure that collected data is entered into statistical databases.

- **Mozambique supervisors** to oversee and maintain quality standards during the data collection and entry process. The supervisors are essential as they provide oversight of the various teams of enumerators who are contracted by the EGRA+QIM project during the data collection process in the field as well as manage the data entry clerks during the data entry process.

- **Mozambican research assistants**, to assist in the collection of qualitative data using qualitative methods, such as focus groups and key informant interviews.

The enumerators responsible for conducting the actual EGRA assessments shall not be hired and managed by the IE Contractor, but instead, by the EGRA+QIM Contractor. IE Contractor supervisors shall provide overall quality control oversight of these enumerators during the various rounds of IE data collection and data entry.

D. Logistics

For purposes of the IE/RCT, a total of three EGRA and School Management Assessments (henceforth, referred to as EGRA Assessment) shall be administered in 180 schools by the EGRA+QIM Contractor: Baseline assessment in Jan 2013; Post academic year 1 assessment

in Sept 2013; and, Post academic year 2 assessment in Sept 2014. Logistical guidelines (to be negotiated upon award) for carrying out the IE/RCT shall be as follows:

1. **Hiring and Training:** EGRA+QIM Contractor shall be responsible for hiring all EGRA assessment field staff (enumerators and field managers). Ten (10) teams shall be hired with each team consisting of three (3) enumerators: Two (2) enumerators to conduct the EGRA reading portion of the assessment, and a separate enumerator to conduct data collection on school management. The EGRA+QIM Contractor shall provide training to the enumerator teams, along with local education institution counterparts.

2. **Deployment and Data Collection:** All EGRA assessment teams shall deploy simultaneously and work over a period of one month (21 work days) in the field conducting assessments in 180 schools (one work day per school). In addition to the EGRA+QIM Contractor's supervision of each team, the IE Contractor shall be responsible for providing additional support to ensure the quality of data collected. The EGRA+QIM Contractor shall collaborate as requested by the IE Contractor to implement safeguards to maintain data quality in the collection process.

3. **Data Entry:** Data entry clerks and supervisors for the IE shall be the responsibility of the IE Contractor. The IE Contractor shall train up to ten (10) data entry clerks. Data entry shall begin within the first week of the commencement of data collection.

4. **Dissemination of IE Results:** In collaboration with USAID, MINED, and local education institutions, the IE Contractor shall be responsible for organizing and facilitating in November of each academic year 2013 and 2014, three "EGRA Assessment Results Dissemination Conferences:" One in Maputo, and one in each of the target provinces, Zambezia and Nampula. The EGRA Assessment Results Dissemination Conferences shall be conducted in Portuguese, and include the participation of key education stakeholders: MINED, local education institutions, leaders from the target areas, civil society, education cooperating partners, and USG agencies. Copies of each periodic report shall be made available to dissemination event participants.

References:
USAID/Mozambique Country Assistance Strategy 2009-2014
USAID Education Strategy
USAID EQUIP 2 Aga Khan Foundation Mozambique Case Study
EGRA+QIM Project RFP

[END OF SECTION C]

ATTACHMENT 1

Revised: July 28, 2013

C.5 IMPLEMENTATION AND MANAGEMENT PLAN – IMPACT EVALUATION TIMELINE

1. Coordinated APAL and APAL IE Implementation Timeframe

The IE will be conducted in close collaboration with the EGRA and School Management Assessment[1] timeline coinciding with the first two academic years of the USAID/Aprender a Ler program, 2013 and 2014, as follows in the table below:

Timeframe	Key Tasks	Impact Evaluator of USAID/Aprender a Ler	USAID/Aprender a Ler
July – December 2012	Project Start-up	• Consult with USAID, USAID/Aprender a Ler project staff and MINED • Coordinate with local partner and define roles and responsibilities	• Recruit and train locally hired staff • Consult USAID, IE Contractor and MINED • Orient and train local education institution personnel to participate in relevant aspects of *Aprender a Ler*
September – October 2012 * *Deliverables Oct 1st: Inception Report and Communication and User Engagement Plan*	IE/RCT Design	• Define sampling frame • Finalize sampling plan • Identify relevant information needed and sources where data can be obtained • Prepare Inception Report and Communication and User Engagement Plan	• Coordinate with IE Contractor as necessary • Provide IE Contractor a detailed work plan for 2013, including an implementation plan at the ZIP and school level • Provide IE Contractor target districts by October 15, 2012 and all target ZIPs by October 30, 2012.
Oct—Nov, 2012 *Deliverable November 15: Report on Survey for Sampling*	Conduct Survey to gather data on districts and ZIPs selected by WEI (Selection of districts completed by October 15, selection of ZIPs completed by October 31)	• Survey data available at MINED and at the provincial/district levels • Based on the information gathered by the survey recommend ZIPs in each province where WEI could implement the USAID/Aprender a Ler project in accordance with the RCT model • Select, identify additional	• Use ZIP data report provided by IE Contractor to select the targets ZIPs in consultation with DPEC and inform IE Contractor of the selection. • Adapt and pilot EGRA +School Management Assessment (SMA) instrument and share with IE contractor (November) • Recruit EGRA + SMA enumerators and field

[1] Henceforth in this document, the term "EGRA Assessment" refers to a hybrid assessment which includes an *Early Grade Reading Assessment* component coupled with a *School Management Assessment* component.

Timeframe	Key Tasks	Impact Evaluator of USAID/Aprender a Ler	USAID/Aprender a Ler
		data sources • Develop and pilot IE instruments for data collection at the school level (quantitative and qualitative data) • Recruit and train supervisors to ensure data quality and to collect qualitative data at the province/district/school levels	managers
Dec 2012–Jan, 2013 ** Deliverable: Final versions of data collection instruments*	Refine detailed plans/procedures to be followed for RCT Baseline Study	• Finalize and prepare IE instruments for data collection at the school level (quantitative and qualitative data) • Train supervisors to ensure data quality of reading test results and to collect qualitative and quantitative data at the province/district/school level • Pilot logistic arrangements related to supervisors responsibilities • Coordinate training with USAID/Aprender a Ler	• Finalize and prepare EGRA + SMA instrument for data collection at the school level. • Train local education institution personnel, local EGRA + SMA enumerators and field managers to conduct EGRA + SMA assessments • Coordinate training with IE Contractor
February 2013	Conduct RCT Baseline Study in 180 schools	• Accompany USAID/Aprender a Ler enumerators to schools in order to ensure data quality • Collect qualitative data • Enter and analyze data	• In collaboration with IE, conduct EGRA + SMA assessments in 180 schools • Train local education institution personnel to manage EGRA assessments
March, 2013	Prepare RCT Baseline Study Report	• Enter and analyze data and write baseline report • Provide USAID/Aprender a Ler with EGRA data in electronic form	• Available for consultation • Provide IE results of analyses of SMA data
March – September 2013 ** Deliverable:*	Conduct USAID/Aprender a Ler interventions	• Available for virtual consultation • Review SMA classroom observation instrument submitted on Aug 3, and	• Conduct teacher and school director training, coaching, and support in initial 120 "medium" and "full" treatment schools.

Timeframe	Key Tasks	Impact Evaluator of USAID/Aprender a Ler	USAID/Aprender a Ler
Report on results of Baseline Study (May 3rd)		comment per quality assurance role by Aug 10). • Review SMA field test data received from APAL on Aug 28, and comment on quality to APAL by Sept 2.	• Revise SMA classroom observation instrument (Submit Aug 3). • IE comments received Aug 10 considered/ integrated into revised SMA by Aug 17. • Field test SMA Aug 20-22 in 8-10 schools. • Field test data entered and analyzed (Aug 23-27). • Submit field test data to APAL IE by Aug 28.
September 2013	Conduct post intervention RCT in 180 schools (Collect Sept 16 to Oct 11)	• Retrain supervisors (Sept 9-13). • Conduct second round of data collection	• Make final SMA instrument revisions, formatting and copying, and make available at training venue by Sept 6. • Retrain EGRA + SMA enumerators and field managers to conduct EGRA assessments (Sept 9-13). • In collaboration with IE, conduct EGRA + SMA assessments in 180 schools • Local education institution personnel capacity building activities to be determined
October 2013	Continue RCT; enter/analyze data.	• Enter and analyze EGRA data. • Provide USAID/Aprender a Ler EGRA data in electronic form (Oct 25).	• Provide EGRA+SMA assessment data, quarterly implementation progress reports, Monitoring and Evaluation (M&E) Plan data, and other information to IE Contractor as requested. • Provide SMA data in electronic form to APAL IE (Oct 25).
November 2013 ** Deliverable: Year 1 PPT (November 15); and Draft IE Report(November 29h)*	Prepare Year 1 IE Report Dissemination of Year 1 Results (tentative)	• Present findings, conclusions and recommendations (PPT to USAID and MINED on Nov 15) • Submit draft Year 1 IE report (English and Portuguese) on Nov 29. • Disseminate results to key stakeholders: USAID, MINED, local education	• Provide APAL IE results of analyses of SMA data (November 8). • Participate in dissemination activities • Provide comments on draft Year 1 IE report

Timeframe	Key Tasks	Impact Evaluator of USAID/Aprender a Ler	USAID/Aprender a Ler
		institutions, civil society, cooperating partners by Dec 20. (dissemination activities pending demonstrable progress that is both statistically significant and of practical educational significance).	
December 2013 ** Deliverable: Final Year 1 IE Report (December 23h*	Final Year 1 IE Report	• Submit Final Year 1 IE report in English and Portuguese by Dec 23.	• Available for consultation • Provide APAL IE available data as requested
September 2014	Conduct post 2014 intervention RCT in 180 original schools. (Collect Sept 15 to Oct 10)	• Retrain supervisors (Sept 8-12). • Conduct second round of data collection	• Retrain EGRA + SMA enumerators and field managers to conduct EGRA assessments (Sept 8-12). • In collaboration with APAL IE, conduct EGRA + SMA assessments in 180 schools. • Local education institution personnel capacity building activities to be determined
October 2014	Continue RCT; enter/analyze data.	• Enter and analyze EGRA data. • Provide USAID/Aprender a Ler EGRA data in electronic form (Oct 24).	• Provide EGRA+SMA assessment data, quarterly implementation progress reports, Monitoring and Evaluation (M&E) Plan data, and other information to IE Contractor as requested. • Provide SMA data in electronic form to APAL IE (Oct 24).
November 2014 ** Deliverable:Y ear 2 PPT (November 14th) and Draft IE Report (November 28th)*	Prepare Year 2 IE Report Dissemination of Year 2 Results	• Present findings, conclusions and recommendations (PPT to USAID and MINED on Nov 14) • Submit draft Year 2 IE report (English and Portuguese) on Nov 28. • Disseminate results to key stakeholders: USAID, MINED, local education institutions, civil society, cooperating partners (Dec	• Provide IE results of analyses of SMA data (November 7). • Participate in dissemination activities • Provide comments on draft Year 2 IE report

Timeframe	Key Tasks	Impact Evaluator of USAID/Aprender a Ler	USAID/Aprender a Ler
		19).	
December 2014 * *Deliverable: Final Year 2 IE Report (December 22nd)*	Final Year 2 IE Report.	• Submit final Year 2 IE report in English and Portuguese (Dec 22).	• Provide support to MINED to continue with scheduled *Aprender a Ler* program activity • Available for consultation • Provide APAL IE available data as requested.

[end of Attachment 1]

Annex A.2. Scope of Work for the Follow-On Contract

SECTION C – DESCRIPTION /SPECIFICATIONS/ STATEMENT OF WORK

1. Background

In July 2012, USAID/Mozambique awarded International Business & Technical Consultants, Inc., (IBTCI) Task Order AID-656-TO-12-00002 under the Evaluation Services IQC to conduct the Impact Evaluation (IE) for the USAID/*Aprender a Ler* (ApaL) Project. *Aprender a Ler* is a four-year USAID-funded project designed to improve student reading outcomes in grades 2 and 3 in selected schools in Nampula and Zambézia Provinces in Mozambique. Managed by World Education, Inc. (WEI), *Aprender a Ler* works closely with the Ministry of Education in Mozambique (MINED) to improve the quality and increase the quantity of reading instruction. The project includes teacher training, coaching and support in improved reading instruction methods, and school management training, coaching and support for school directors. These elements will be implemented in a holistic approach designed to improve reading outcomes in the early grades. The IE is being conducted by IBTCI and its Mozambican partner, Global Surveys Corporation (GSCResearch) in close collaboration with the USAID mission, WEI (implementers of the project), and national, provincial, district MINED offices, as well as teacher training institutes in the target provinces. The IE entails conducting qualitative and quantitative analysis of 180 schools along the economic corridors of the two Provinces using Randomized Control Trial (RCT) methodology with 60 randomly selected schools receiving "Full" treatment, 60 schools receiving "Medium" treatment, and 60 no-treatment "Control" schools.

The use of RCT methodology is the most effective way to measure impact of a project or program for three main reasons. First, it allows for direct attribution of the *Aprender a Ler* interventions to improve outcomes because the RCT model controls for all other possible determinants of the outcomes. Second, the random sampling component of RCT eliminates the effects of potential unobservable differences between treatment and control groups on the outcomes. Third, an RCT is the most rigorous evaluation method to obtain accurate and valid results to inform plans to scale up the most effective and cost-effective interventions.

2. Target Geographical Regions and Beneficiaries

The ApaL program and the Impact Evaluation are targeting the provinces of Nampula and Zambezia in central and northern Mozambique. The rationale for selecting these two provinces can be summarized as follows: These two provinces combined contain 42 percent of the school-age population of Mozambique and according to national statistics, these heavily populated and rural provinces have posted weak education performance results compared to national averages. In addition, girls in these two provinces are grossly underserved and disadvantaged. While the national illiteracy rate for women is 63 percent, in Nampula it is 77 percent, and in Zambezia 79 percent (Multiple Indicator Cluster Survey (MICS), 2008). Furthermore, despite clear and pressing needs, there is relatively little bilateral or multilateral support for education in these provinces, with the exception of UNICEF which has been working in only one district in each province.

3. Purpose of IE and Rationale for Extension

The Government of Mozambique and USAID need rigorously generated evidence about the effectiveness of education interventions to help them make informed decisions about the allocation of scarce resources in the struggle to improve early grade reading outcomes. In Mozambique, the government is poised to take pro-active steps towards establishing quality standards in basic AID-656-C-15-00002 education, but is seeking the necessary assessment instruments and skills to effectively measure learning outcomes. This IE is the first external and independent impact evaluation of its kind in USAID Africa. The IE will provide a thorough understanding of the benefits of the ApaL program as it relates to improving early grade reading outcomes.

The IE utilizes a randomized control trial (RCT) methodology, in line with the new USAID evaluation policy and general best practice in measuring causal impacts. A carefully designed RCT can provide the rigorous evidence needed to advocate effectively with the government and donor agencies about the best way to invest scarce resources to improve reading outcomes in the early grades. Given the lack of rigorous external and independent evaluation evidence on ApaL type interventions in USAID Africa, this IE has the potential to raise awareness of the importance of such interventions as well as the importance of conducting well thought out impact evaluations that can relate interventions to student outcomes.

Best practice for IEs using RCT design calls for data collection at baseline prior to implementation of project interventions, midline(s) as appropriate, and end-line at a predetermined time that provides for at least one full cycle of project interventions in order to effectively measure the impact of specified program interventions throughout the lifetime of the program. While baseline data were collected near the beginning of the 2013 school year, WEI was not able to begin implementation of ApaL in the schools until well after the beginning of classes, and the Midline EGRA measurement conducted in September 2013 reflects student reading outcomes after approximately three months of project intervention. Implementation and measurement for a full academic year of ApaL intervention was necessarily postponed to the 2014 school year.

4. Development Hypothesis

The IE is testing USAID/Mozambique's development hypothesis that reading outcomes in grades 2 and 3 will improve if the quality and quantity of reading instruction in those grades is improved through better teacher training and coaching, and will improve more with strengthened school management to support the increased time on task and the teaching-learning process.

5. Objectives

The main objective of the IE remains to measure the causal effect of ApaL program treatment interventions on improving early grade reading outcomes in grades 2 and 3 in targeted schools. The evaluation is testing two treatment interventions against a control group that will not receive any of the interventions. One treatment intervention, the medium treatment model, includes training, coaching, classroom materials and support in improved reading instruction methodology for teachers. The second treatment intervention, the full treatment model, includes training, coaching and support in improved reading instruction methodology for teachers with additional school management training, coaching and support for school directors. The main objective of the IE for 2015 is to determine the sustainability of the ApaL interventions, i.e.

whether ApaL activities implemented at 120 project schools have been sustained without the presence of assistance from ApaL.

6. Impact Evaluation Questions

The main evaluation questions to be addressed by the Impact Evaluation (IE) for the third year extension remain the same as previous years:

To what extent have USAID/Aprender a Ler treatment interventions improved early grade reading outcomes for students in second and third grades in the target schools in the Nampula and Zambézia Provinces?

From this general guiding question flows a set of focused questions to be answered by the Impact Evaluation.

Question 1: To what extent does the "reading instruction support" treatment intervention cause early grade reading outcomes to improve for students in grades 2 and 3 in target schools whose teachers have received training, coaching and support?

Question 2: To what extent does the treatment intervention of additional "school management" training, coaching and support to school directors cause a significant and additional improvement in early grade reading outcomes when coupled with "reading instruction support" in target schools?

Question 3: Cost-effectiveness - To what extent are the "Medium" and "Full" treatment interventions cost-effective? Specifically, what are the most significant reading outcome effects and unit costs per student, per teacher, per school director, per school of the key treatment interventions?

Question 4: Management Sustainability – Of the most cost-effective interventions, which fall within the existing technical and financial management capacity of local education institution personnel? What capacity-building activities would be required to ensure sufficient MINED technical and financial management capacity to implement the interventions?

The original Inception Report submitted by the Impact Evaluation Contractor specified in detail how each of the evaluation questions was to be addressed and answered. The focus of the IE for 2015 will be to determine the sustainability of the ApaL intervention: Have gains observed in 2014 as a result of the intervention persisted or disappeared?

Responses to Questions 1, 2, and 3 will be updated and refined to incorporate the findings from the 2014 endline data collection. There are two parts to Question 4. The first requires that the IE determine whether gains (by students and teachers) observed in 2014 as a result of the intervention have persisted or disappeared. That means comparing gains obtained by students and teachers in 2014 to those observed in 2015 and providing a summary of what was previously observed since Baseline. The second part of the question entails a determination of the existing technical and financial management capacity of ZIP coordinators and school directors and a determining the capacity building activities necessary to ensure that activities introduced by the ApaL intervention—which have shown to benefit students—will continue to be implemented.

The third year extension will utilize the methodology and approach stipulated in an updated Inception Report to answer Evaluation Question 4.
It is important to emphasize that the IE RCT model requires that IBTCI apply a mixed method approach where both qualitative and quantitative information is collected and analyzed in order to properly address and answer the above questions. The Final IE Report as with the previous Midline
Reports is thus expected to correlate and integrate EGRA data with relevant qualitative data generated from the ApaL M&E system such as the SMA tools (e.g. the Classroom Observation tool and the School Management tool) as well as qualitative information collected by the IE supervisors. These qualitative tools provide essential information for interpreting and explaining EGRA scores along the various treatment groups.

7. Methodology and Approach.

The original Inception Report submitted by the Impact Evaluation Contractor specified in detail how each of the evaluation questions is to be addressed and answered. The third year extension will utilize that methodology and approach to answer the evaluation questions as proposed in the Inception Report.

The IE RCT model entails a mixed method approach where both qualitative and quantitative information is collected and analyzed in order to address the above questions. As has been the case with the Baseline and the 2013 and 2014 Midline reports, the Final IE Report will integrate data from various sources, primarily (1) Student scores on the EGRA; (2) Student and teacher interviews; (3) relevant quantitative and qualitative data generated and analyzed by the ApaL M&E system such as data collected using the SMA tools (e.g. Classroom Observation tool and the School Management tool); and (4) qualitative information collected by the IE supervisors. These tools provide essential information for interpreting and explaining EGRA scores across the Full Treatment, Medium Treatment, and Control groups.

The Impact Evaluation uses a Randomized Control Trial (RCT) design which implies that participating entities have been randomly assigned to either a treatment (intervention) group or to a control group. In the specific case of USAID/ApaL, the IE will evaluate the impact of one of two treatments on the reading performance of students in second and third grade, relative to that of students in Control schools. The RCT allows for direct attribution of the ApaL intervention to improved learning outcomes because the model controls for other possible determinant s of the outcome. The reading scores obtained through the EGRA tool pre- and post-intervention will be compared to results obtained from the non-intervention Control group; the Control group reflects the probable performance of students in the Medium or Full treatment groups absent the relevant interventions.

The evaluation design is based upon a number of data collection events for all groups: the February/March 2013 baseline; the October 2013 midline; the September 2014 end-line; and the September 2015 data collection after the ApaL intervention is no longer being implemented at the schools that were selected to be part of the treatment groups. The ApaL project has been providing training, coaching and other reading instruction support (“Medium” treatment) to 30 schools in
Nampula Province and 30 schools in Zambézia Province during part of the 2013 school year and during the 2014 school year. In addition to the Medium treatment activities, the project has been

providing a school management improvement intervention to school directors ("Full" treatment) in the same number of schools in each Province. These schools will receive no USAID-assisted interventions in the 2015 school year. During the 2013, 2014 or 2015 school years, the Control schools (30 schools of comparable demographics in each Province) did not and will not receive the interventions previously mentioned.

The quantitative data that are collected through the EGRA administration will answer questions such as *who* was involved (second and third grade students), *what* was done (characteristics of the ApaL program), *where* (180 schools in two provinces), and *how much (*scores obtained and gain in scores). This will be complemented with qualitative data in order to answer "why" and "how" questions.

Collecting and analyzing qualitative data usually involves talking to and observing teachers, classrooms and school directors. It is important that the ApaL M&E analyses of the qualitative data be incorporated in the Final IE Report produced by IBTCI to explain EGRA results. USAID is not requesting two separate reports, but one Final IE report integrating qualitative and quantitative data generated through the data collection process. This requires strong coordination and communication between IBTCI and the ApaL implementer. Both parties are expected to facilitate the necessary flow of information sharing and communication, but it is the overall responsibility of IBTCI to promote and support coordination and quality assurance for the impact evaluation including logistical planning, sharing of necessary information and general communications.

IBTCI will ensure that USAID evaluation policy is adhered to and that rigorous impact evaluation standards are maintained.

Specifically it is the overall responsibility of IBTCI to:

1. Provide overall oversight and coordination of the IE process, including in terms of ensuring necessary communication and information sharing between IBTCI and implementer;

2. Provide quality assurance of all qualitative and quantitative data collection instruments for the mid-line data collection process as well as for the end-line data collection process;

3. Provide coordination, technical assistance and overall guidance and quality assurance of the planning process leading up to the end-line data collection, i.e. in terms of logistics plan and training and retraining of local enumerators and supervisors;

4. In close collaboration with the implementer to collect, process, exchange and analyze quantitative and qualitative data during the end-line data collection necessary to answer all of the evaluation questions;

5. Provide USAID with a Final IE Report covering the entire RCT period, from 2012 to the conclusion of the Endline data collection and analysis, and incorporating responses to all of the evaluation questions;

8. 2015 – 2016 Evaluation Work Plan.

The Evaluation Work plan for 2015 will for the most part replicate the 2014 schedule but the emphasis will be on measuring sustainability. The work plan schedule may change slightly as it has to accommodate the APAL implementation schedule as well as any potential changes to the school calendar made by the Mozambican authorities.

Timeframe	Key Tasks	Impact Evaluator	EGRA+SMA Implementer
July-August 2015	**Logisitics preparation and retraining for End-line Data Collection**	• Fine tune EGRA IE tools and advise on fien tuning of SMA IE tools • Retrain IE supervisors • Coordinate with WEI staff on logistics plan for end-line data collection • Participate in (re)training of enumerators and WEI supervisors • Ensure coordination and communication with WEI and USAID necessary for successful training and planning for End-line Data Collection	• Recruit and (re)train EGRA+SMA enumerators • Fine tune SMA IE tools • Develop End-line data collection logistics plan in collaboration with Impact Evaluator • Local education institution personnel capacity building activities conducted (TBD)
September 2015	**End line data collection (Data collection to start on or about August 31)**	• Conducts IE/RCT Accompanies EGRA+SMA staff to conduct EGRA assessments across 180 schools, assuring quality and data collection	• In collaboration with IE conducts EGRA assessments in 180 schools • Conducts EGRA assessments for performance monitoring and evaluation purposes for 300

		• Collects qualitative and contextual information (SD interviews) to gather relevant information on the sustainability of the intervention • Enters, cleans and processes EGRA data and resource inventory data	additional schools external to impact evaluation effort • Conduct SMA in 180 schools • Provide IE with completed QA'ed EGRA forms
October 2015	**Data Analysis and exchange**	• Coordinates and facilitates necessary communication and data exchange with WEI to enable analysis of EGRA+SMA IE data • Provide USAID/Aprender a Ler EGRA data in electronic form on Oct. 18 2015 • Analyzes results of the EGRA and integrates SMA results and analyses as provided by WEI, and other sources including qualitative data collected by IE supervisors, into one cohesive Draft IE Report that describes findings and provides	• Provide qualitative and contextual information (SMA relevant program monitoring data) to Impact Evaluators as requested by IBTCI • Provide IBTCI with electronic version of cleaned SMA dataset by October 18 • Available for consultation to provide feedback and general contextual information based on data collection fieldwork to IE team

		explanation for the EGRA scores.	
November to December 2015	**Presentation of preliminary findings and submit Draft IE Report**	• Present findings, conclusions and recommendations (PPT to USAID MINED on November 13 2015) • Submit Draft IE report to USAID (English and Portuguese versions) on November 20, 2015	Provide support to MINED to continue with scheduled Aprender a Ler program activity.

9. 2015-2016 Deliverables

Deliverables	**Due Date**
Draft IE Report with summary of data collected findings, results, conclusions and recommendations (USAID/Mozambique Evaluation Manager approval required)	November 20, 2015
Draft IE Report with summary of data collected, findings, results, conclusions and recommendations (USAID/Mozambique Evaluation Manager approval required)	Submit Draft Final English language IE Report to USAID by December 14 2015, and the final Portuguese language report within 10 working days following receipt of comments.
Dissemination events for key stakeholders – In-country PowerPoint presentation of preliminary findings to USAID/Mozambique and MINED partners (Education Donors, Civil Society, Private Sector, and U.S. Government Agencies). The IE evaluator will present preliminary findings from data collection and analysis to USAID/Mozambique and Mined partners. (USAID/Mozambique Evaluation Manager approval required) **1 PPT presentation to MINED** The IE evaluator will present preliminary findings from data collection and anaylsis to MINED **1 PPT presentation to USAID/Mozmbique** The evaluator will present preliminary findings from data collection and analysis to USAID/Mozambique	December, 21, 2015 (Exact dates TBD)

Final IE Report with summary of data collected, findings, results, conclusions, and recommendations. (USAID/Mozambique Evaluation Manager approval required)	Submit Final English language IE report to USAID by January 16, 2016, and the final Portuguese language report within 10 working days following receipt of comments.

10. GUIDING PRINCIPLES

The following are guiding principles that will help in the successful implementation of this evaluation. The Contractor should review the themes listed below carefully, to ensure they are considered throughout the evaluation.

a. USAID Forward

IBTCI will adhere to key reform initiatives described under USAID Forward which emphasizes new partnerships, local capacity building, innovation, and rigorous monitoring and evaluation to achieve results. The Contractor should focus on ways to strengthen host country systems and build local technical and managerial capacity to ensure sustainability.

b. USAID Evaluation Policy

IBTCI should ensure that the Impact Evaluation follows the USAID Evaluation Policy requirements for rigorous impact evaluations. In addition, the IE should lead to more focused and collaborative education investments aimed at identifying low unit costs and at improving learning outcomes and institutional sustainability in the host country.

c. Data Quality Standards

IBTCI must ensure that the Impact Evaluation adheres to USAID‘s requirements for data quality. USAID data quality standards are detailed in Automated Directives System (ADS) 578 and ADS 203, which will be provided to IBTCI.

d. Data Analysis

The qualitative and quantitative data that is collected must undergo separate, but complementary analyses. The analysis of qualitative data will consist of four components:

1) data reduction;
2) displaying data;
3) drawing conclusions; and
4) verification through data triangulation.

Qualitative data should undergo analysis using a coding system to be developed by the team‘s Statistician/Data Expert. The IBTCI will utilize a variety of techniques, including computer-based tools to draw conclusions from the data such as noting patterns, themes, and relations between variables, assessing plausibility, and uncovering intervening variables. The consultant will protect against bias by testing explanations, examining exceptions, and confirming findings. Quantitative data from the EGRA administration must be reviewed for missing information and when possible corrected. The data must be cleaned and inputted into SPSS, CSPro or similar statistical program to begin analysis.

e. Consultation with Key Stakeholders

IBTCI will consult with key education stakeholders throughout the evaluation process and create opportunities for input and information sharing. Primary stakeholders include MINED, local education institutions, civil society, and education cooperating partners donors. Transparent and consistent communication with key stakeholders will be critical for building interest and momentum around the IE findings to ignite higher level policy changes and inform GOM and donor resource allocation decision-making, especially as it relates to scaling up early grade reading interventions. USAID/Mozambique will provide a list of key stakeholders with contact information to IBTCI prior to the commencement of the IE.

f. Stay Results-Focused

IBTCI should remain cognizant at all times and during program planning that the demonstration of concrete results at all levels is important in building critical support for educational reform related to changes in reading assessment, evaluation and improving early grade reading outcomes.

11. PERSONNEL AND LOGISTICS

A. Staffing

In order to successfully conduct the IE, a highly qualified and highly motivated team will be required. The Contractor is encouraged to field a team that it believes will best accomplish the IE objectives. The Contractor will be required to partner with a local indigenous NGO, Mozambican university, or private sector company to recruit, train, and manage locally hired staff for IE activities.

A maximum of four (4) key personnel may be proposed. A minimum of one (1) of the key personnel is required to be a Mozambican citizen or permanent resident of Mozambique.

Key personnel on the evaluation team shall be comprised of a mixture of international and local experts to ensure that the necessary technical skills for designing and running a rigorous education impact evaluation, as well as the necessary country knowledge and experience, are covered. Key personnel shall minimally include an Evaluation Team Leader, and a Statistician/Data Specialist.

IBTCI shall ensure that additional personnel who are Mozambican citizens or permanent residents with qualifications to cover the following technical areas are included on the team:

- Education impact evaluation skills and experience
- Experience in developing country and Mozambique context
- Survey, sampling, and statistical skills
- Early grade reading assessment
- School management assessment
- Financial and cost-benefit analysis
- Scheduling and Logistics
- Superior written and oral communication skills in English and Portuguese are essential.

B. Key Personnel

A maximum of four (4) key personnel may be proposed. The following three positions are required:

Article I. Evaluation Team Leader

The Evaluation Team Leader is responsible for overall management of the impact evaluation and provides overall technical leadership support for the IE. S/he is the primary liaison with USAID/Mozambique, MINED, APAL implementing partner, and all participating local institutions and key stakeholders.

Required qualifications include:
• Advanced degree (Master's/PhD) in evaluation with an emphasis on education evaluation, policy and planning;
• Minimum 15 years' experience and expertise leading, supervising and managing education evaluation teams, including managing impact evaluations in the education sector; at least 10 years of this experience in developing countries.
• Ability to work with various counterparts, implementing partners, and host country government stakeholders;
• Ability to travel to remote and challenging areas to conduct evaluation activities and provide technical expertise;
• Excellent interpersonal skills and team work;
• Superior written and oral communication skills in English and Portuguese; and
• Strong computer skills.

Article II Statistician/Data Specialist

The Statistician/Data Specialist is responsible for the overall survey design, including sampling design and the actual conduct of the various survey rounds, including training and oversight of the survey staff, i.e. enumerators, data entry clerks and supervisors. The Statistician/Data Specialist is also responsible for the statistical data analysis programs.

Required qualifications include:
• Advanced degree (Master's/PhD) in statistics, Evaluation, Monitoring and Evaluation (M&E) or related field;
• Minimum 15 years' experience in qualitative and quantitative data collection and analysis methods and in designing education evaluations, and at least seven of these years in a developing country context;
• Minimum 10 years' experience in designing and leading education national surveys, including expert knowledge of state-of-the-art sampling or census methods;
• Minimum 10 years' experience in running statistical analysis programs;
• Ability to work with various counterparts, implementing partners, and host country government stakeholders;
• Ability to travel to remote and challenging areas to conduct data collection and analysis activities;
• Excellent interpersonal skills and team work;
• Superior written and oral communication skills in English and Portuguese; and

• Strong computer skills.

Article III Deputy Team Leader

The Deputy Team Leader is responsible for planning and management of in-country data collection and analysis efforts and for providing routine liaison between the IE implementer, the ApaL implementer, and USAID/Mozambique.

Required qualifications include:
• Citizenship or permanent residency in Mozambique
• Advanced degree in a field related to business administration, marketing, statistics/data analysis, or other relevant discipline
• At least six years' experience conducting large-scale data collection and analysis studies in Mozambique
• At least six years' experience managing and conducting research activities on behalf of Mozambican and/or U.S.G. agencies, which may include large-scale data collection and analysis studies
• Ability to work with various counterparts, implementing partners, and host country government stakeholders;
• Ability to travel to remote and challenging areas to conduct data collection and analysis activities;
• Excellent interpersonal skills and team work;
• Superior written and oral communication skills in English and Portuguese; and
• Strong computer skills.

C. Non-Key Personnel

RCT Survey Staff and Research Assistants

In addition to the above key personnel, IBTCI is expected to hire and manage the following:

• Mozambican data entry clerks to ensure that collected data is entered into statistical databases.

• Mozambique supervisors to oversee and maintain quality standards during the data collection and entry process. The supervisors are essential as they provide oversight of the various teams of enumerators who are contracted by the ApaL project during the data collection process in the field as well as manage the data entry clerks during the data entry process.

• Mozambican research assistants, to assist in the collection of qualitative data using qualitative methods, such as focus groups and key informant interviews.

The enumerators responsible for conducting the actual EGRA assessments shall not be hired and managed by the IBTCI, but instead, by the APAL Contractor. IBTCI supervisors shall provide overall quality control oversight of these enumerators during the various rounds of IE data collection and data entry.

D. Logistics

For purposes of the 2015 IE/RCT, one EGRA and School Management Assessment (henceforth, referred to as EGRA Assessment) shall be administered in 180 schools by IBTCI:

A post academic year 2 assessment in Sept 2015. Logistical guidelines (to be negotiated upon award) for carrying out the IE/RCT shall be as follows:

1. Hiring and Training: The ApaL Contractor shall be responsible for hiring all EGRA assessment field staff (enumerators and field managers). Ten (10) teams shall be hired with each team consisting of three (3) enumerators: Two (2) enumerators to conduct the EGRA reading portion of the assessment, and a separate enumerator to conduct data collection on school management. The ApaL Contractor shall provide training to the enumerator teams, along with local education institution counterparts.

2. Deployment and Data Collection: All EGRA assessment teams shall deploy simultaneously and work over a period of one month (21 work days) in the field conducting assessments in 180 schools (estimated to require one work day per school). In addition to the ApaL Contractor's supervision of each team, IBTCI shall be responsible for providing additional support to ensure the quality of data collected. The ApaL Contractor shall collaborate as requested by IBTCI to implement safeguards to maintain data quality in the collection process.

3. Data Entry: Data entry clerks and supervisors for the IE shall be the responsibility of IBTCI.
IBTCI shall train up to ten (10) data entry clerks. Data entry shall begin within the first week of the commencement of data collection.

4. Dissemination of IE Results: In collaboration with USAID, MINED, and local education institutions, IBTCI shall be responsible for organizing and facilitating in November of each academic year 2013 and 2014, three "EGRA Assessment Results Dissemination Conferences": One in Maputo, and one in each of the target provinces, Zambezia and Nampula. The EGRA Assessment Results Dissemination Conferences shall be conducted in Portuguese, and include the participation of key education stakeholders: MINED, local education institutions, leaders from the target areas, civil society, education cooperating partners, university academics, private sector representatives, and USG agencies. Copies of all IE reports shall be made available to dissemination event participants.

References:

USAID/Mozambique Country Assistance Strategy 2009-2014
USAID Country Development Cooperation Strategy 2014-2018
USAID Education Strategy
USAID EQUIP 2 Aga Khan Foundation Mozambique Case Study
APAL Project RFP
USAID Technical Note on Impact Evaluations, September 2013

Annex B. Reports Submitted for Work Done Under the Task Order

International Business & Technical Consultants, Inc.

8618 WESTWOOD CENTER DRIVE ■ SUITE 400 ■ VIENNA, VA 22182 ■ USA
TEL: 1-703-749-0100 ■ FAX: 1-703-749-0110

Impact Evaluation for the USAID/Aprender a Ler Project in Mozambique AID-656-TO-12-00002

Development Experience Clearinghouse (DEC) REPORT SUBMISSIONS

DELIVERABLE	LINK
Inception Report	http://pdf.usaid.gov/pdf_docs/PA00KB21.pdf
Baseline Report (English)	http://pdf.usaid.gov/pdf_docs/PA00JDZT.pdf
Baseline Report (Portuguese)	http://pdf.usaid.gov/pdf_docs/PA00JDZV.pdf
Midline 1 Report (English)	http://pdf.usaid.gov/pdf_docs/PA00JZK7.pdf
Midline 1 Report (Portuguese)	http://pdf.usaid.gov/pdf_docs/PA00JZK8.pdf
Midline 2 Report (English)	http://pdf.usaid.gov/pdf_docs/PA00KB79.pdf
Midline 2 Report (Portuguese)	http://pdf.usaid.gov/pdf_docs/PA00KB94.pdf
Midline 2 Executive Summary (English)	http://pdf.usaid.gov/pdf_docs/PA00KBD7.pdf
Midline 2 Executive Summary (Portuguese)	http://pdf.usaid.gov/pdf_docs/PA00KBD6.pdf

Annex C. EGRA Instrument (Student Interview)

USAID/Aprender a Ler
EGRA: Questionário para alunos e Instrumento

COMPLETE TODOS OS CAMPOS NA TABELA ABAIXO ANTES DE INICIAR A ENTREVISTA.

Nome do Inquiridor: _____ Data: ____/____/ 2015

Tipo de Tratamento oferecido à escola: Completo ☐ Médio ☐ Controlo ☐

Província: Nampula ☐ Zambézia ☐ Distrito: _____

Nome da Escola: _____ Código da Escola: _____

Nome da Escola Sede da ZIP: _____

Classe: Segunda ☐ Terceira ☐

Número de Identificação do Professor: _____

Secção1. Formulário de Informação sobre os antecedentes do(a) aluno(a)

Diga na **LÍNGUA MATERNA DO ALUNO**: Começarei por fazer algumas perguntas sobre ti, tá bom?

1. Sexo do Aluno Feminino ☐ Masculino ☐

2. Quantos anos tens?_____ Não Sabe/Não Responde ☐

3. Fale em Português: [Marque com 'X' apenas uma resposta por linha]

	Quase Sempre	Ocasionalmente	Quase Nunca	Nunca	Não Aplicável
Com a sua mãe?					
Com o seu pai?					
Com os seus irmãos/irmãs?					
Com os seus amigos?					

4. Vives com a tua mãe? SIM ☐ NÃO ☐

5. Vives com o teu pai? SIM ☐ NÃO ☐

⇨ [Se o(a) aluno(a) vive com os dois pais, passe para a Questão 7]

6. Por quê não vives com teu pai e/ou tua mãe?

Orfã(o) de mãe ☐ Orfã(o) de pai ☐ Orfã(o) de pai e mãe ☐ Outro ☐ _____

7. Repetiste a:

1ª classe? SIM ☐ NÃO ☐ Não Sabe/Não se aplica ☐

2ª classe? SIM ☐ NÃO ☐ Não Sabe/Não se aplica ☐

3ª classe? SIM ☐ NÃO ☐ Não Sabe/Não se aplica ☐

8. Estudaste na segunda classe nesta escola no ano passado? SIM ☐ NÃO ☐ Não Sabe/Não se aplica ☐

9. Estudaste na terceira classe nesta escola no ano passado? SIM ☐ NÃO ☐ Não Sabe/Não se aplica ☐

10. Usas os "livrinhos"para leitura quando estás na escola? SIM ☐ NÃO ☐

11. Usas os "livrinhos" quando estás na sala de aulas? SIM ☐ NÃO ☐

12. Levas os "livrinhos" para ler na tua casa? SIM ☐ NÃO ☐

NÃO PERGUNTAR AO ALUNO (PERGUNTAS 13-16) VERIFICAR NO LIVRO DA TURMA

13. Quantos dias o aluno esteve presente na aula em Julho 2015? _____ dias

14. Quantos dias o aluno esteve presente na aula em Agosto 2015? _____ dias

15. Quantos dias de aula teve esta turma durante o mês de Julho 2015? _____ dias

16. Quantos dias de aula teve esta turma durante o mes de Agosto 2015? _____ dias

Annex D. SMA Package (Teacher Interview)

USAID / Aprender a Ler
SMA: Entrevista com o/a Professor/a

APENAS para o/a professor/a das turmas onde o EGRA foi aplicado (2º ou 3º classe)

Antes de Iniciar a Entrevista (A SER PREENCHIDO PELO SUPERVISOR APAL)

A.1 Nome do inquiridor: _____	**A.2 Data:** ____/____/ 2015
A.3 Tipo de Tratamento oferecido à escola: **Completo** ☐ **Médio** ☐ **Controlo** ☐	
A.4 Província: Nampula ☐ **Zambézia** ☐	**A.5 Distrito:** _____
A.6 Nome da Escola: _____	**A.7 Código da Escola:** _____
A.8 ZIP: _____	
A.9 Classe: Segunda ☐ **Terceira** ☐	
A.10 Nome do/a entrevistado/a (apenas para referência): _____	

Secção A: Informações sobre o/a Professor/a:

MARCAR UM X A RESPOSTA CORRECTA. SE COMETER ALGUM ERRO RISQUE NA RESPOSTA INICIALMENTE MARCADA E FAÇA UM X NA OPÇÃO QUE CONSIDERAR CORRECTA

A.11 Sexo: Feminino ☐ Masculino ☐

A.12 Idade: _____anos

A.13 Número de Identificação do Professor: _____

A.14 Língua materna: Português ☐ Macua ☐ Chuabo ☐ Lomwe ☐ Outra ☐ _____

A.15 Teve formação como professor? SIM ☐ NÃO ☐ [Passe para a questão A.17]

A.16 Se respondeu SIM na pergunta anterior, que tipo de curso de formação possui? [Apenas uma resposta]

EHPP ☐	9ª + 3 Anos (IMP) ☐	Bacharelato (UP/UCM) ☐
Magistério Primário (MP) ☐	CFPP 6ª/7ª + 3 Anos ☐	Licenciatura (UP/UCM) ☐
UEM/CFP 7ª/9ª ☐	Instituto Magistério Primário (IMP) ☐	ADPP ☐
UEM/CFP 10ª/11ª ☐	10ª + 1 Ano ☐	
Curso 6ª + 3 Anos ☐	12ª + 1 Ano ☐	
Outra ☐	Especificar: _____	

A.17 Durante 2015, teve algum tipo de formação/capacitação em exercício?

SIM ☐ NÃO ☐ [Passe para a questão A.21]

A.18 Durante 2015, beneficiou de algum tipo de formação/capacitação em exercício sobre a leitura nas classes iniciais? SIM ☐ NÃO ☐ [Passe para a questão A.21]

A.19 Quantos dias de formação/capacitação em exercício recebeu sobre a leitura nas classes iniciais durante 2015? _____ dias [Anote 0 (zero) se não recebeu formação durante 2015 e Passe para a A.21]

A.20 Se recebeu formação/capacitação durante 2015, de quem recebeu essa capacitação?

MINED ☐ IFP ☐ DPEC ☐ Programa de extenção universitário ☐

SDEJT ☐ ONG ☐ Director pedagógico ☐ Coordenação da ZIP ☐

A.21 Quantos anos de experiência tem como professor/a? _____anos completos

A.22 Há quantos anos está nesta escola? _____anos completos (coloque zero se é o primeiro ano)

A.23 Ensinou nesta escola a 2ª. ou 3ª. classes em 2014? SIM ☐ NÃO ☐

A.24 Os seus alunos dispõem dos "livrinhos" para a leitura individual (ApaL) en número suficiente?

SIM ☐ NÃO ☐

A.25 Tem o programa de ensino básico? SIM ☐ NÃO ☐

A.26 Tem o guia/guião do professor? SIM ☐ NÃO ☐

A.27 Tem o manual do Projecto APAL? SIM ☐ NÃO ☐

A.28 Tem alunos na sala de aulas que têm necessidades educativas especiais? SIM ☐ NÃO ☐

A.29 Se SIM, quantos? _____meninas _____meninos

A.30 Utiliza a língua local do aluno(a) para facilitar o ensino-aprendizagem da leitura em Português?

SIM ☐ NÃO ☐

A.31 Segundo o/a professor/a, quantos dias faltou em Julho 2015? _____dias

A.32 Segundo o/a professor/a, quantos dias faltou em Agosto 2015? _____dias

Indique o número de dias que faltou em Agosto por tipo de motivo:

A.32a Doença própria _____dias

A.32b Doença na família _____dias

A.32c Problemas familiares _____dias

A.32d Formação/capacitação _____dias

A.32e Resolver problemas administrativos da escola_____dias

A.33 Pedir o/a professor/a para ver o caderno de desempenho. Foi possivel ver o caderno desempenho?

SIM ☐ NÃO ☐

A.34 Se o/a professor/a tiver o caderno desempenho, está em uso? SIM ☐ NÃO ☐ N/A ☐

Annex E. Field-level Sampling Strategy for Classrooms & Students for EGRA Administration

A simple, unbiased method for selection of classrooms is to list, for each grade and in alphabetical order the name of the all the second and third grade teachers present in the school on the day of the visit. The classroom to be assessed is selected based on the day of the month of the visit, as follows:

- If only one classroom of the grade, that one will be assessed.
- If two classrooms:
    - Day of the month is 01-15, select the first from the list.
    - Day of the month is 16-31, select the second.
- If three classrooms:
    - Day of the month is 01-10, select the first from the list.
    - Day of the month is 11-20, select the second.
    - Day of the month is 21-31, select the third.
- If four or more classrooms:
    - Day of the month is 01-07, select the first from the list.
    - Day of the month is 08-14, select the second.
    - Day of the month is 15-21, select the third.
    - Day of the month is 22-31, select the fourth.

This procedure will effectively eliminate any bias in selection of the classroom to be assessed that is based on a director's or teacher's knowledge of classroom performance or teacher characteristics.

The next step requiring randomization is the selection of students within the selected classroom who will be administered the EGRA assessment instrument. The objective of a procedure to be applied in the field is to eliminate teacher or enumerator bias in student selection. Time constraints in terms of the length of the school day, the time limitations of the enumerator in the classroom and the time it takes to administer the EGRA assessment suggest that ten students per classroom selected (per grade) be assessed. Again, we use a random selection procedure that can be applied in the field that will eliminate teacher or enumerator bias, as follows:

- First, students in attendance will be organized into rows, if possible, to facilitate counting off. If rows cannot easily be created (such as outdoors), the students should form a line. The rows or line can be made in any order, or none at all, as simple random sampling will be employed. The total number of students in attendance will be determined by counting of the students present in the rows or line.
- If there are ten or fewer students present, all will be assessed.
- If there are more than ten students, the following table shall be employed by first selecting the row in the table corresponding to the total number of students in attendance. Students will then be "count off" and if the count corresponds to any of the ten numbers to the right in the selected row in the table, then the student is asked for their name and has been selected for assessment.
- If a selected student cannot or will not participate in the assessment, the enumerator may choose one before or one after in the rows or line of students that has not already been chosen. If, as the EGRA assessments are initiated, a student does not wish to participate or is no longer present, the enumerator may choose another student of the same sex.
- Shown below is a fragment of the selection table showing which students are selected when the number of students present ranges from 40-44. The full table covered cases with attendance ranging from 11 to 75.

Students in Attendance	Corresponding Numbers for the Ten Selected Students									
40	5	10	11	12	13	16	26	27	36	39
41	1	2	6	7	15	16	23	25	31	38
42	3	4	10	14	21	23	28	35	37	40
43	1	2	3	7	11	15	26	28	31	35
44	1	4	9	14	22	26	28	33	41	42

Annex F. Semi-Structured Interview Protocol

Protocolo de entrevistas para directores de escola
USAID / Aprender a Ler

Entreviste o director da escola se o mesmo estiver presente no dia da visita. Em alguns casos, o director pode sugerir que o director pedagógico é a pessoa correcta para entrevistar. Isso ocorre quando o director é novo na função. Caso o director da escola não se encontre presente, entreviste o director pedagógico. Anote abaixo a função do entrevistado.

Director da escola ☐ **Director pedagógico** ☐

A.1 Você era director ou director pedagogico desta escola em 2014? **SIM** ☐ **NÃO** ☐	**A.2 Data:** ___/___/ **2015**
A.3 Tipo de Tratamento oferecido à escola: **Completo** ☐ **Médio** ☐	
A.4 Província: Nampula ☐ **Zambézia** ☐	**A.5 Distrito:** _____
A.5 Nome da Escola:	**A.6 Código da Escola:**
A.7 Nome da Escola Sede do ZIP:	
A.8 Nome do/a entrevistado/a (apenas para referência):	

Fóco da Entrevista

Investigar quais as acções implementadas pelo ApaL em 2014 que foram mantidas em 2015 quando o ApaL não mais estava na escola.

1. ***Formação em leitura para as classes iniciais***
2. ***Formação para o desenvolvimento de materiais didácticos***
3. ***Formação para a gestão escolar***
4. ***Processos para melhor gerir a escola***
5. ***Disponibilidade de materiais didácticos***
   ("livrinhos" para leitura individual, manual do professor, materiais didácticos)

1. Como o ApaL modificou os procedimentos anteriormente utilizados nessa escola? Por exemplo: O que o ApaL trouxe de novo para a escola? Cite uma actividade que você não fazia antes mas que faz agora como resultado de ter participado no projeto ApaL em 2014. Quais as acções e os procedimentos que no momento fazem parte do quotidiano da escola mas que não existiam antes do ApaL?

2. Pensando em tudo o que o ApaL trouxe para essa escola, quais foram, na sua opinião, as acções mais importantes para melhorar o ensino da leitura nas classes iniciais?

3. De todas as acções, actividades e procedimentos implementados pelo ApaL nessa escola quais as que permanecem?

4. Em 2015 foi realizada alguma formação nesta escola direccionada aos professores da segunda e terceira classes na área de ensino de leitura para as classes iniciais?

5. Quem realizou a formação? Director da escola ☐ Director pedagógico ☐

   Coordenação da ZIP ☐ Districto/MINED ☐ Outra organização ☐

6. Em 2015 houve necessidade de desenvolver ou adicionar materiais didácticos para o ensino da leitura nas classes iniciais? Quais?

7. "livrinhos" para a leitura individual ☐ Manual para professores ☐ Manual de apoio didáctico ☐

8. Se isso foi feito, quais os recursos utilizados para essa finalidade?

   Recursos fornecidos pela ZIP ☐ Recursos fornecidos pelo Distrito ☐ Fundo escolar ☐

   Recursos fornecidos por ONG ☐ Outra fonte de recursos ☐

   (Se outra QUAL?) _____

9. O Distrito e a ZIP apoiaram a escola de alguma maneira no esforço de melhorar a leitura nas classes iniciais?

   Sim ☐ Não ☐ Se afirmativo, Como? _____

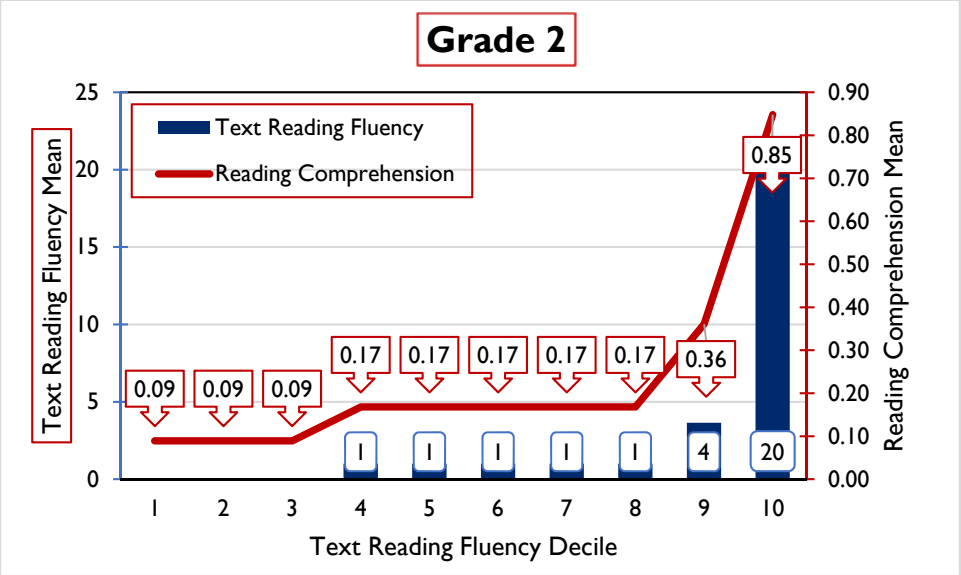
   _____

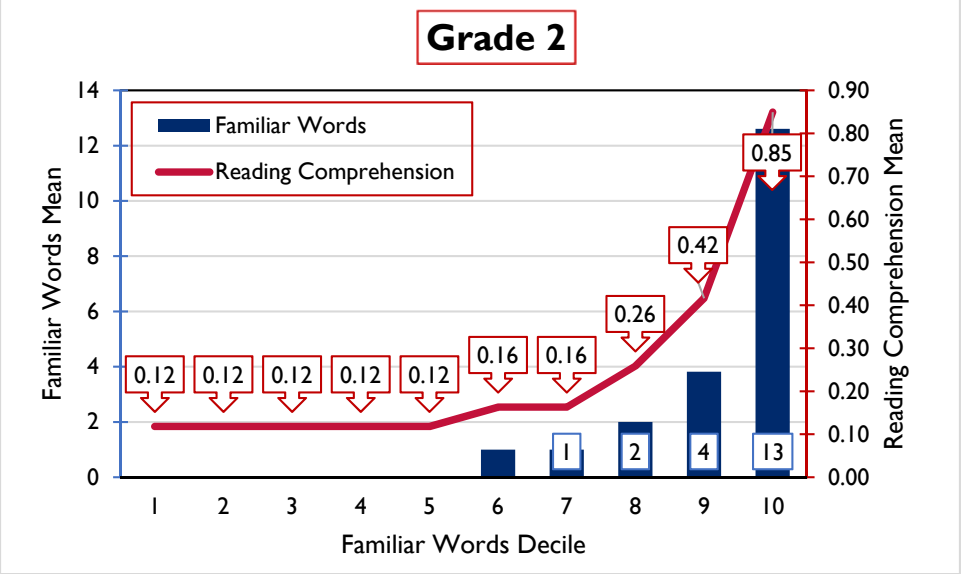
10. Dos professores que ensinavam as 2ª e 3ª Classes em 2014, quantos continuam na escola, ensinando as mesmas classes?
    (Entrevistador, peça que o director seja específico. Não queremos respostas tipo "quase todos" ou "a maioria." Peça ao director que consulte sua lista de professores e indique quais os professores de 2ª. e 3ª. Classes em 2014 que já não estão mais na escola ou que mesmo estando na escola, não mais ensinam essas classes).

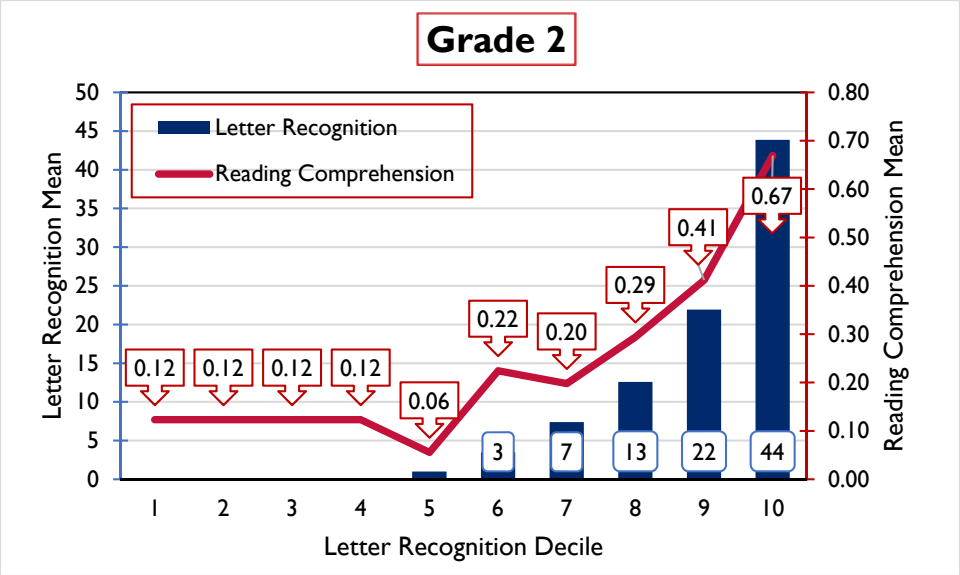
Annex G. Statistical Annexes

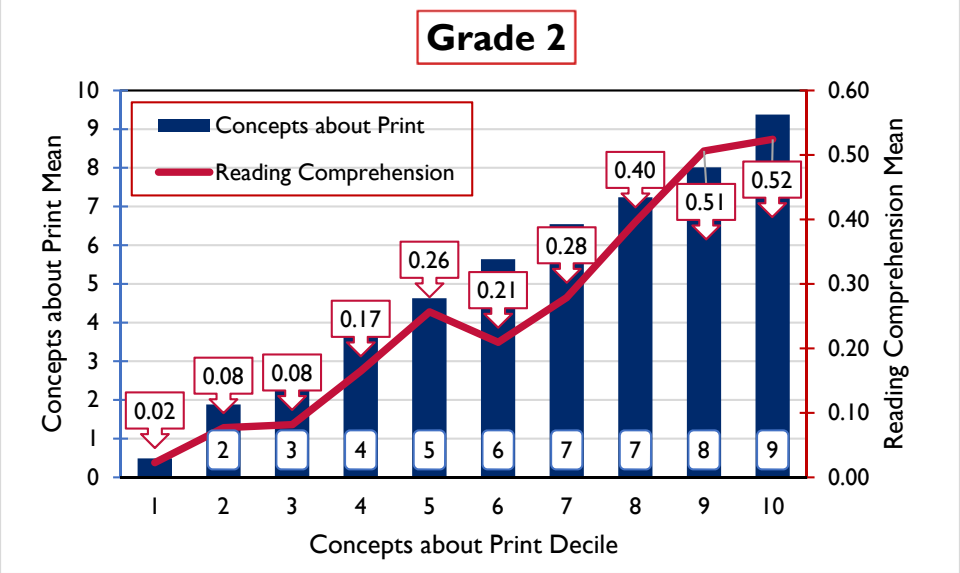
Annex G.1 Deciles Histogram Grade 2 (Text Reading Comprehension, Reading Fluency, Familiar Word Reading Fluency, Letter Recognition, Concepts about Print)

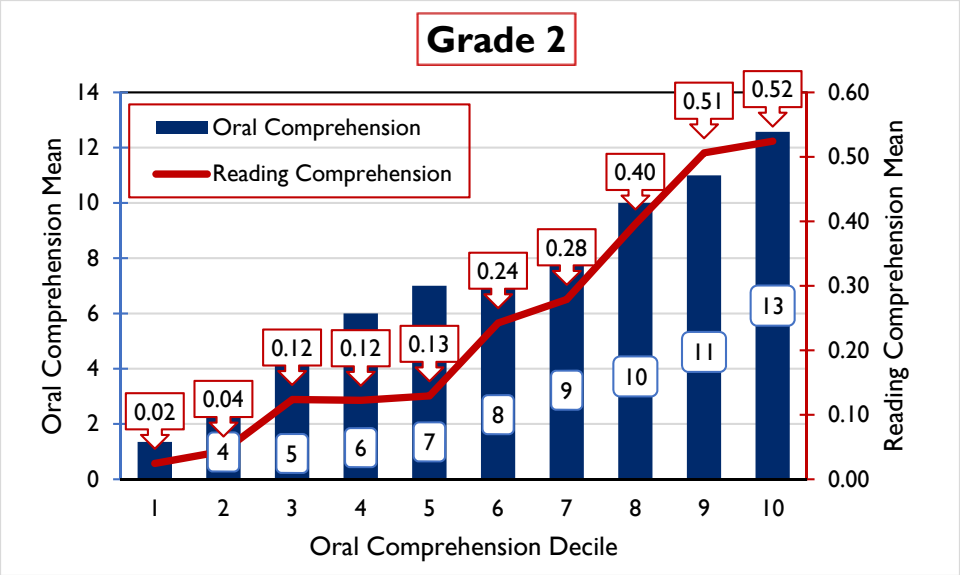
Text Reading Comprehension – Grade 2



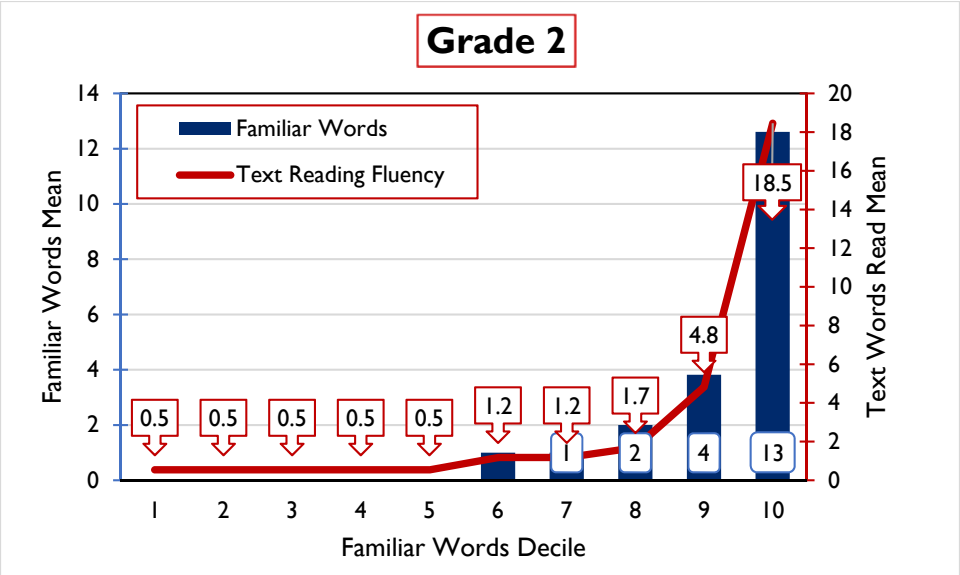


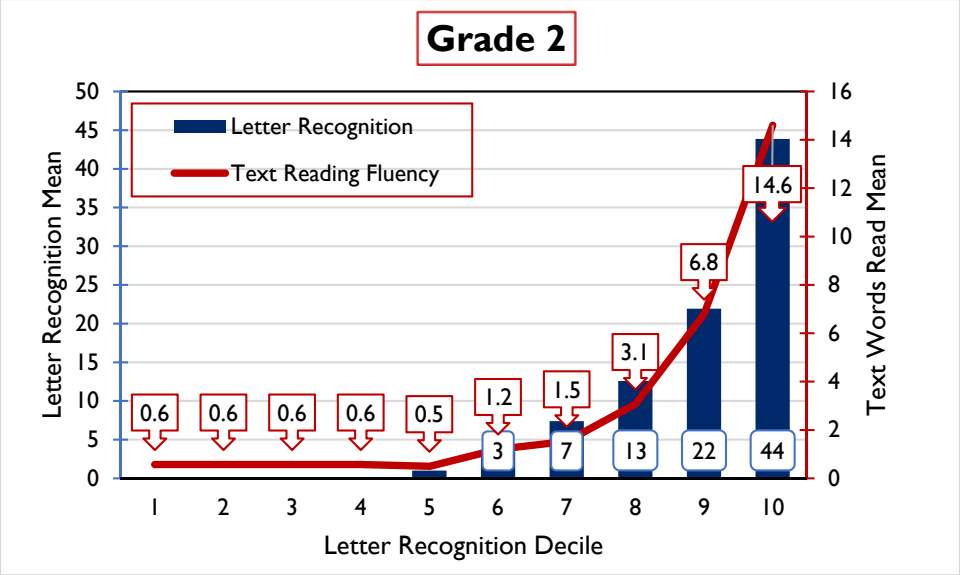

Grade 2
Letter Recognition
Reading Comprehension
Letter Recognition Mean
Reading Comprehension Mean
Letter Recognition Decile
0.12
0.12
0.12
0.12
0.06
0.22
0.20
0.29
0.41
0.67
3
7
13
22
44

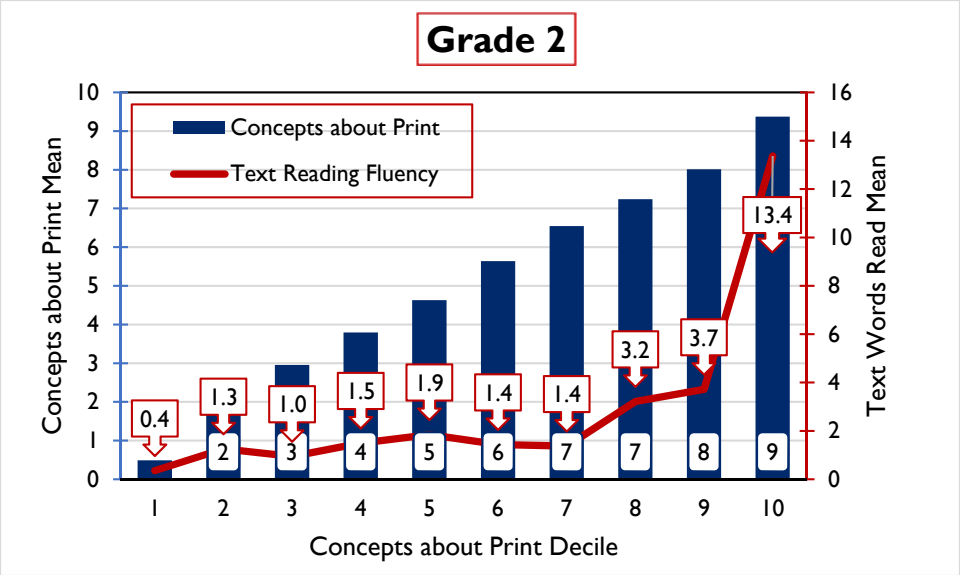

Grade 2
Concepts about Print
Reading Comprehension
Concepts about Print Mean
Reading Comprehension Mean
Concepts about Print Decile
0.02
0.08
0.08
0.17
0.26
0.21
0.28
0.40
0.51
0.52
2
3
4
5
6
7
7
8
9

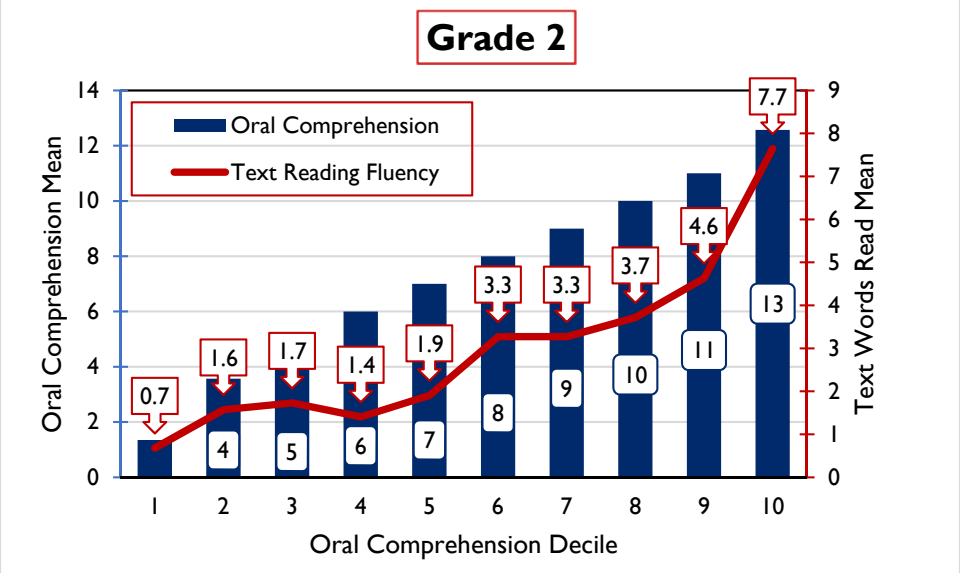

Grade 2
Oral Comprehension
Reading Comprehension
Oral Comprehension Mean
Reading Comprehension Mean
Oral Comprehension Decile
0.02
0.04
0.12
0.12
0.13
0.24
0.28
0.40
0.51
0.52
4
5
6
7
8
9
10
11
13

Text Reading Fluency – Grade 2

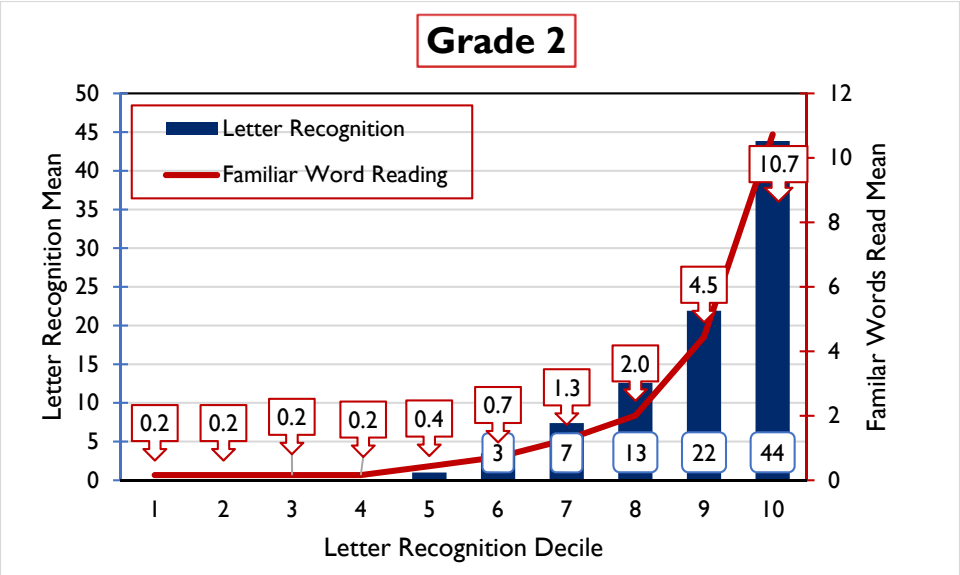


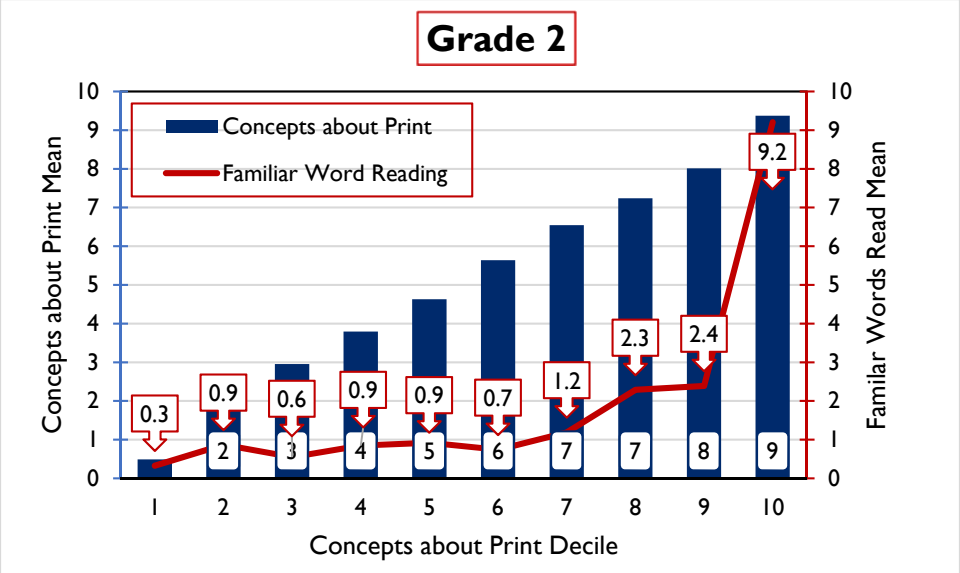


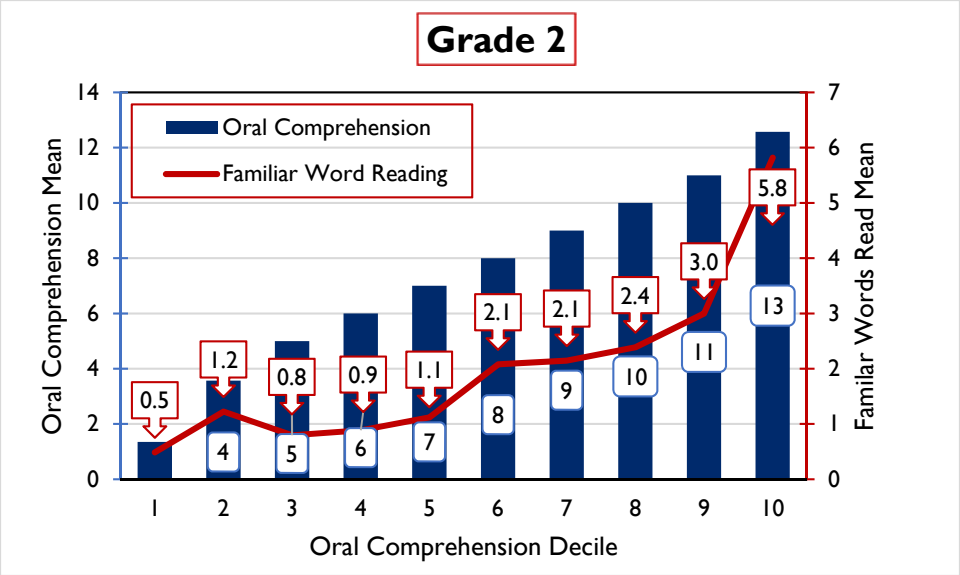

Grade 2
Concepts about Print
Text Reading Fluency
Concepts about Print Mean
Text Words Read Mean
Concepts about Print Decile
0.4
1.3
1.0
1.5
1.9
1.4
1.4
3.2
3.7
13.4
2
3
4
5
6
7
7
8
9


Grade 2
Oral Comprehension
Text Reading Fluency
Oral Comprehension Mean
Text Words Read Mean
Oral Comprehension Decile
0.7
1.6
1.7
1.4
1.9
3.3
3.3
3.7
4.6
7.7
4
5
6
7
8
9
10
11
13

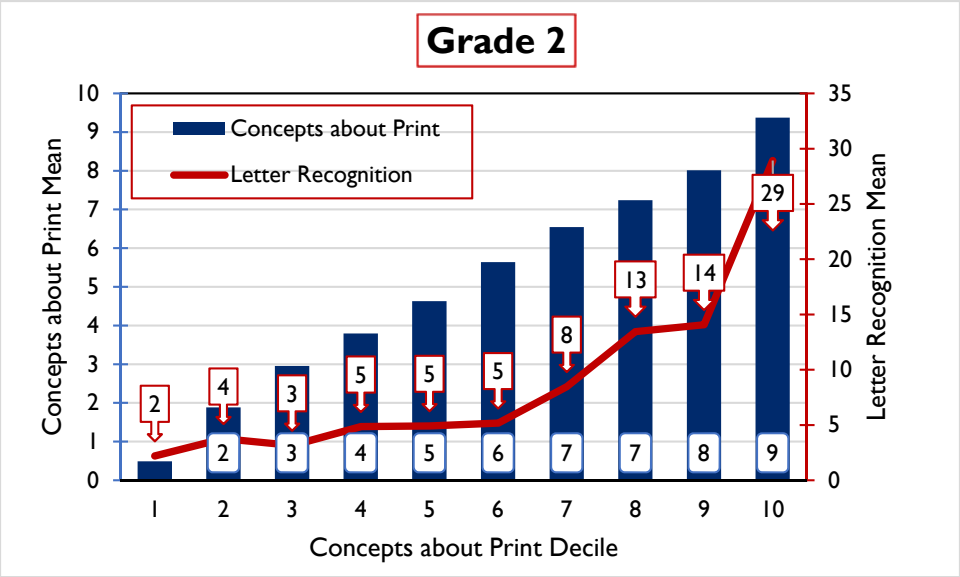
Familiar Word Reading Fluency – Grade 2

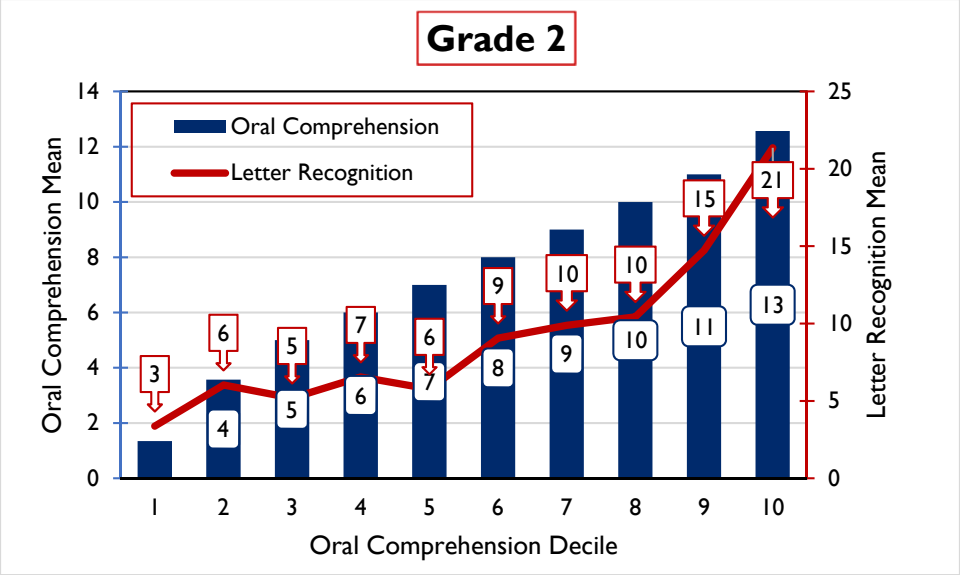




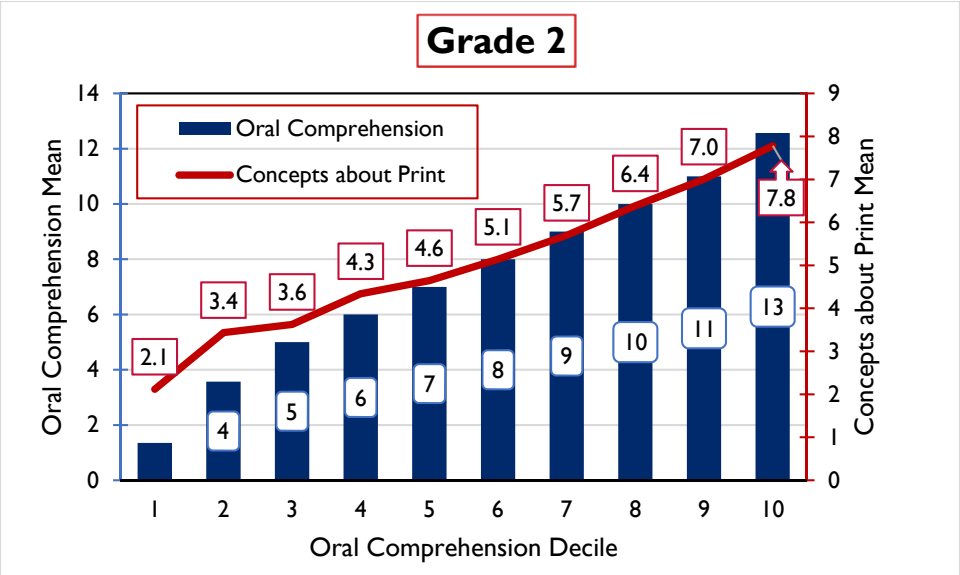

Grade 2
Oral Comprehension
Familiar Word Reading
Oral Comprehension Mean
Familar Words Read Mean
Oral Comprehension Decile
0.5
1.2
0.8
0.9
1.1
2.1
2.1
2.4
3.0
5.8
4
5
6
7
8
9
10
11
13

Letter Recognition – Grade 2



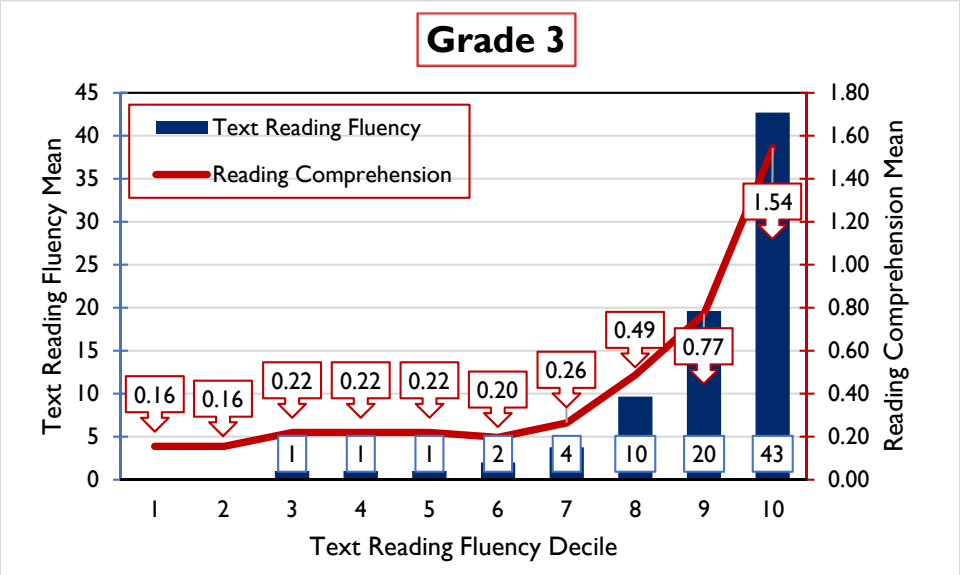


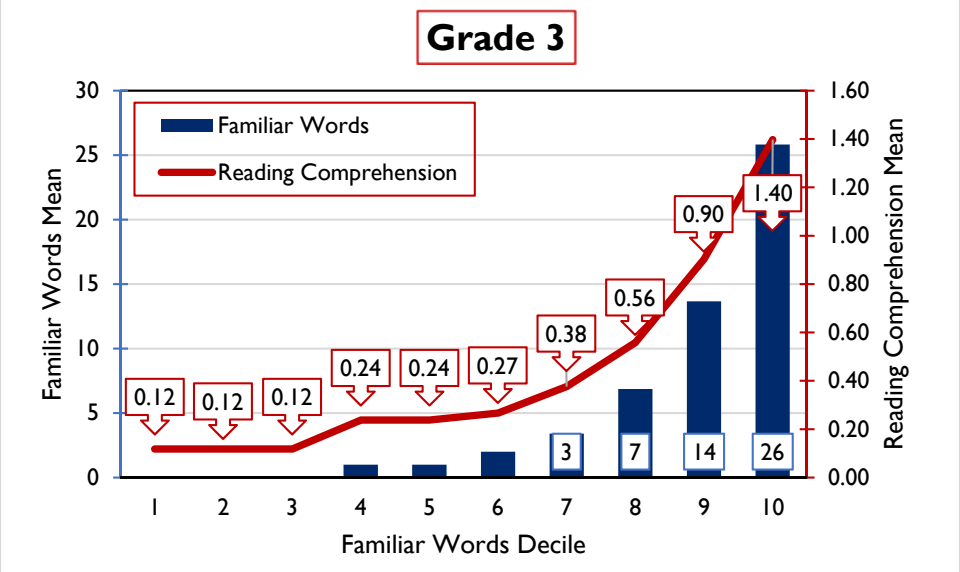
Concepts about Print – Grade 2

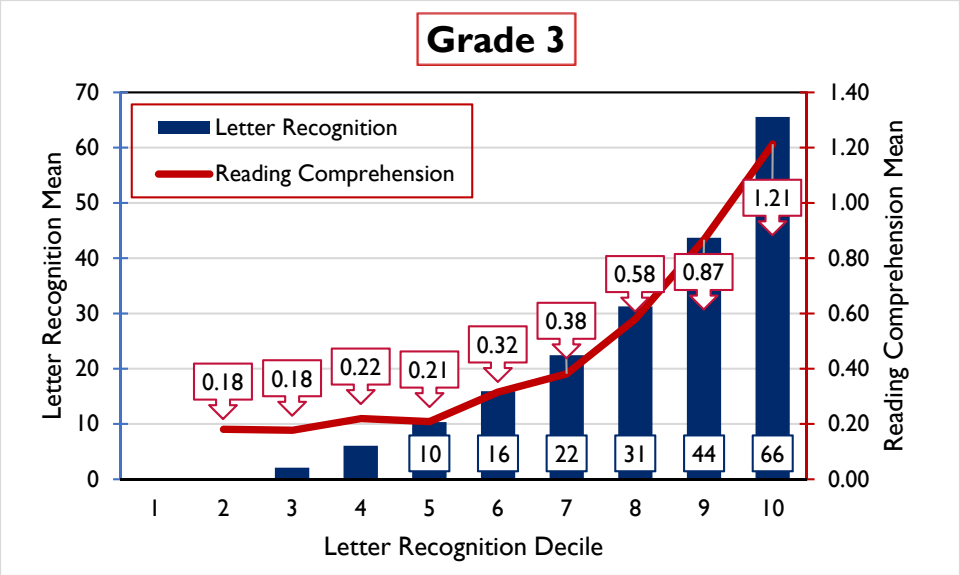


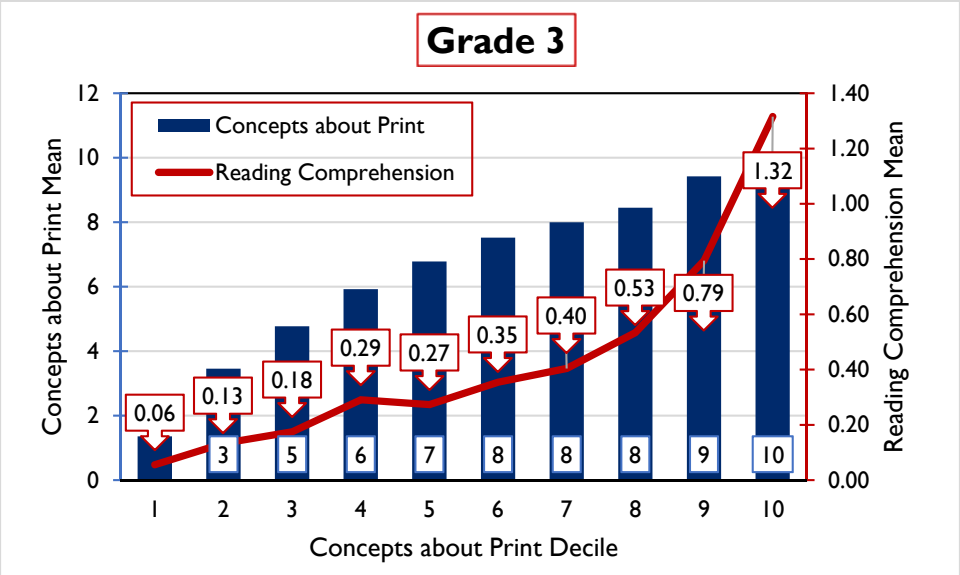
Annex G.2 Deciles Histogram Grade 3 (Text Reading Comprehension, Reading Fluency, Familiar Word Reading Fluency, Letter Recognition, Concepts about Print)

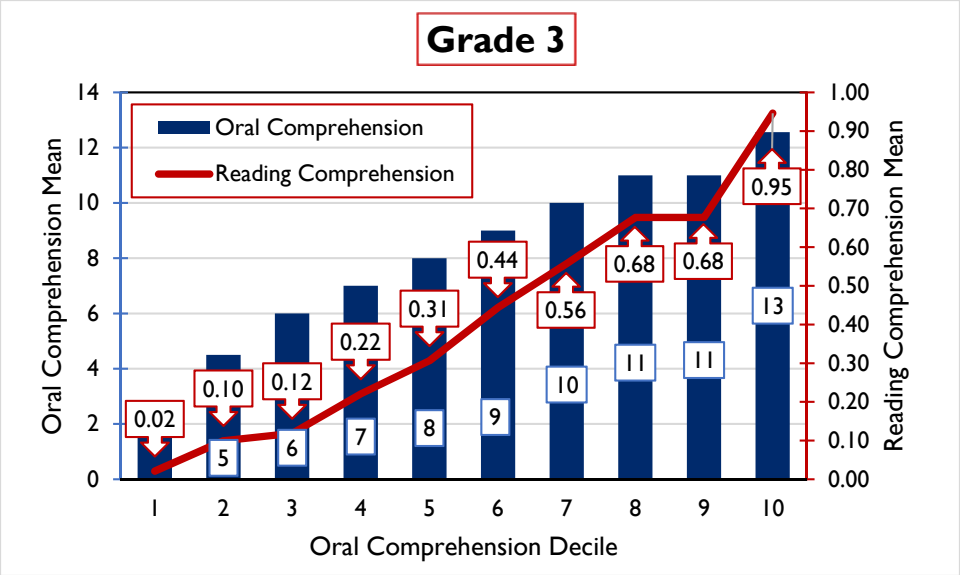
Text Reading Comprehension – Grade 3



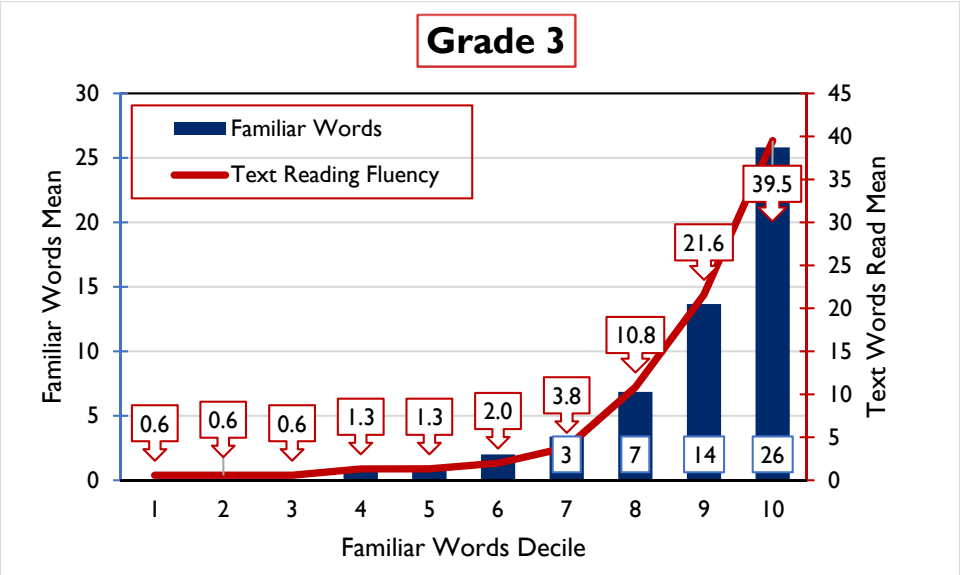


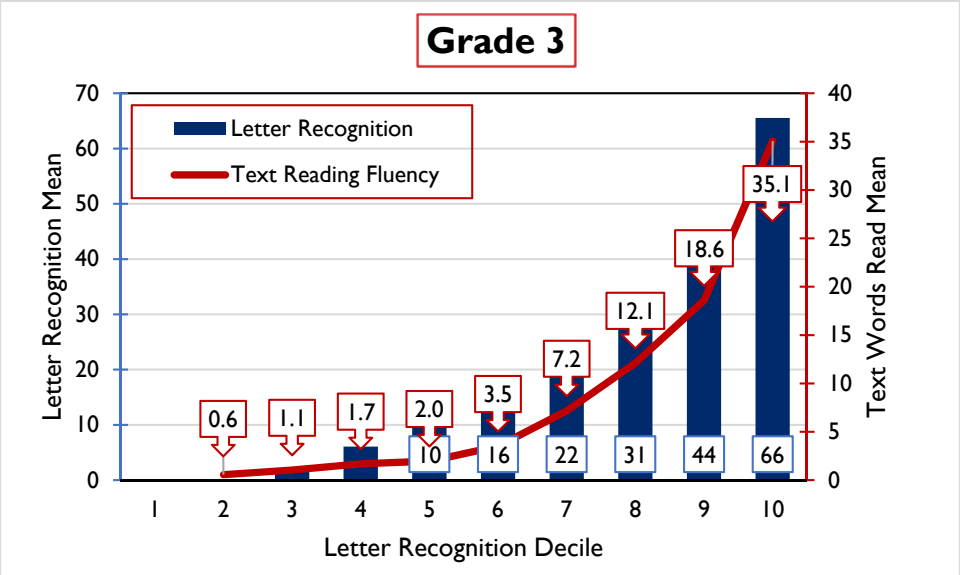

Grade 3
Letter Recognition
Reading Comprehension
Letter Recognition Mean
Reading Comprehension Mean
Letter Recognition Decile
0.18
0.18
0.22
0.21
0.32
0.38
0.58
0.87
1.21
10
16
22
31
44
66

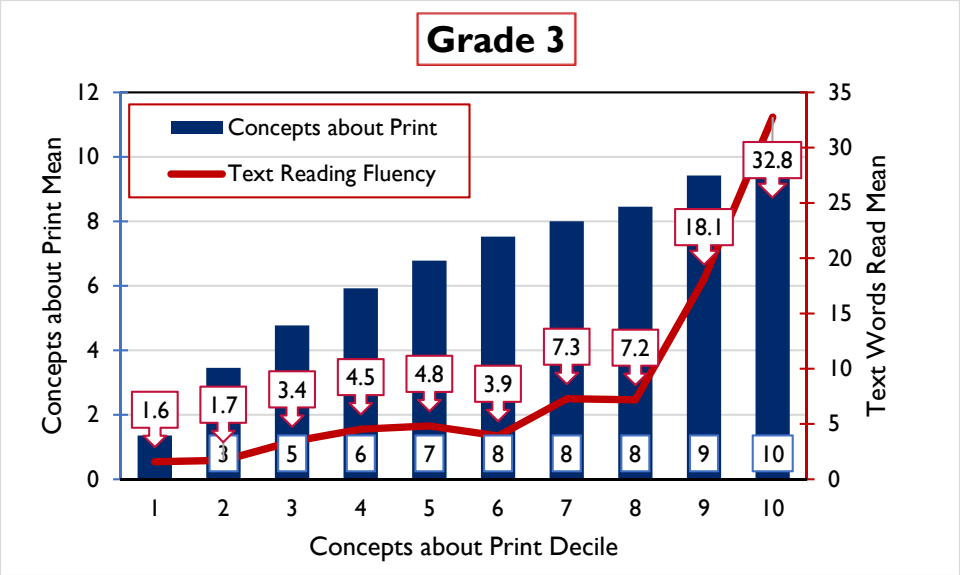

Grade 3
Concepts about Print
Reading Comprehension
Concepts about Print Mean
Reading Comprehension Mean
Concepts about Print Decile
0.06
0.13
0.18
0.29
0.27
0.35
0.40
0.53
0.79
1.32
3
5
6
7
8
8
8
9
10

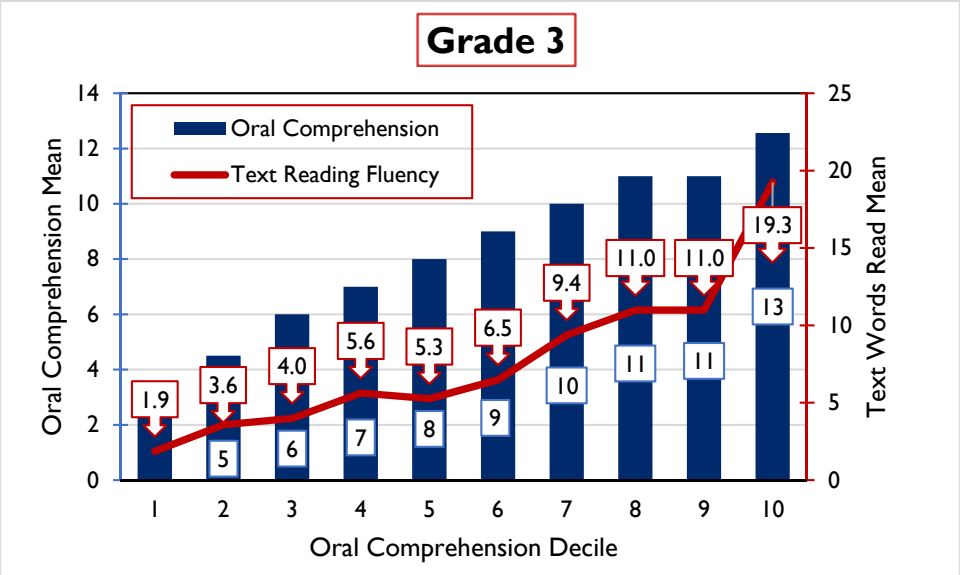

Grade 3
Oral Comprehension
Reading Comprehension
Oral Comprehension Mean
Reading Comprehension Mean
Oral Comprehension Decile
0.02
0.10
0.12
0.22
0.31
0.44
0.56
0.68
0.68
0.95
5
6
7
8
9
10
11
11
13

Text Reading Fluency – Grade 3

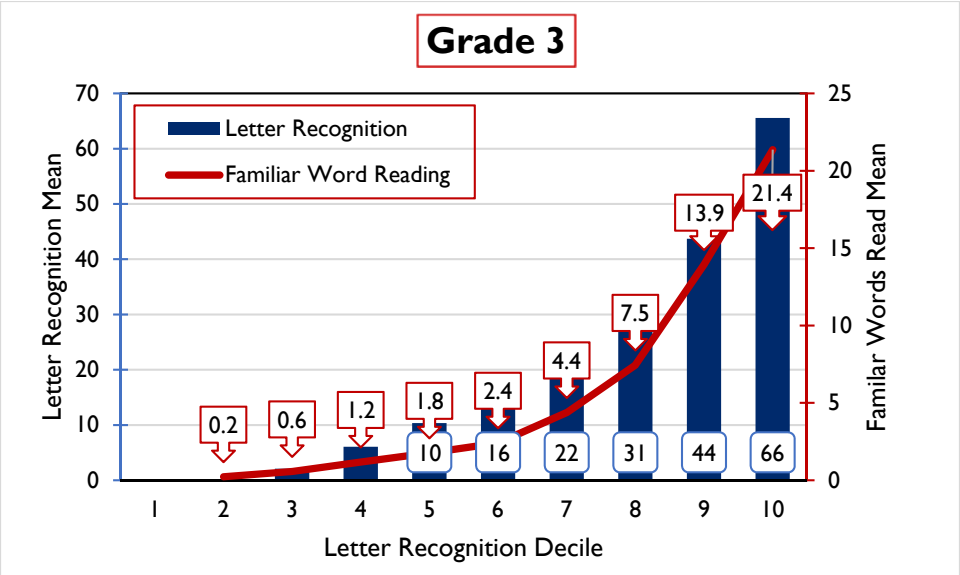


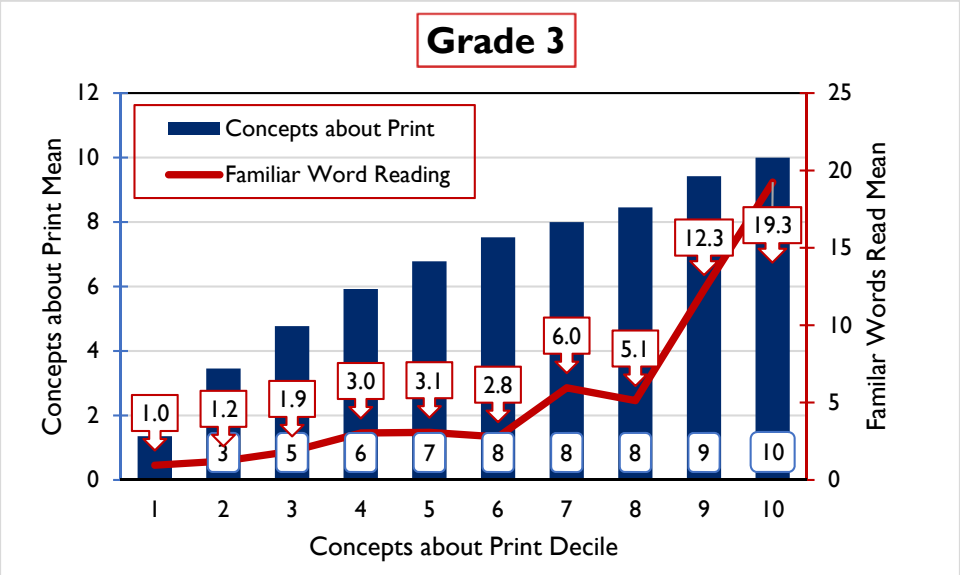


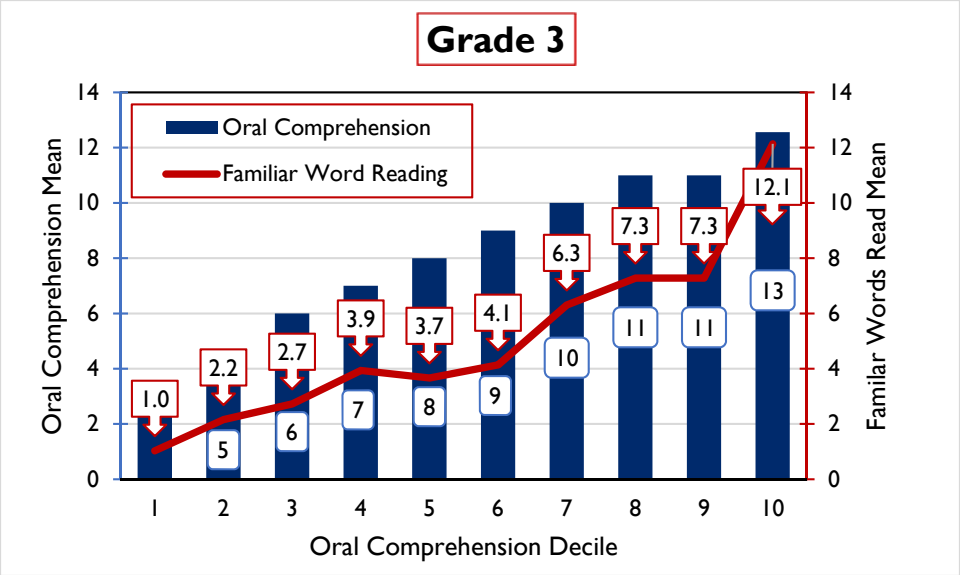

Grade 3
Concepts about Print
Text Reading Fluency
Concepts about Print Mean
Text Words Read Mean
Concepts about Print Decile
1.6
1.7
3.4
4.5
4.8
3.9
7.3
7.2
18.1
32.8
3
5
6
7
8
8
8
9
10


Grade 3
Oral Comprehension
Text Reading Fluency
Oral Comprehension Mean
Text Words Read Mean
Oral Comprehension Decile
1.9
3.6
4.0
5.6
5.3
6.5
9.4
11.0
11.0
19.3
5
6
7
8
9
10
11
11
13

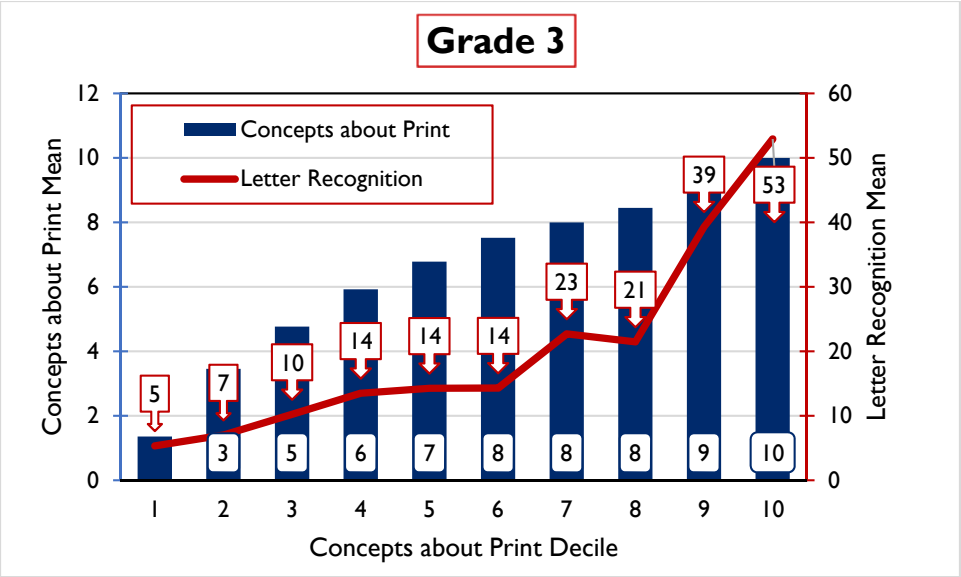
Familiar Word Reading Fluency – Grade 3

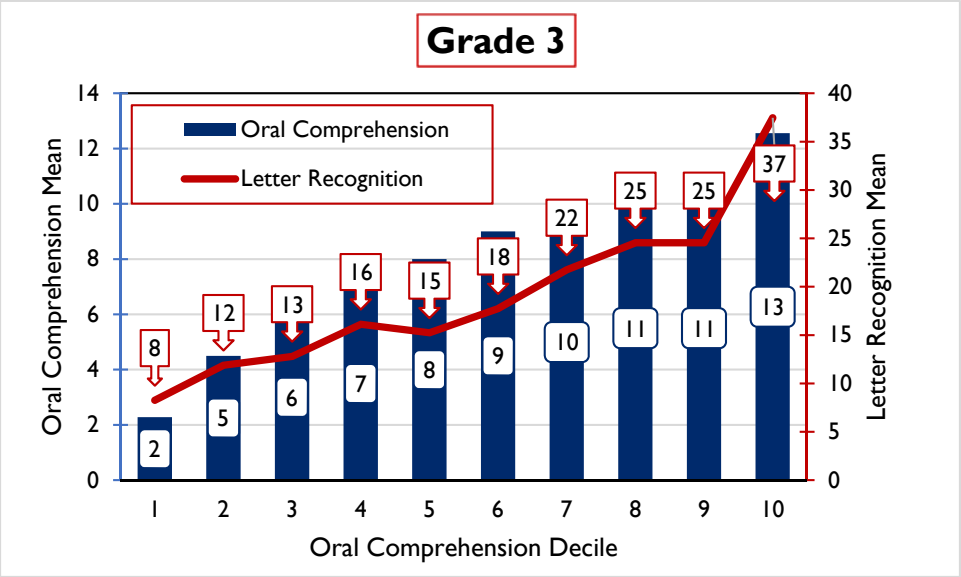




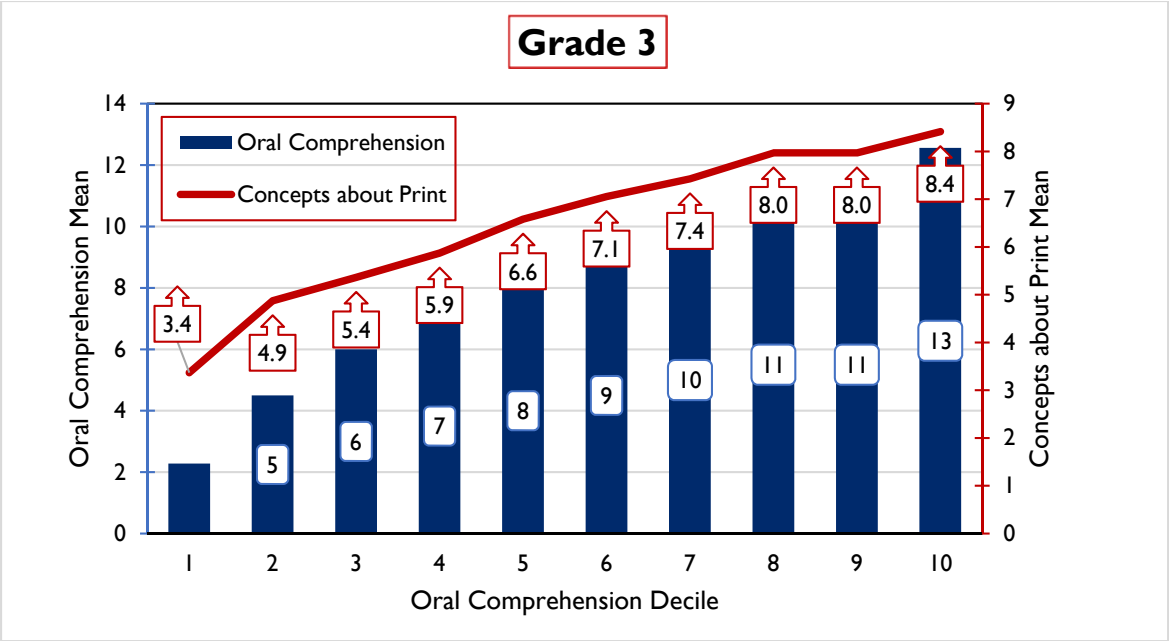

Grade 3
Oral Comprehension
Familiar Word Reading
Oral Comprehension Mean
Familar Words Read Mean
Oral Comprehension Decile
0
2
4
6
8
10
12
14
1
2
3
4
5
6
7
8
9
10
1.0
2.2
2.7
3.9
3.7
4.1
6.3
7.3
7.3
12.1
5
6
7
8
9
10
11
11
13

Letter Recognition – Grade 3





Concepts about Print – Grade 3



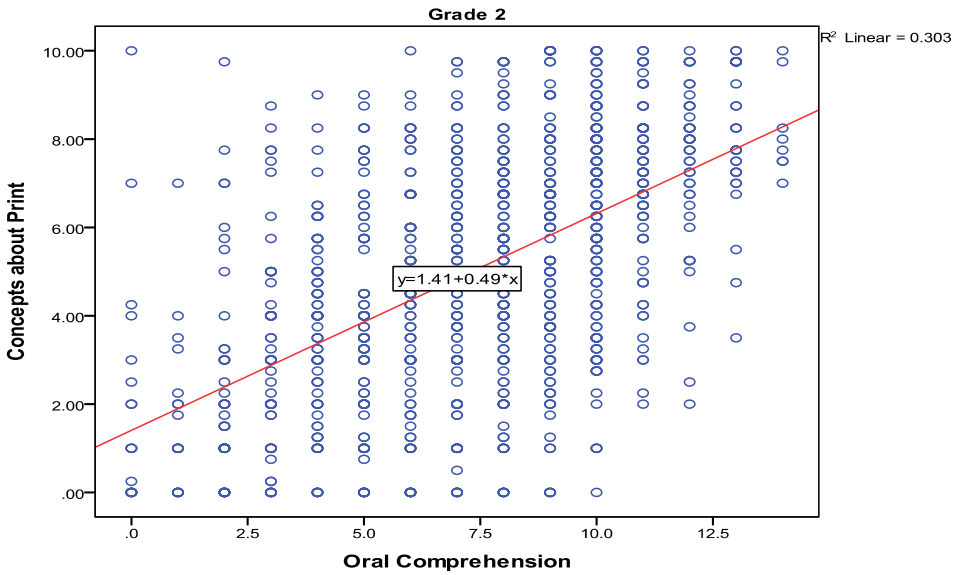
Annex G.3 Scatter Plots – Grades 2 & 3

Oral Comprehension (OC) predicting Concepts about Print (CAP) – Grade 2

CAP= -1.41+ 0.49OC R^2=30.3%

SE (0.17) (0.02)

P (0.00) (0.000)

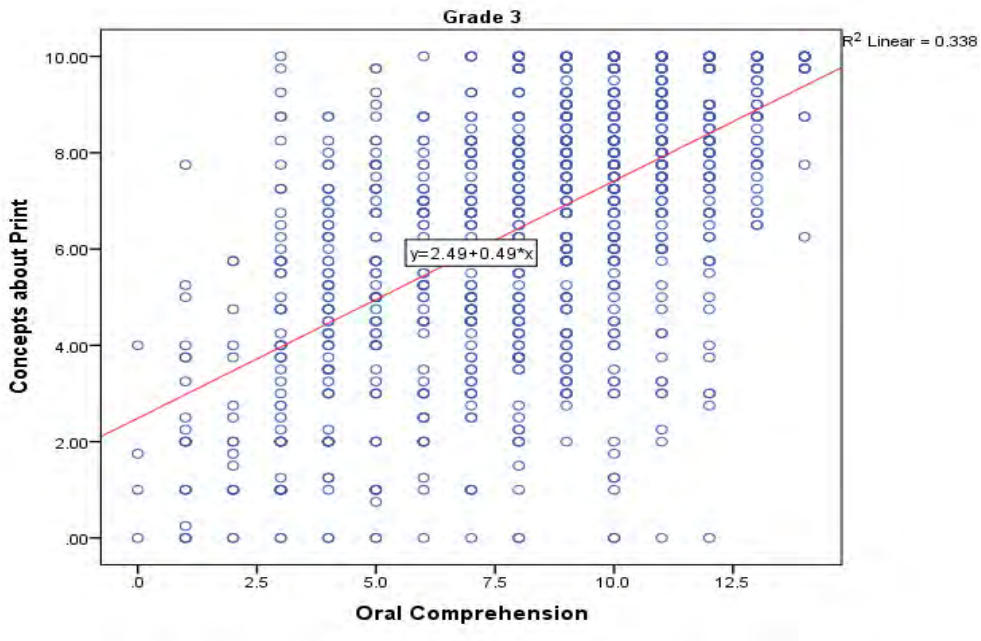


Oral Comprehension (OC) predicting Concepts about Print (CAP) – Grade 3

CAP= -2.49+ 0.49OC **R^2=33.8%** **moderate predictor**

SE (0.18) (0.02)

P (0.00) (0.000)

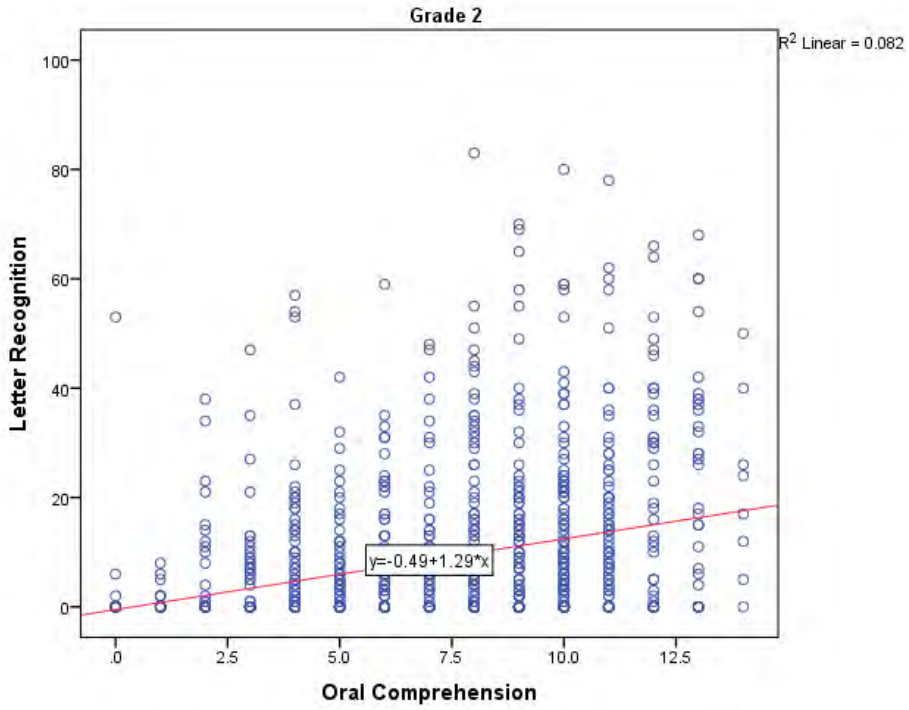


Oral Comprehension (OC) predicting Letter Recognition (LR) – Grade 2

LR= -0.49+ 1.29 OC R^2=8.2%

SE (1.0) (0.13)

P (0.64) (0.000)

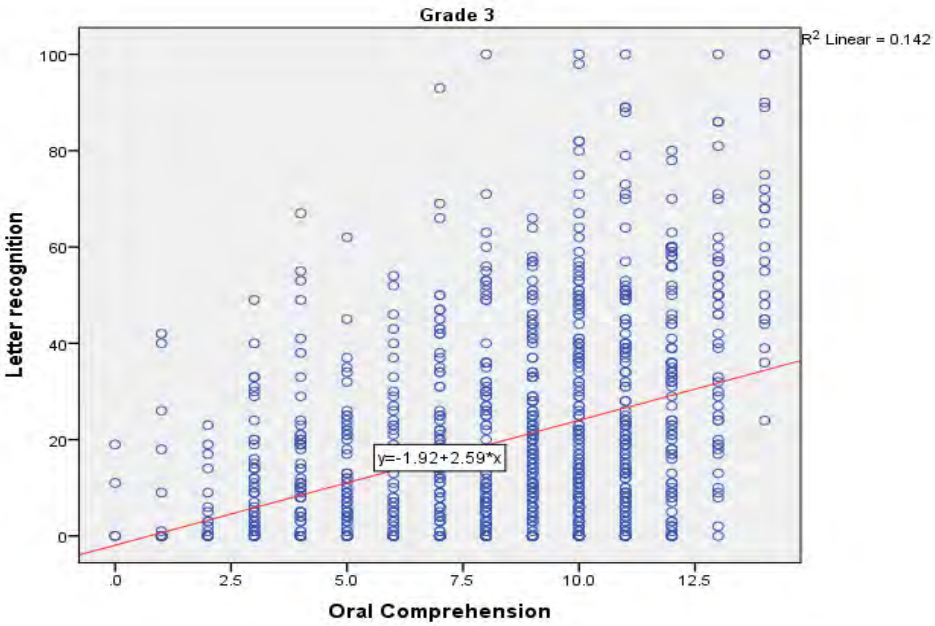


Oral Comprehension (OC) predicting Letter Recognition (LR) – Grade 3

LR= -1.92+ 2.2.6OC R^2=14.2%

SE (1.7) (0.19)

P (0.26) (0.000)

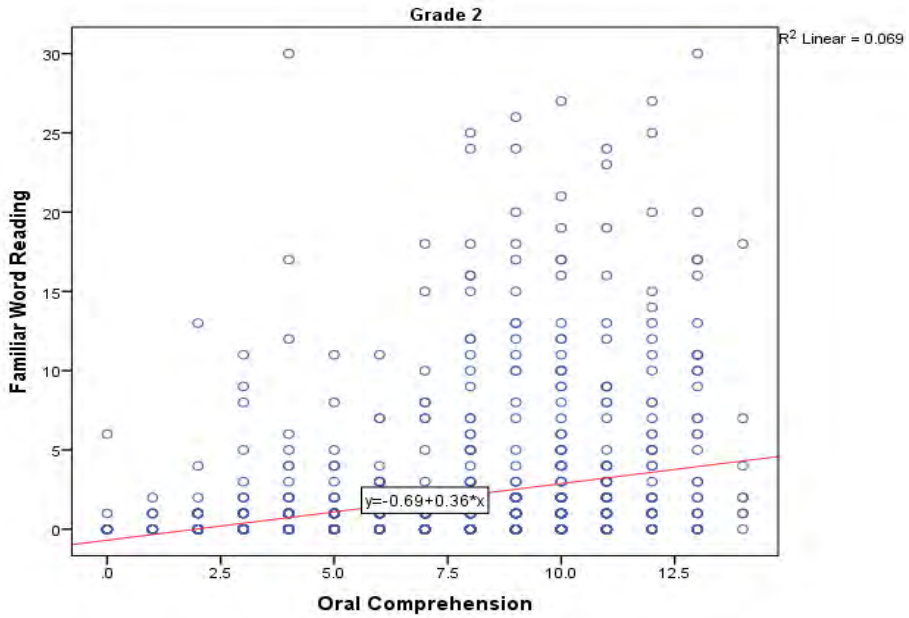


Oral Comprehension (OC) predicting Familiar Word Reading (FWR) – Grade 2

FWE= -0.69+ 0.36OC R^2=6.9%

SE (0.31) (0.04)

P (0.029) (0.000)

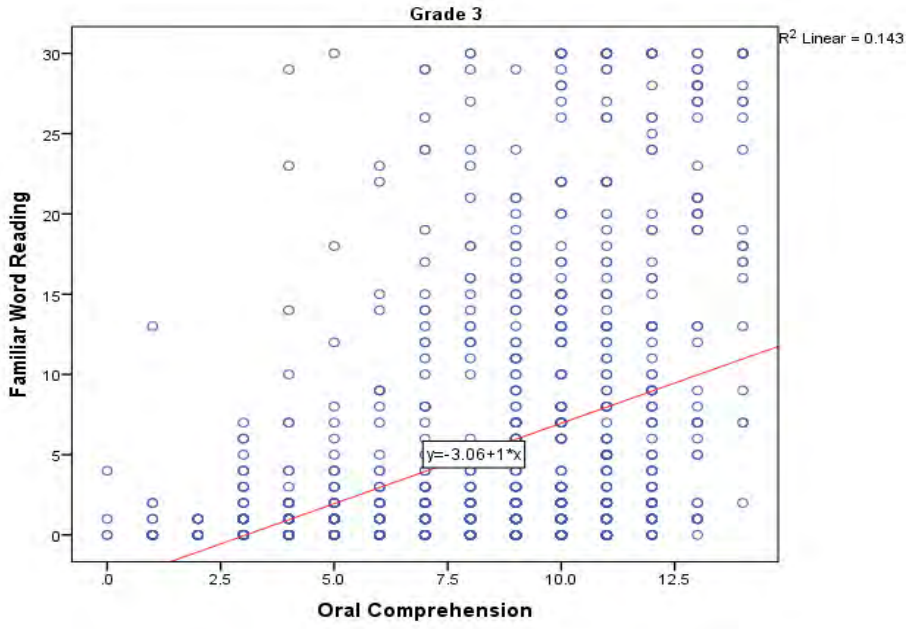


Oral Comprehension (OC) predicting Familiar Word Reading (FWR) – Grade 3

FWE= -3.06+ OC R^2=14.3%

SE (0.65) (0.07)

P (0.00) (0.000)

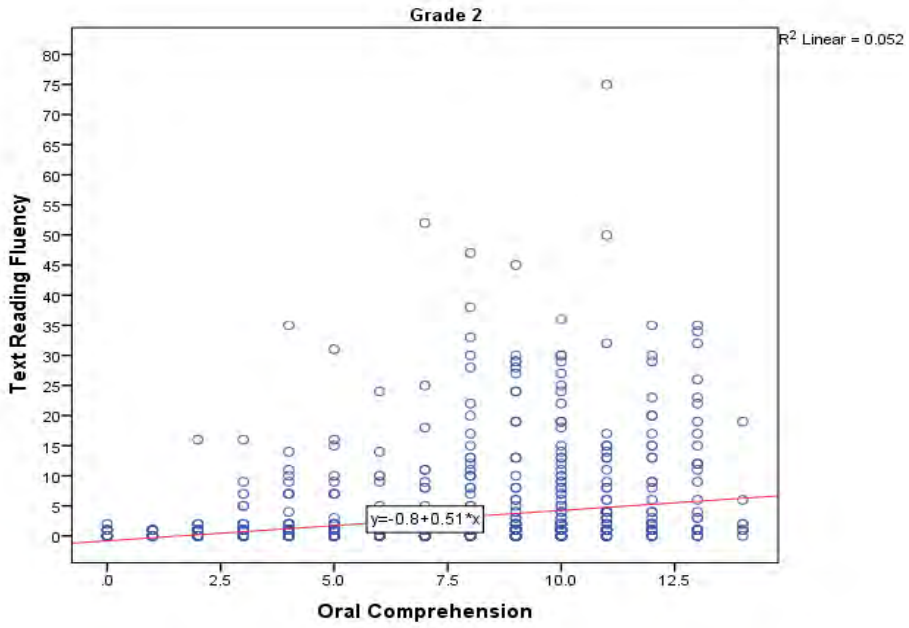


Oral Comprehension (OC) predicting Oral Reading Fluency (ORF) – Grade 2

ORF= -0.8+ 0.51OC R^2=5.2%

SE (0.52) (0.06)

P (0.12) (0.000)

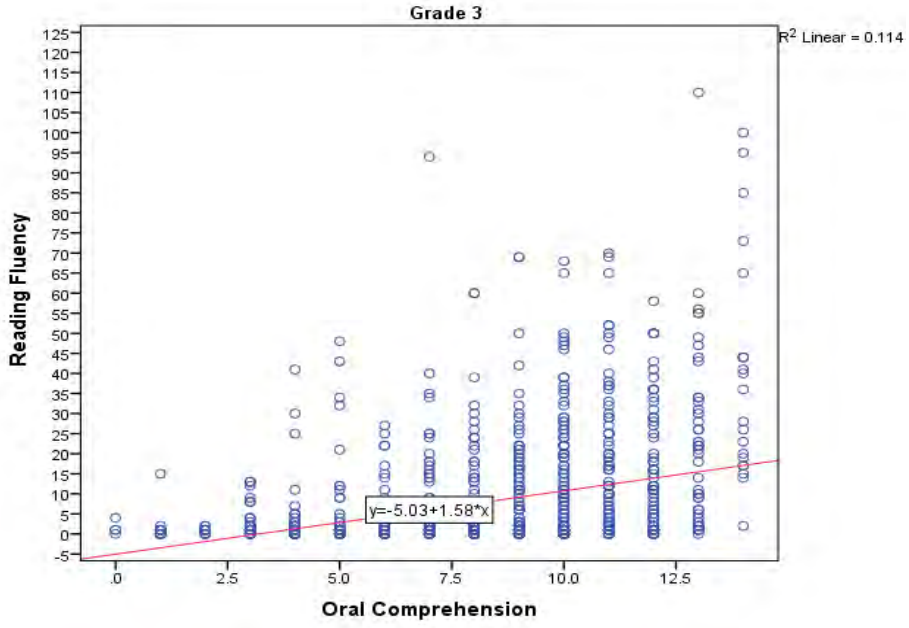


Oral Comprehension (OC) predicting Oral Reading Fluency (ORF) – Grade 3

ORF= -5.0+ 1.58OC R^2=11.4%

SE (1.1.) (0.13)

P (0.00) (0.00)

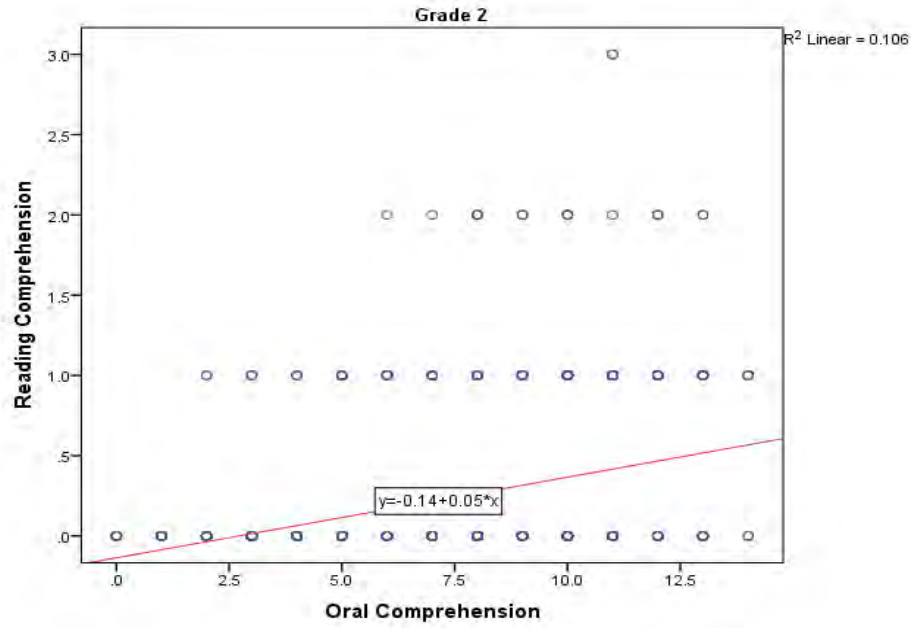


Oral Comprehension (OC) predicting Reading Comprehension (RC) – Grade 2

RC= -0.14+ 0.05OC R^2=10.6%

SE (0.03) (0.00)

P (0.00) (0.00)

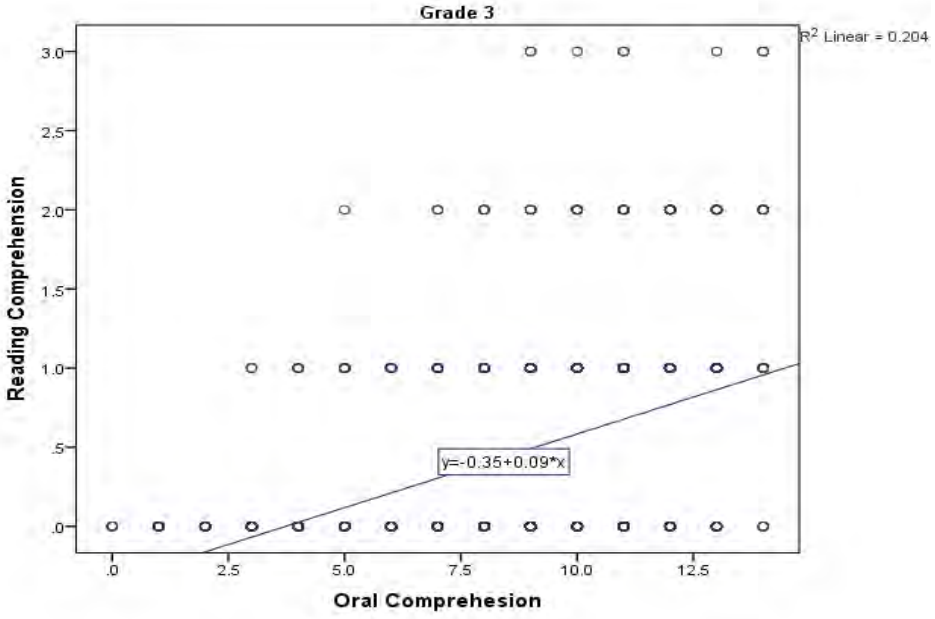


Oral Comprehension (OC) predicting Reading Comprehension (RC) – Grade 3

RC= -0.35+ 0.09OC R^2=20.4 %

SE (0.04) (0.00)

P (0.00) (0.00)

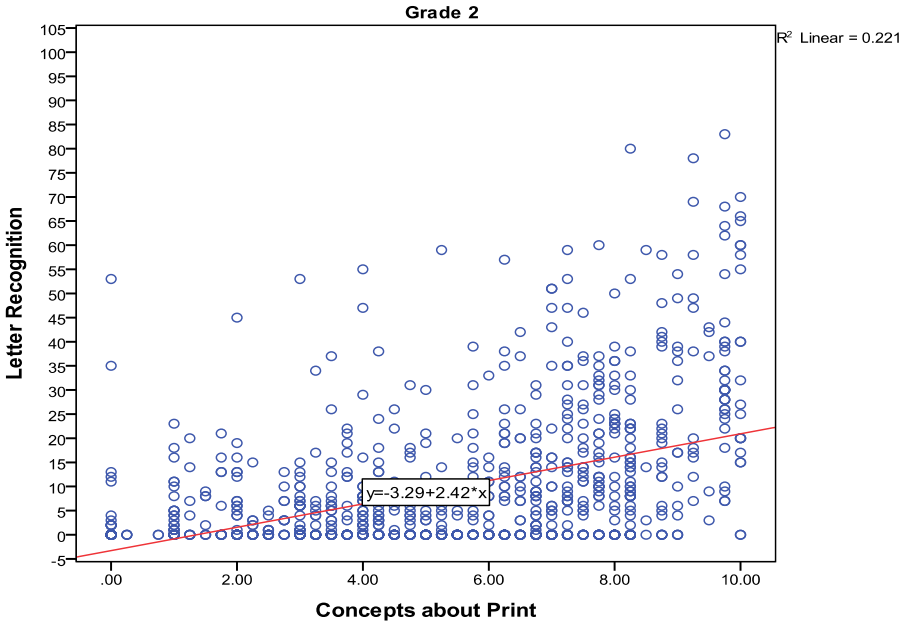


Concepts about Print (CAP) predicting Letter Recognition (LR) – Grade 2

LR= -3.29+ 2.42CAP R^2=22.1%

SE (0.8) (0.13)

P (0.00) (0.00)

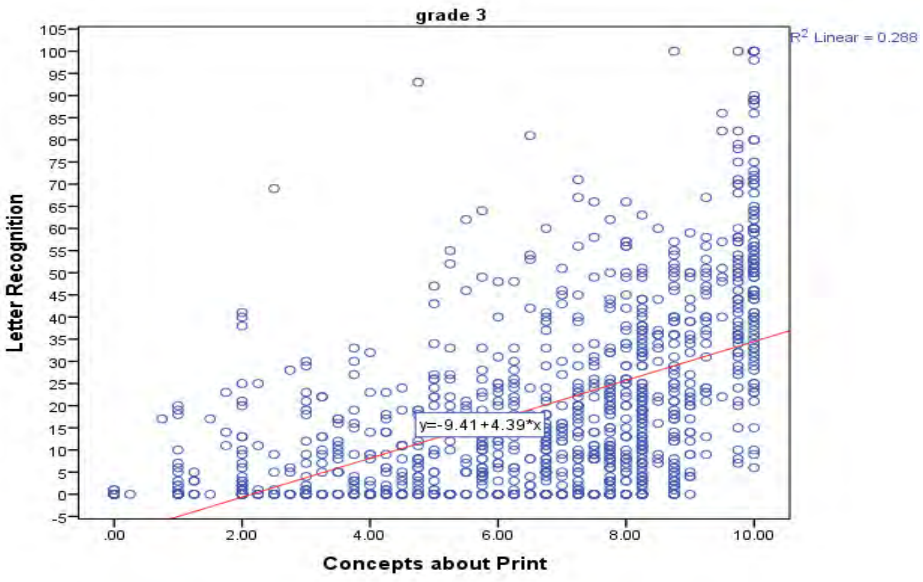


Concepts about Print (CAP) predicting Letter Recognition (LR) – Grade 3

LR= -9.41+ 4.39CAP R^2=28.8%

SE (1.4) (0.20)

P (0.00) (0.00)

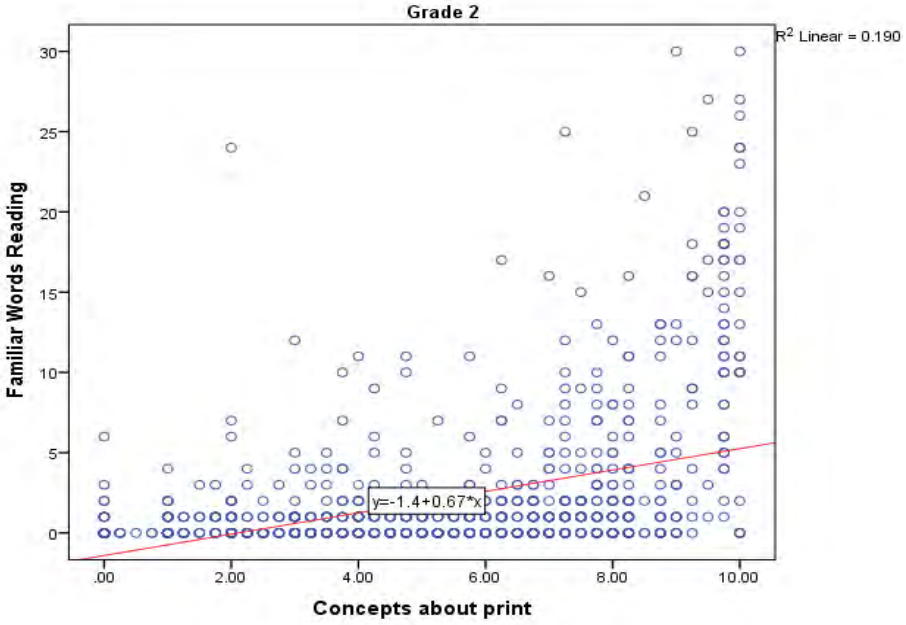


Concepts about Print (CAP) predicting Familiar Words Reading (FWR) – Grade 2

FWR= -1.4+ 0.67CAP R^2=19.0%

SE (023) (0.04)

P (0.00) (0.00

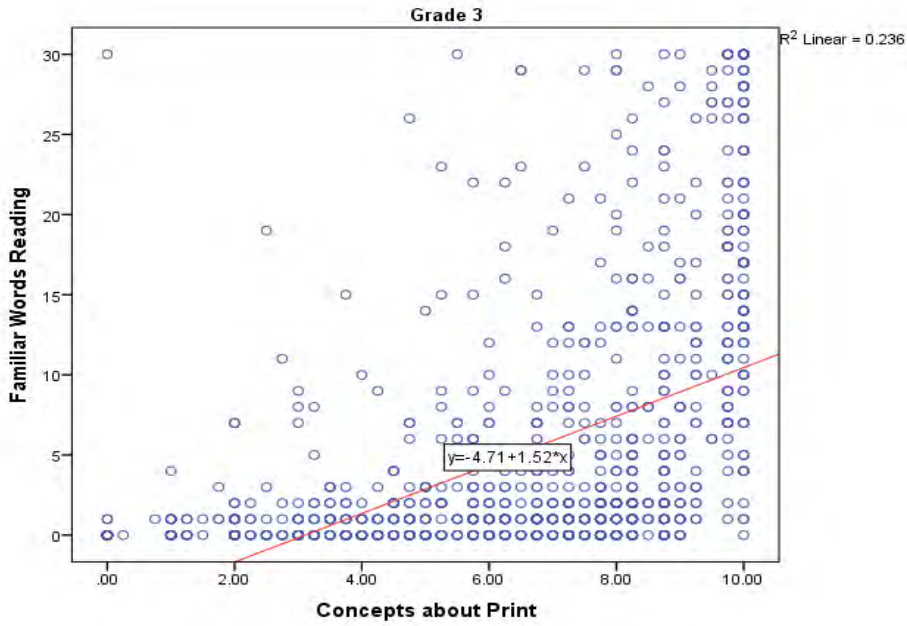


Concepts about Print (CAP) predicting Familiar Words Reading (FWR) – Grade 3

LR= -4.71+ 1.52CP R^2=23.6%

SE (0.51) (0.0.8)

P (0.00) (0.00)

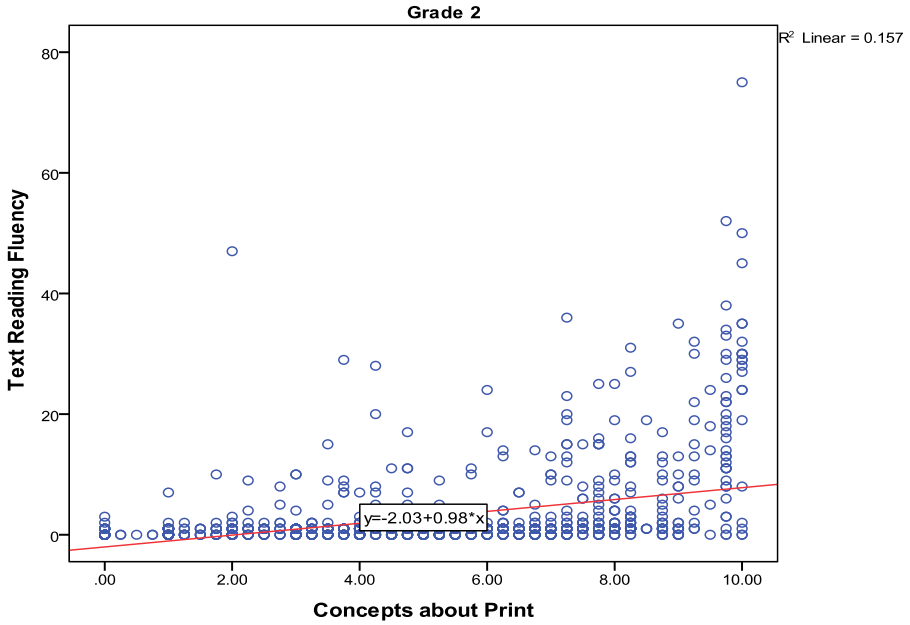


Concepts about Print (CAP) predicting Oral Reading Fluency (ORF) – Grade 2

TRF= -2.0+ 0.98CAP R^2=15.7%

SE (0.39) (0.06)

P (0.00) (0.00)

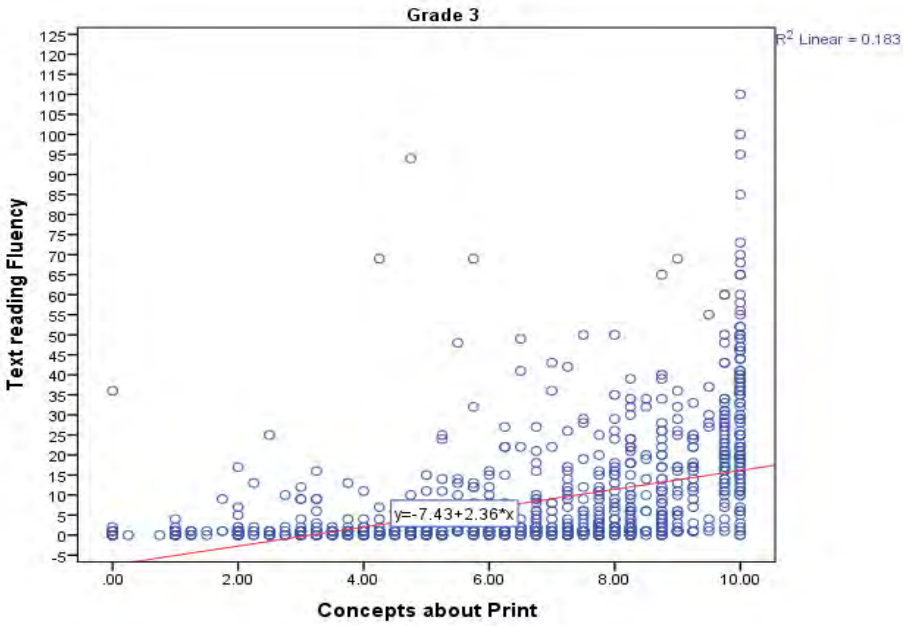


Concepts about Print (CAP) predicting Oral Reading Fluency (ORF) – Grade 3

ORF= -7.43+ 2.36 CAP R^2=18.3%

SE (1.06) (0.15)

P (0.00) (0.00)

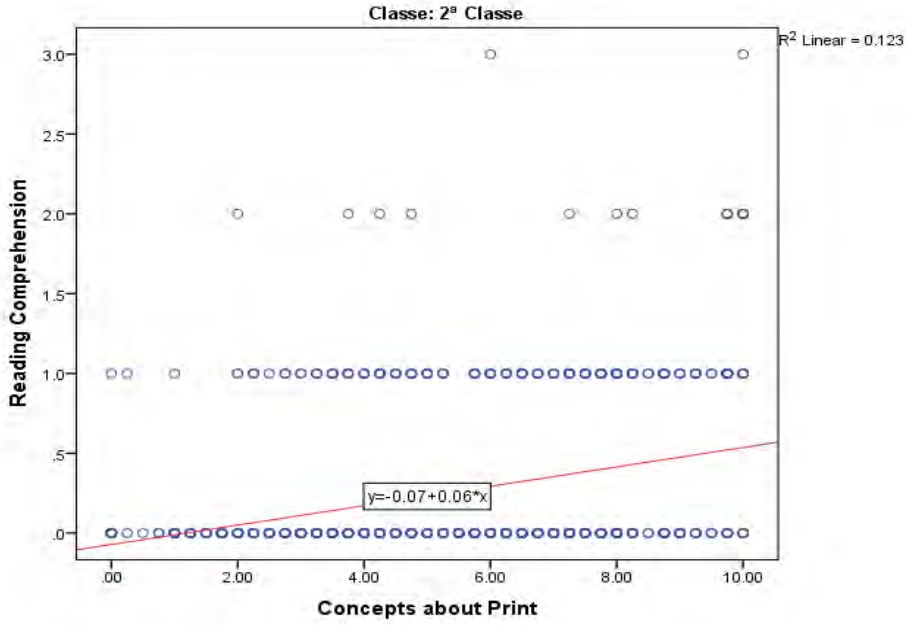


Concepts about Print (CAP) predicting Reading Comprehension (RC) – Grade 2

RC= -0.07+ 0.06CAP R^2=12.3%

SE (0.028) (0.006)

P (0.00) (0.00)

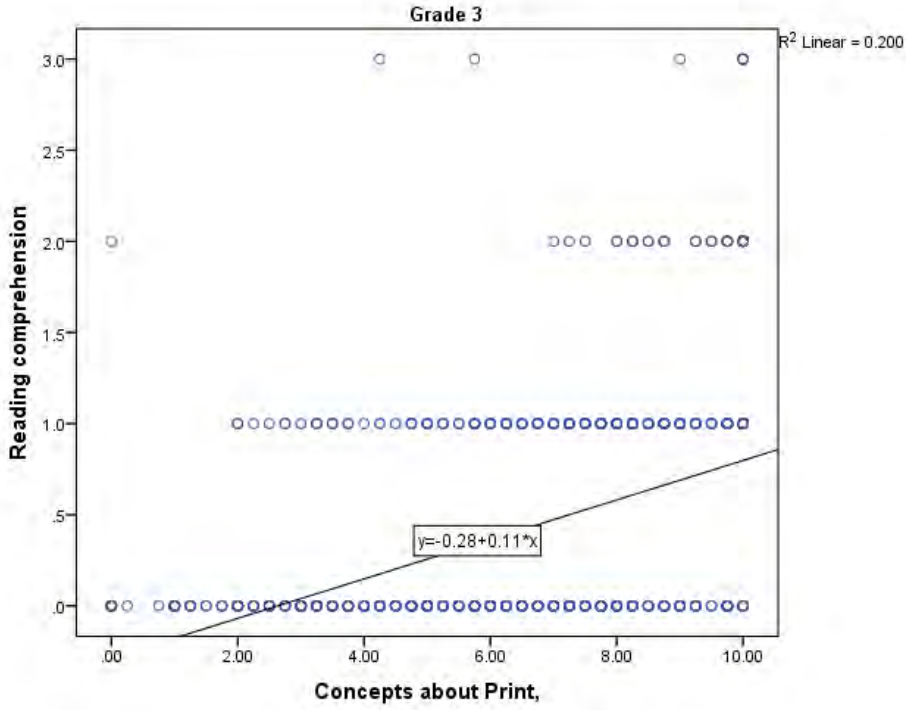


Concepts about Print (CAP) predicting Reading Comprehension (RC) – Grade 3

RC= -0.28+ 0.11CP R^2=20.0%

SE (0.046) (0.007)

P (0.00) (0.00)

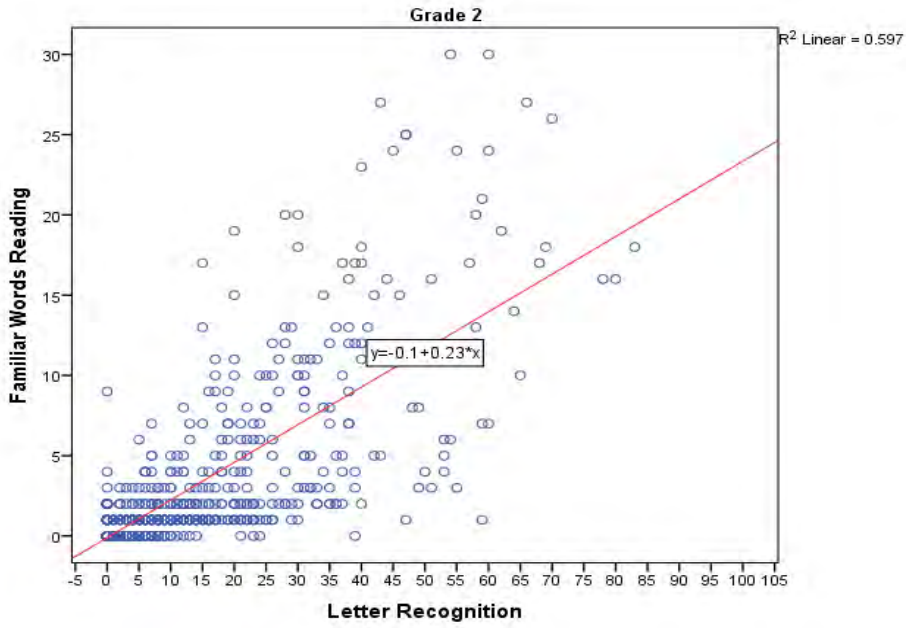


Letter Recognition (LR) predicting Familiar Words Reading (FWR) – Grade 2

FWR= -0.1+ 0.23LR **R^2=59.7% strong predictor**

SE (0.09) (0.006)

P (0.29) (0.00)

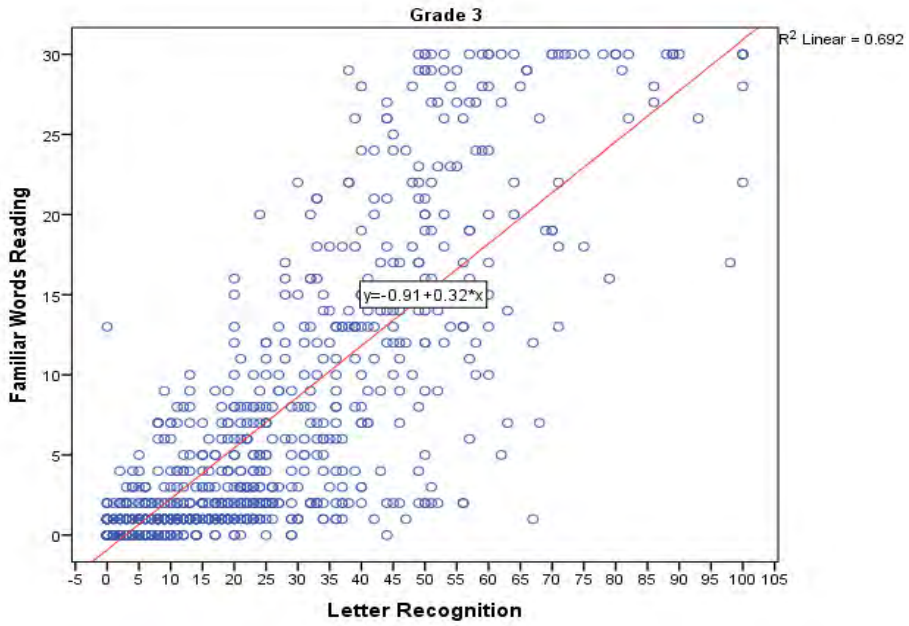


Letter Recognition (LR) predicting Familiar Words Reading (FWR) – Grade 3

FWR= -0.91+ 0.32LR **R^2=69.2% strong predictor**

SE (0.18) (0.006)

P (0.00) (0.00)

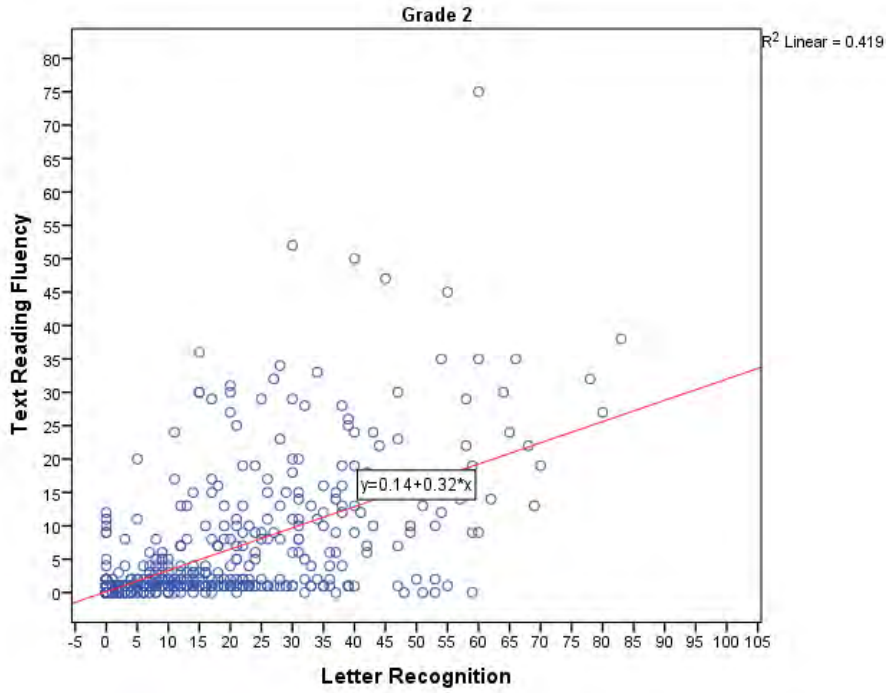


Letter Recognition (LR) predicting Oral Reading Fluency (ORF) – Grade 2

ORF= -0.14+ 0.32LR **R^2=41.9% moderately strong predictor**

SE (0.19) (0.01)

P (0.45) (0.00)

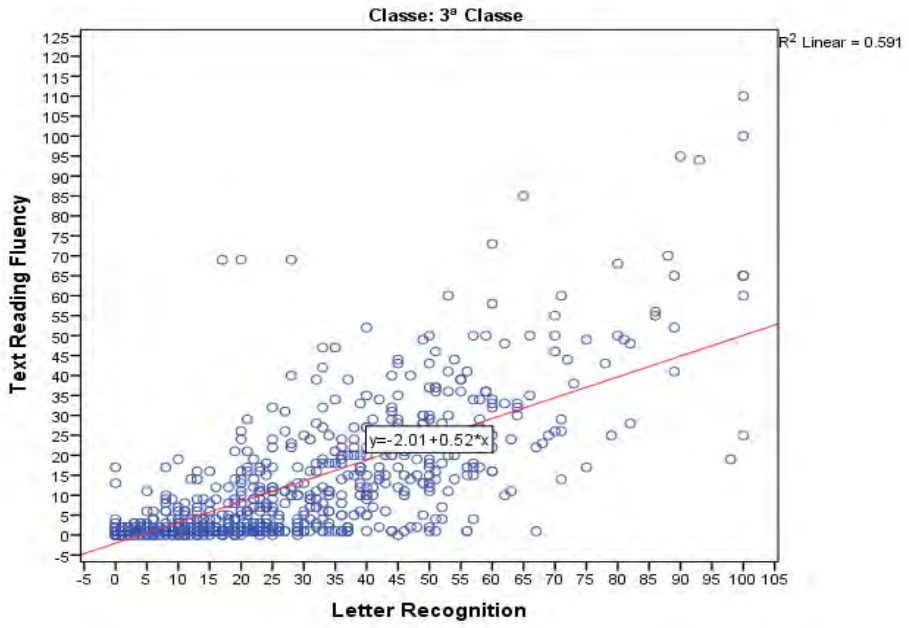


Letter Recognition (LR) predicting Oral Reading Fluency (ORF) – Grade 3

ORF= -2.01+ 0.52LR **R^2=59.1% strong predictor**

SE (0.38) (0.01)

P (0.00) (0.00)

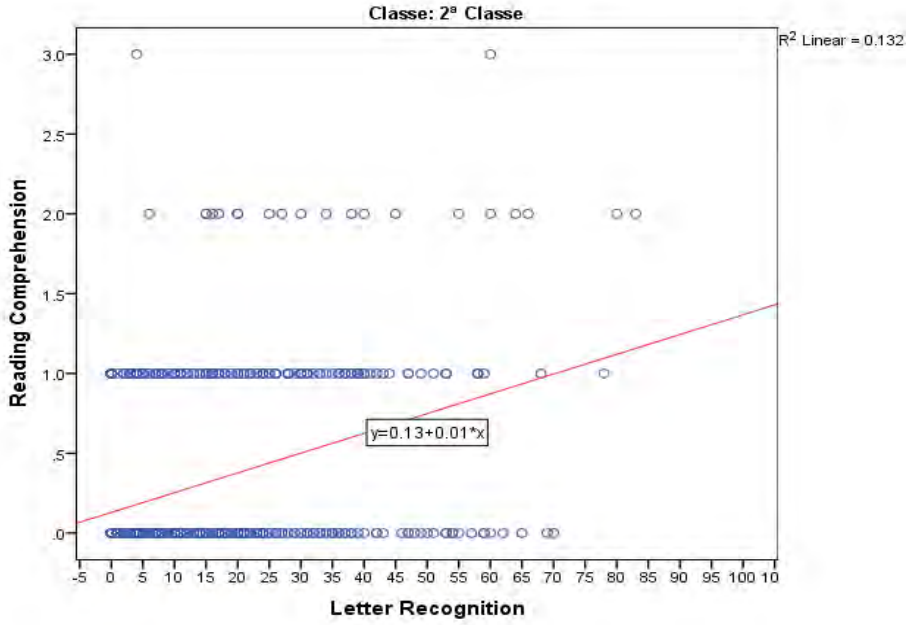


Letter Recognition (LR) predicting Reading Comprehension (RC) – Grade 2

RC= -0.13+ 0.01LR R^2=13.2%

SE (0.016) (0.001)

P (0.00) (0.00)

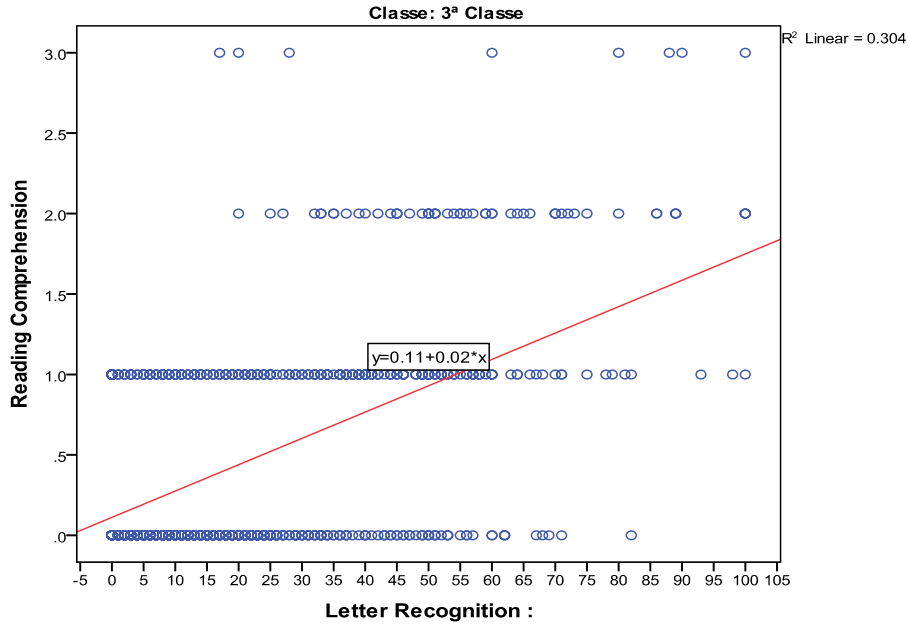


Letter Recognition (LR) predicting Reading Comprehension (RC) – Grade 3

RC= -0.11+ 0.02LR **R^2=30.4% moderate predictor**

SE (0.02) (0.001)

P (0.00) (0.00)

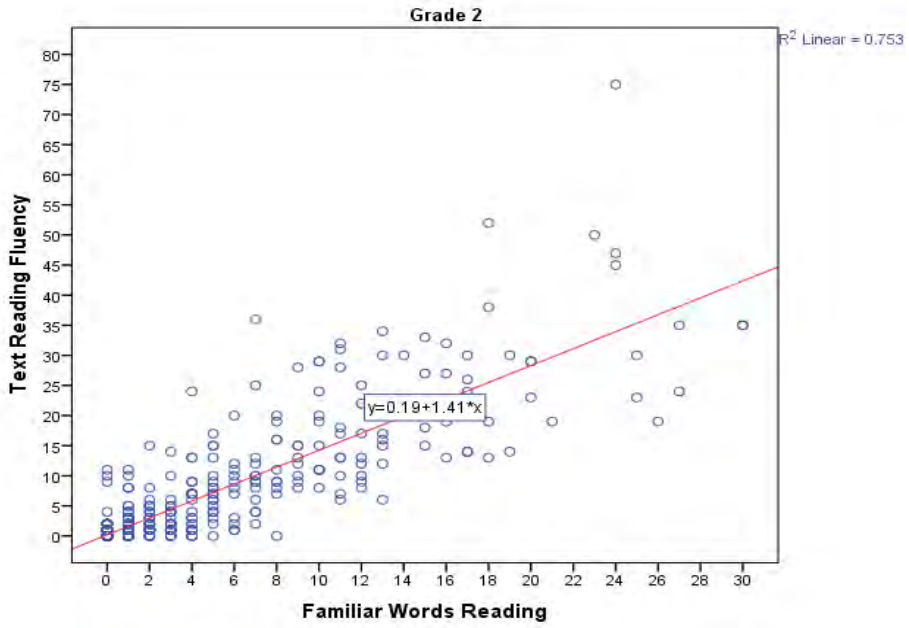


Familiar Word Reading (FWR) predicting Oral Reading Fluency (ORF) – Grade 2

TRF= -0.19+ 1.4FWR **R^2=75.3% strong predictor**

SE (0.11) (0.024)

P (0.09) (0.00)

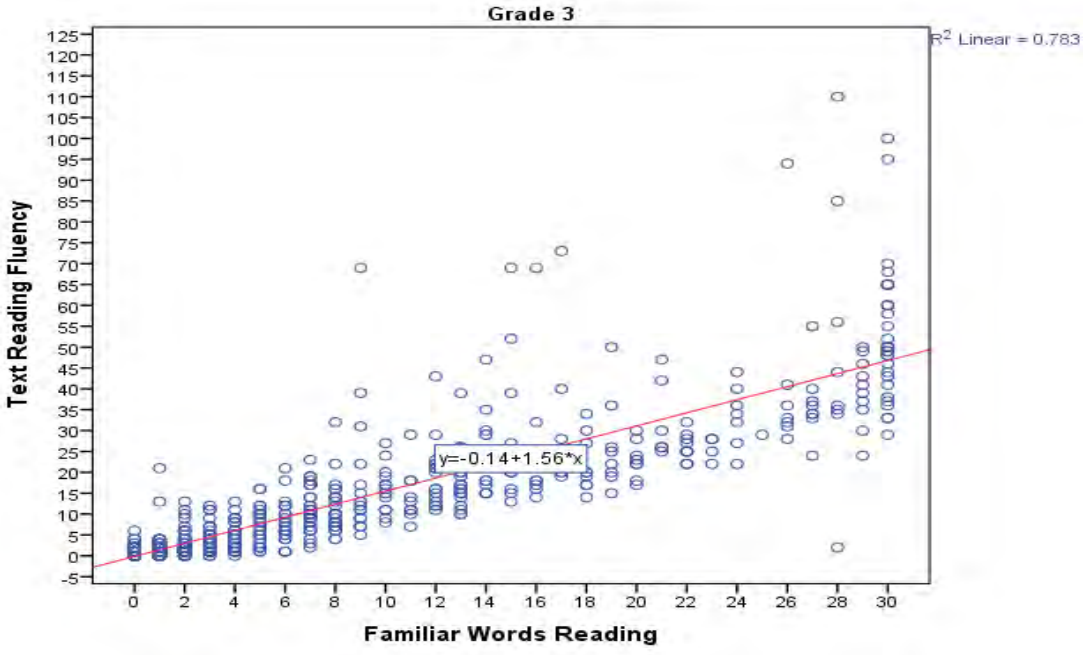


Familiar Word Reading (FWR) predicting Oral Reading Fluency (ORF) – Grade 3

ORF= -0.14+ 1.56FWR **R^2=78.3% strong predictor**

SE (0.23) (0.02)

P (0.56) (0.00)

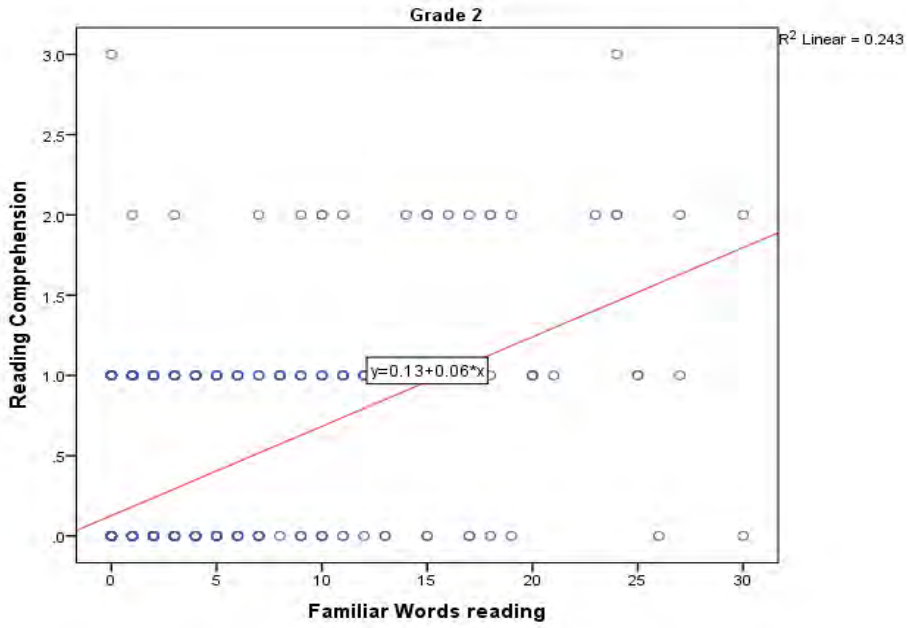


Familiar Word Reading (FWR) predicting Reading Comprehension (RC) – Grade 2

RC= -0.13+ 0.06FWR R^2=24.3%

SE (0.01) (0.003)

P (0.00) (0.00)

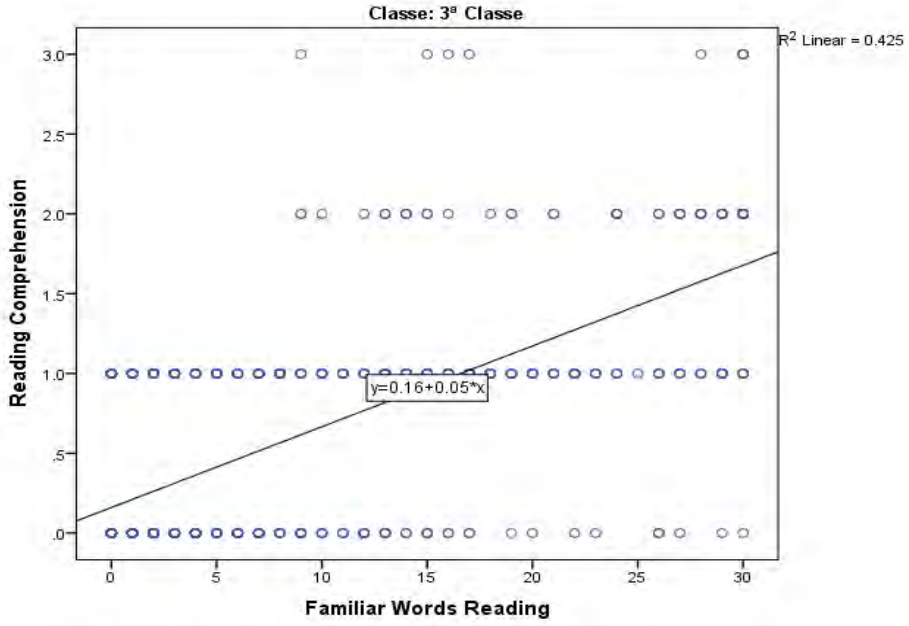


Familiar Word Reading (FWR) predicting Reading Comprehension (RC) – Grade 3

RC= -0.16+ 0.05FWR **R^2=42.5% moderately strong predictor**

SE (0.01) (0.02)

P (0.00) (0.00)

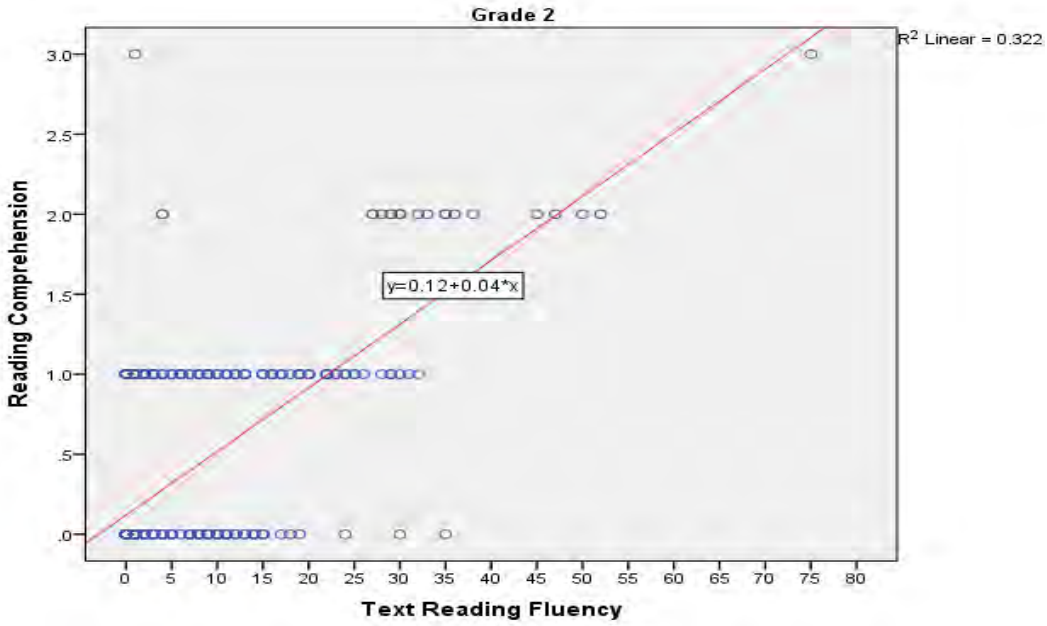


Oral Reading Fluency (ORF) predicting Reading Comprehension (RC) – Grade 2

RC= -0.12+ 0.04ORF **R^2=32.2% moderate predictor**

SE (0.01) (0.002)

P (0.00) (0.00)

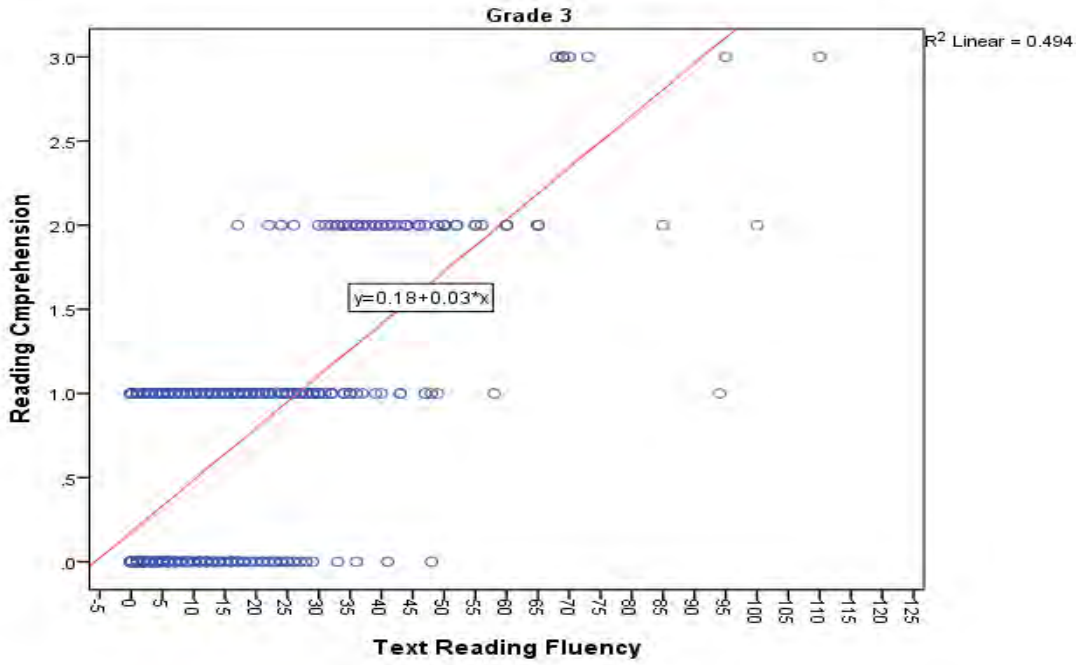


Oral Reading Fluency (ORF) predicting Reading Comprehension (RC) – Grade 3

RC= -0.18+ 0.03ORF **R^2=49.4% strong predictor**

SE (0.01) (0.001)

P (0.00) (0.00)



Annex H. Urban-Rural Differentiations

Annex H.1 Comparison of 2015 EGRA Mean Scores by Urban/Rural Locality

EGRA Subtest – Grade 2	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Oral Comprehension	8.6	6.6	11.62	728.30	**0.000**
Concepts about Print	5.7	4.4	8.60	766.62	**0.000**
Letter Recognition (lpm)	11.7	6.5	6.20	543.19	**0.000**
Common Words Read Correctly (wpm)	2.6	1.4	4.50	611.06	**0.000**
Text Reading Fluency (wpm)	3.7	2.2	3.68	679.72	**0.000**
Text Reading Comprehension	0.31	0.17	5.14	648.18	**0.000**

EGRA Subtest - Grade 3	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Oral Comprehension	9.7	7.6	13.83	826.67	**0.000**
Concepts about Print	7.4	5.8	11.83	826.88	**0.000**
Letter Recognition (lpm)	25.2	14.4	8.43	627.34	**0.000**
Common Words Read Correctly (wpm)	7.2	3.7	7.05	588.31	**0.000**
Text Reading Fluency (wpm)	10.8	5.7	5.93	596.23	**0.000**
Text Reading Comprehension	0.58	0.32	7.30	652.52	**0.000**

Annex H.2 Intermediate variables that may affect EGRA scores by Urban/Rural Locality

Other Variables	Urban	Rural	t	df	Sig. (2-tailed, unequal variances)
Student Age	8.6	9.2	-5.25	782.02	**0.000**
Student Days Attendance in July	11.9	8.3	6.20	673.74	**0.000**
Class Days Offered in July	20.9	11.8	22.67	1,166.74	**0.000**

Annex H.3 Use of Portuguese with Family and Friends

Use with Parents	Urban	Rural	Urban as a % of Rural
Almost Always	57.8%	20.2%	286.1%
Occasionally	14.8%	27.8%	53.2%
Almost Never	9.9%	27.2%	36.4%
Never	17.4%	24.7%	70.4%
N=1,530; Chi-square=200.92, df=3, p=0.000			
Portuguese Use with Siblings/Friends	**Urban**	**Rural**	**Urban as a % of Rural**
Almost Always	54.9%	21.2%	259.0%
Occasionally	16.9%	31.3%	54.0%
Almost Never	10.1%	25.0%	40.4%
Never	18.1%	22.4%	80.8%
N=1,675; Chi-square=183.35, df=3, p=0.000			

Annex H.4 Comparison of Treatment Group Effects within Urban/Rural Areas

Grade 2	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Oral Comprehension	Urban	8.7	8.9	7.7	0.789 NS	0.098 NS	**0.037**
	Rural	7.4	6.0	6.4	**0.000**	**0.000**	0.194 NS
Concepts about Print	Urban	5.9	5.6	5.2	0.558 NS	0.096 NS	0.428 NS
	Rural	5.3	4.0	4.1	**0.000**	**0.000**	0.949 NS
Letter Recognition	Urban	12.5	10.7	11.4	0.510 NS	0.852 NS	0.953 NS
	Rural	9.5	5.7	5.1	**0.000**	**0.000**	0.772 NS
Familiar Word Reading	Urban	2.6	2.2	3.2	0.770 NS	0.622 NS	0.341 NS
	Rural	2.3	1.1	1.0	**0.000**	**0.000**	0.917 NS
Text Word Reading Fluency	Urban	3.3	3.7	4.4	0.849 NS	0.548 NS	0.816 NS
	Rural	3.5	1.8	1.7	**0.001**	**0.000**	0.917 NS
Reading Comprehension	Urban	0.26	0.39	0.26	**0.045**	0.998 NS	0.167 NS
	Rural	0.24	0.15	0.13	**0.008**	**0.000**	0.772 NS

Grade 3	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Oral Comprehension	Urban	9.8	10.0	8.8	0.742 NS	**0.014**	**0.003**
	Rural	7.8	7.3	7.6	**0.000**	0.614 NS	0.282 NS
Concepts about Print	Urban	7.6	7.5	6.6	0.931 NS	**0.004**	**0.014**
	Rural	6.3	5.9	5.5	0.202 NS	**0.000**	**0.017**
Letter Recognition	Urban	26.5	23.8	24.7	0.546 NS	0.857 NS	0.963 NS
	Rural	16.4	17.2	10.8	0.860 NS	**0.000**	**0.000**
Familiar Word Reading	Urban	8.0	6.5	6.4	0.337 NS	0.997 NS	0.519 NS
	Rural	4.4	4.1	2.9	0.849 NS	**0.002**	**0.013**
Text Word Reading Fluency	Urban	12.2	9.6	9.7	0.480 NS	0.519 NS	1.000 NS
	Rural	6.8	6.5	4.3	0.908 NS	**0.006**	**0.013**
Reading Comprehension	Urban	0.63	0.55	0.51	0.450 NS	0.413 NS	0.940 NS
	Rural	0.42	0.28	0.27	**0.002**	**0.001**	0.992 NS

Annex H.5 Effects of Treatment Group on Grade 2 Urban Students - Detail

Student Age				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	182	8.74	0.132	Medium	0.27	.194	0.347 NS
				Control	0.25	.267	0.616 NS
Medium	142	8.47	0.141	Control	-0.02	.276	0.997 NS
Control	55	8.49	0.236	**Overall ANOVA**	**F**	**Significance**	
Total	379	8.60	0.089		1.10	0.333 NS	

Student Days Attended in July				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	167	12.35	0.639	Medium	0.98	1.048	0.618 NS
				Control	0.92	1.300	0.758 NS
Medium	110	11.37	0.832	Control	-0.06	1.385	0.999 NS
Control	58	11.43	1.174	**Overall ANOVA**	**F**	**Significance**	
Total	335	11.87	0.466		0.53	0.588 NS	
Class Days Offered in July				**Contrast With**			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	187	21.49	0.164	Medium	1.70	0.463	**0.001**
				Control	-0.08	0.578	0.990 NS
Medium	138	19.80	0.540	Control	-1.77	0.605	**0.010**
Control	70	21.57	0.212	**Overall ANOVA**	**F**	**Significance**	
Total	395	20.91	0.211		7.79	**0.000**	

Oral Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	208	8.65	0.220	Medium	-0.22	0.336	0.789 NS
				Control	0.91	0.441	0.098 NS
Medium	159	8.87	0.207	Control	1.13	0.457	**0.037**
Control	70	7.74	0.508	**Overall ANOVA**	**F**	**Significance**	
Total	437	8.59	0.153		3.14	**0.044**	

Concepts about Print				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	208	5.93	0.173	Medium	0.29	0.284	0.558 NS
				Control	0.77	0.372	0.096 NS
Medium	159	5.64	0.218	Control	0.48	0.386	0.428 NS
Control	70	5.16	0.370	**Overall ANOVA**	**F**	**Significance**	
Total	437	5.70	0.129		2.23	0.109 NS	

Letter Recognition (lpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	206	12.54	1.145	Medium	1.83	1.653	0.510 NS
				Control	1.17	2.162	0.852 NS
Medium	158	10.71	1.161	Control	-0.66	2.244	0.953 NS
Control	70	11.37	1.849	**Overall ANOVA**	**F**	**Significance**	
Total	434	11.68	0.750		0.63	0.533 NS	

Common Words Read Correctly (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	208	2.61	0.338	Medium	0.36	0.523	0.770 NS
				Control	-0.64	0.685	0.622 NS
Medium	159	2.25	0.356	Control	-1.00	0.712	0.341 NS
Control	70	3.24	0.730	**Overall ANOVA**	**F**	**Significance**	
Total	437	2.58	0.237		0.99	0.373 NS	

Text Reading Fluency (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	207	3.34	0.446	Medium	-0.41	0.753	0.849 NS
				Control	-1.03	0.987	0.548 NS
Medium	159	3.75	0.576	Control	-0.62	1.024	0.816 NS
Control	70	4.37	1.042	**Overall ANOVA**	**F**	**Significance**	
Total	436	3.65	0.342		0.57	0.566 NS	

Text Reading Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	207	0.26	0.033	Medium	-0.13	0.054	**0.045**
				Control	0.00	0.071	0.998 NS
Medium	159	0.39	0.045	Control	0.13	0.073	0.167 NS
Control	70	0.26	0.060	**Overall ANOVA**	**F**	**Significance**	
Total	436	0.31	0.025		3.27	**0.039**	

Portuguese Use with Parents	**Full**	**Medium**	**Control**
Almost Always	57.8%	61.5%	50.0%
Occasionally	19.5%	11.9%	7.8%
Almost Never	9.2%	11.9%	7.8%
Never	13.5%	14.8%	34.3%

N=384; Chi-square=20.18, df=6, p=0.003

Portuguese Use with Siblings/Friends	**Full**	**Medium**	**Control**
Almost Always	55.3%	56.5%	50.0%
Occasionally	19.7%	16.4%	10.0%
Almost Never	8.2%	14.5%	5.7%
Never	16.8%	12.6%	34.3%

N=437; Chi-square=21.49, df=6, p=0.001

Annex H.6 Effects of Treatment Group on Grade 2 Rural Students - Detail

Student Age				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	243	8.86	0.111	Medium	-0.52	0.172	**0.004**
				Control	-0.41	0.153	**0.019**
Medium	202	9.42	0.130	Control	-0.14	0.161	0.669 NS
Control	331	9.28	0.101	**Overall ANOVA**	**F**	**Significance**	
Total	776	9.18	0.065		5.92	**0.003**	

Student Days Attended in July				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	191	10.93	0.611	Medium	4.01	0.856	**0.000**
				Control	3.18	0.803	**0.000**
Medium	224	6.92	0.578	Control	-0.83	0.766	0.529 NS
Control	303	7.75	0.510	**Overall ANOVA**	**F**	**Significance**	
Total	718	8.34	0.330		12.16	**0.000**	
Class Days Offered in July				**Contrast With**			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	208	13.29	0.635	Medium	2.98	0.910	**0.003**
				Control	1.37	0.840	0.235 NS
Medium	237	10.31	0.607	Control	-1.61	0.807	0.114 NS
Control	346	11.92	0.535	**Overall ANOVA**	**F**	**Significance**	
Total	791	11.80	0.342		5.41	**0.005**	

Oral Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	357	7.36	0.158	Medium	1.34	0.220	**0.000**
				Control	0.99	0.206	**0.000**
Medium	377	6.02	0.141	Control	-0.35	0.203	0.194 NS
Control	504	6.37	0.140	**Overall ANOVA**	**F**	**Significance**	
Total	1,238	6.55	0.086		20.13	**0.000**	

Concepts about Print				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	357	5.25	0.150	Medium	1.20	0.196	**0.000**
				Control	1.15	0.184	**0.000**
Medium	377	4.04	0.131	Control	-0.06	0.181	0.949 NS
Control	504	4.10	0.116	**Overall ANOVA**	**F**	**Significance**	
Total	1,238	4.41	0.077		24.77	**0.000**	

Letter Recognition (lpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	337	9.54	0.849	Medium	3.83	0.929	**0.000**
				Control	4.48	0.868	**0.000**
Medium	360	5.71	0.515	Control	0.66	0.851	0.772 NS
Control	489	5.06	0.509	**Overall ANOVA**	**F**	**Significance**	
Total	1,186	6.53	0.360		14.48	**0.000**	

Common Words Read Correctly (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	357	2.29	0.262	Medium	1.18	0.267	**0.000**
				Control	1.28	0.250	**0.000**
Medium	377	1.11	0.130	Control	0.10	0.246	0.917 NS
Control	504	1.01	0.141	**Overall ANOVA**	**F**	**Significance**	
Total	1,238	1.41	0.104		14.87	**0.000**	

Text Reading Fluency (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	355	3.47	0.448	Medium	1.65	0.454	**0.001**
				Control	1.79	0.426	**0.000**
Medium	377	1.82	0.248	Control	0.15	0.418	0.936 NS
Control	503	1.68	0.223	**Overall ANOVA**	**F**	**Significance**	
Total	1,235	2.24	0.176		10.08	**0.000**	

Text Reading Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	353	0.24	0.027	Medium	0.09	0.031	**0.008**
				Control	0.11	0.029	**0.000**
Medium	377	0.15	0.020	Control	0.02	0.028	0.772 NS
Control	503	0.13	0.016	**Overall ANOVA**	**F**	**Significance**	
Total	1,233	0.17	0.012		8.11	**0.000**	

Portuguese Use with Parents	**Full**	**Medium**	**Control**
Almost Always	30.7%	9.2%	21.0%
Occasionally	33.1%	26.4%	25.1%
Almost Never	21.2%	33.0%	27.2%
Never	14.9%	31.3%	26.8%

N=1,146; Chi-square=73.02, df=6, p=0.000

Portuguese Use with Siblings/Friends	**Full**	**Medium**	**Control**
Almost Always	30.5%	10.6%	22.6%
Occasionally	36.7%	30.5%	28.2%
Almost Never	19.6%	28.1%	26.6%
Never	13.2%	30.9%	22.6%

N=1,238; Chi-square=71.51, df=6, p=0.000

Annex H.7 Effects of Treatment Group on Grade 3 Urban Students - Detail

Student Age				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	195	10.26	0.139	Medium	0.74	0.193	**0.000**
				Control	0.44	0.263	0.217 NS
Medium	158	9.53	0.131	Control	-0.30	0.270	0.514 NS
Control	62	9.82	0.218	**Overall ANOVA**	**F**	**Significance**	
Total	415	9.92	0.090		7.39	**0.001**	

Student Days Attended in July				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	174	15.05	0.517	Medium	1.79	0.918	0.217 NS
				Control	0.93	1.038	0.642 NS
Medium	101	13.26	0.852	Control	-0.86	1.141	0.733 NS
Control	70	14.11	0.792	**Overall ANOVA**	**F**	**Significance**	
Total	345	14.33	0.396		1.94	0.145 NS	

Class Days Offered in July				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	185	19.91	0.271	Medium	-0.83	0.382	0.078 NS
				Control	-1.03	0.478	0.080 NS
Medium	139	20.74	0.286	Control	-0.20	0.499	0.914 NS
Control	70	20.94	0.309	**Overall ANOVA**	**F**	**Significance**	
Total	394	20.39	0.173		3.48	**0.032**	

Oral Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	205	9.82	0.179	Medium	-0.21	0.287	0.742 NS
				Control	1.06	0.377	**0.014**
Medium	160	10.03	0.186	Control	1.27	0.390	**0.003**
Control	70	8.76	0.450	**Overall ANOVA**	**F**	**Significance**	
Total	435	9.73	0.132		5.55	**0.005**	

Concepts about Print				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	205	7.58	0.152	Medium	0.09	0.243	0.931 NS
				Control	1.02	0.319	**0.004**
Medium	160	7.49	0.172	Control	0.93	0.331	**0.014**
Control	70	6.56	0.346	**Overall ANOVA**	**F**	**Significance**	
Total	435	7.39	0.112		5.36	**0.005**	

Letter Recognition (lpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	204	26.48	1.774	Medium	2.68	2.554	0.546 NS
				Control	1.78	3.350	0.857 NS
Medium	160	23.79	1.757	Control	-0.91	3.466	0.963 NS
Control	70	24.70	2.990	**Overall ANOVA**	**F**	**Significance**	
Total	434	25.20	1.160		0.57	0.566 NS	

Common Words Read Correctly (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	205	7.96	0.714	Medium	1.42	1.007	0.337 NS
				Control	1.53	1.321	0.997 NS
Medium	160	6.54	0.697	Control	0.11	1.368	0.519 NS
Control	70	6.43	1.083	**Overall ANOVA**	**F**	**Significance**	
Total	435	7.19	0.458		1.26	0.286 NS	

Text Reading Fluency (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	205	12.15	1.266	Medium	2.56	1.742	0.480 NS
				Control	2.49	2.281	0.519 NS
Medium	159	9.59	1.114	Control	-0.07	2.364	1.000 NS
Control	70	9.66	1.974	**Overall ANOVA**	**F**	**Significance**	
Total	434	10.81	0.792		1.29	0.278 NS	

Text Reading Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	205	0.63	0.048	Medium	0.09	0.072	0.450 NS
				Control	1.20	0.094	0.413 NS
Medium	159	0.55	0.052	Control	0.03	0.098	0.940 NS
Control	70	0.51	0.088	**Overall ANOVA**	**F**	**Significance**	
Total	434	0.58	0.033		1.15	0.317 NS	

Portuguese Use with Parents	**Full**	**Medium**	**Control**
Almost Always	60.5%	65.2%	56.7%
Occasionally	22.6%	18.1%	15.0%
Almost Never	7.9%	9.4%	13.3%
Never	9.0%	7.2%	15.0%

N=375; Chi-square=6.31, df=6, p=0.390 NS

Portuguese Use with Siblings/Friends	**Full**	**Medium**	**Control**
Almost Always	68.8%	62.5%	54.3%
Occasionally	18.5%	26.3%	12.9%
Almost Never	4.9%	6.9%	12.9%
Never	7.8%	4.4%	20.0%

N=435; Chi-square=25.80, df=6, p=0.000

Annex H.8 Effects of Treatment Group on Grade 3 Rural Students - Detail

Student Age				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	298	10.00	0.098	Medium	-0.60	0.148	**0.000**
				Control	-0.46	0.137	**0.002**
Medium	294	10.60	0.112	Control	0.13	0.138	0.603 NS
Control	404	10.45	0.089	**Overall ANOVA**	**F**	**Significance**	
Total	996	10.36	0.057		9.13	**0.000**	

Student Days Attended in July				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	164	13.46	0.629	Medium	4.49	0.884	**0.000**
				Control	4.39	0.832	**0.000**
Medium	250	8.97	0.526	Control	-0.10	7.280	0.990 NS
Control	351	9.07	0.503	**Overall ANOVA**	**F**	**Significance**	
Total	765	9.98	0.324		16.37	**0.000**	
Class Days Offered in July				**Contrast With**			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	185	18.23	0.582	Medium	7.44	0.948	**0.000**
				Control	5.46	0.881	**0.000**
Medium	251	10.78	0.648	Control	-1.98	0.800	**0.036**
Control	370	12.77	0.534	**Overall ANOVA**	**F**	**Significance**	
Total	806	13.40	0.358		32.25	**0.000**	

Oral Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	359	7.79	0.166	Medium	0.51	0.221	**0.000**
				Control	0.20	0.209	0.614 NS
Medium	383	7.28	0.145	Control	-0.31	0.205	0.282 NS
Control	491	7.59	0.138	**Overall ANOVA**	**F**	**Significance**	
Total	1,233	7.55	0.086		2.72	0.066 NS	

Concepts about Print				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	359	6.28	0.144	Medium	0.33	0.195	0.202 NS
				Control	0.83	0.184	**0.000**
Medium	383	5.95	0.130	Control	0.50	0.181	0.017
Control	491	5.45	0.121	**Overall ANOVA**	**F**	**Significance**	
Total	1,233	5.85	0.076		10.54	**0.000**	

Letter Recognition (lpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	353	16.44	1.011	Medium	-0.71	1.361	0.860 NS
				Control	5.65	1.288	**0.000**
Medium	369	17.15	0.977	Control	6.36	1.272	**0.000**
Control	470	10.79	0.800	**Overall ANOVA**	**F**	**Significance**	
Total	1,192	14.43	0.536		15.54	**0.000**	

Common Words Read Correctly (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	359	4.39	0.382	Medium	0.27	0.485	0.849 NS
				Control	1.54	0.459	**0.002**
Medium	383	4.13	0.340	Control	1.27	0.450	**0.013**
Control	491	2.85	0.273	**Overall ANOVA**	**F**	**Significance**	
Total	1,233	3.70	0.189		6.81	**0.001**	

Text Reading Fluency (wpm)				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	357	6.83	0.713	Medium	0.36	0.863	0.908 NS
				Control	2.25	0.817	**0.006**
Medium	383	6.46	0.612	Control	2.16	0.800	**0.019**
Control	490	4.31	0.454	**Overall ANOVA**	**F**	**Significance**	
Total	1,230	5.71	0.336		5.91	**0.003**	

Text Reading Comprehension				Contrast With			
Treatment	**Valid N**	**Mean**	**Std. Error**	**Treatment**	**Mean Dif.**	**Std. Error**	**Tukey Sig.**
Full	357	0.42	0.033	Medium	0.14	0.041	**0.002**
				Control	0.14	0.039	**0.001**
Medium	381	0.28	0.028	Control	0.01	0.038	0.992 NS
Control	490	0.27	0.024	**Overall ANOVA**	**F**	**Significance**	
Total	1,228	0.32	0.016		8.16	**0.000**	

Portuguese Use with Parents	**Full**	**Medium**	**Control**
Almost Always	30.3%	16.3%	26.2%
Occasionally	33.1%	27.9%	34.3%
Almost Never	19.4%	27.0%	17.8%
Never	17.2%	28.8%	21.6%

N=1,103; Chi-square=35.88, df=6, p=0.000

Portuguese Use with Siblings/Friends	**Full**	**Medium**	**Control**
Almost Always	32.9%	19.1%	27.9%
Occasionally	32.0%	31.6%	35.4%
Almost Never	20.3%	23.5%	16.7%
Never	14.8%	25.8%	20.0%

N=1,233; Chi-square=31.32, df=6, p=0.000

Annex H.9 Student Days Attended in July 2015 by Grade, Area and Treatment Group

Grade	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Grade 2	Urban	12.4	11.4	11.4	0.618 NS	0.758 NS	0.999 NS
	Rural	10.9	6.9	7.8	**0.000**	**0.000**	0.529 NS
Grade 3	Urban	15.1	13.3	14.1	0.217 NS	0.642 NS	0.733 NS
	Rural	13.5	9.0	9.1	**0.000**	**0.000**	0.990 NS

Annex H.10. Total Enrollment, Grades 2 and 3, by Urban-Rural Treatment

Data provided by ApaL from MINEDH sources. We are providing these for ease in reference for possible reader analysis.

Treatment Group	Total Schools	Total Enrollment		Average Enrollment per School	
		Grade 2	Grade 3	Grade 2	Grade 3
Full	60	13,012	11,726	217	195
Medium	65	12,511	11,365	192	175
Control	54	6,466	5,472	120	101
Total	179	31,989	28,563	179	160

Treatment Group	Urban Schools	Total Urban Enrollment		Average Urban Enrollment per School	
		Grade 2	Grade 3	Grade 2	Grade 3
Full	24	8,227	7,698	343	321
Medium	17	6,956	6,687	409	393
Control	6	1,065	998	178	166
Total	47	16,248	15,383	346	327

Treatment Group	Rural Schools	Total Rural Enrollment		Average Rural Enrollment per School	
		Grade 2	Grade 3	Grade 2	Grade 3
Full	36	4,785	4,028	133	112
Medium	48	5,555	4,678	116	97
Control	48	5,401	4,474	113	93
Total	132	15,741	13,180	119	100

Annex H.11 Enrollment, Grades 2 and 3, by Urban-Rural Treatment - Detail

		Total Enrollment by Grade and Sex						Average Enrollment per School by Grade and Sex					
		Grade 2			Grade 3			Grade 2			Grade 3		
Treatment	Schools	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total
Full	60	6,531	6,481	13,012	5,981	5,745	11,726	109	108	217	100	96	195
Medium	65	6,429	6,082	12,511	5,940	5,425	11,365	99	94	192	91	83	175
Control	54	3,096	3,370	6,466	2,666	2,806	5,472	57	62	120	49	52	101
Total	**179**	**16,056**	**15,933**	**31,989**	**14,587**	**13,976**	**28,563**	**90**	**89**	**179**	**81**	**78**	**160**

		Total Urban Enrollment by Grade and Sex						Average Enrollment per Urban School by Grade and Sex					
		Grade 2			Grade 3			Grade 2			Grade 3		
Treatment	Schools	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total
Full	24	4,163	4,064	8,227	4,028	3,670	7,698	173	169	343	168	153	321
Medium	17	3,658	3,298	6,956	3,714	2,973	6,687	215	194	409	218	175	393
Control	6	556	509	1,065	527	471	998	93	85	178	88	79	166
Total	**47**	**8,377**	**7,871**	**16,248**	**8,269**	**7,114**	**15,383**	**178**	**167**	**346**	**176**	**151**	**327**

		Total Rural Enrollment by Grade and Sex						Average Enrollment per Rural School by Grade and Sex					
		Grade 2			Grade 3			Grade 2			Grade 3		
Treatment	Schools	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total
Full	36	2,368	2,417	4,785	1,953	2,075	4,028	66	67	133	54	58	112
Medium	48	2,771	2,784	5,555	2,226	2,452	4,678	58	58	116	46	51	97
Control	48	2,540	2,861	5,401	2,139	2,335	4,474	53	60	113	45	49	93
Total	**132**	**7,679**	**8,062**	**15,741**	**6,318**	**6,862**	**13,180**	**58**	**61**	**119**	**48**	**52**	**100**

	Girls Percentage of Total Enrollment by Grade		Girls Percentage of Urban Enrollment by Grade		Girls Percentage of Rural Enrollment by Grade	
Treatment	**Grade 2**	**Grade 3**	**Grade 2**	**Grade 3**	**Grade 2**	**Grade 3**
Full	50.2%	51.0%	50.6%	52.3%	49.5%	48.5%
Medium	51.4%	52.3%	52.6%	55.5%	49.9%	47.6%
Control	47.9%	48.7%	52.2%	52.8%	47.0%	47.8%
Total	**50.2%**	**51.1%**	**51.6%**	**53.8%**	**48.8%**	**47.9%**

Annex H.12 Retention of Teacher Practices

Teacher Practices	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
Teacher-Student Interaction	Urban	8.2	8.0	4.9	0.970 NS	**0.028**	**0.052**
	Rural	7.7	8.0	6.9	0.900 NS	0.353 NS	0.175 NS
Teaches Decoding	Urban	4.4	3.5	1.4	0.418 NS	**0.021**	0.164 NS
	Rural	4.9	3.8	1.4	0.071 NS	**0.000**	**0.000**
Teaches Comprehension	Urban	4.3	3.2	5.8	0.436 NS	0.423 NS	0.123 NS
	Rural	4.2	4.0	3.4	0.972 NS	0.238 NS	0.363 NS
Planning and Sequencing	Urban	3.6	3.5	2.6	0.992 NS	0.481 NS	0.554 NS
	Rural	3.5	3.6	2.7	0.963 NS	0.122 NS	0.072 NS

Annex H.13 Retention of Portuguese Reading Booklets

Grades 2 and 3	Area	Full	Medium	Control	Sig. Full-Medium	Sig. Full-Control	Sig. Med.-Control
% Students with Reading Book	Urban	30.7%	21.1%	12.1%	0.635 NS	0.475 NS	0.851 NS
	Rural	54.6%	41.9%	7.0%	0.194 NS	**0.000**	**0.000**

Annex H.14 Retention of Classroom Inventory Materials (TLA)

Classroom Inventory	Area	Full	Medium	Control	Chi-Square	df	Sig. (p =)
Permanent Alphabet Chart	Urban	34.2%	19.2%	0.0%	4.848	2	0.089 NS
	Rural	61.7%	51.8%	6.8%	49.738	2	**0.000**
Materials to Create Words from Letters	Urban	26.3%	19.2%	0.0%	2.838	2	0.242 NS
	Rural	41.7%	39.3%	8.1%	23.675	2	**0.000**
Movable Letters or Words	Urban	28.9%	19.2%	0.0%	3.415	2	0.181 NS
	Rural	46.7%	30.4%	4.1%	32.958	2	**0.000**
Fixed Letter or Word Posters	Urban	23.7%	7.7%	0.0%	4.673	2	0.097 NS
	Rural	28.3%	28.6%	4.1%	17.507	2	**0.000**
Permanent Blackboard	Urban	89.5%	76.9%	87.5%	1.932	2	0.381 NS
	Rural	95.0%	94.6%	91.9%	0.664	2	0.717 NS
Student-made Materials Displayed	Urban	7.9%	0.0%	0.0%	2.801	2	0.246 NS
	Rural	6.7%	5.4%	0.0%	4.776	2	0.092 NS
Teacher-made Materials Displayed	Urban	21.1%	11.5%	0.0%	2.703	2	0.259 NS
	Rural	40.0%	25.0%	5.4%	23.416	2	**0.000**
Non sex-based Seating	Urban	76.3%	76.9%	75.0%	0.013	2	0.994 NS
	Rural	71.7%	82.1%	78.4%	1.886	2	0.389 NS
Row or Group Seating	Urban	92.1%	96.2%	100.0%	1.012	2	0.603 NS
	Rural	88.3%	92.9%	87.8%	0.974	2	0.615 NS

Annex I. References

Abadzi, H. (2013). *"Literacy for All in a 100 days?"* Global Partnership for Education. Series on learning No. 7 (2009). "Instructional Time Loss in Developing Countries: Concepts, Measurement, and Implications." World Bank Research Observer. 24 (2): 267-290

Aga Khan Foundation (2010) Cabo Delgado, Mozambique EQUIP2

Aggarwala, N.K. (2004). Evaluation Report: Quality assessment of primary and middle education in mathematics and science. Retrieved from http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/Aggarwala_UNDP_Evaluation_Report.pdf. Accessed 2013 June 15.

Bruns, B., Mingat, A., & Rakotomalala, R. (2003). Achieving universal primary education by 2015: A chance for every child. Washington, D.C: The World Bank. doi:10.1596/0-8213-5345-4

Cellini, Stephanie Riegg and James Edwin Kee, "Cost-Effectiveness and Cost-Benefit Analysis," Chapter 25 of Handbook of Practical Program Evaluation, Third Edition, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer. San Francisco: Jossey-Bass, 2010.

Collins, P. and Messaoud-Galusi, S. (2012). Student Performance on the Early Grade Reading Assessment (EGRA) in Mozambique. EdData II report prepared by RTI International for USAID. Retrieved from http://www.eddataglobal.org/documents/index.cfm/4-

Cutler I. End games: The challenge of sustainability. Baltimore: Annie E. Casey Foundation; 2002.

Dhaliwal, I. *et al.* (2011). Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. MIT Abdul Latif Jameel Poverty Action Lab (J-PAL), August 10, 2011

Dijkman, M., Harting, J. van Tol, L. & van der Wal, M. Sustainability of the good behaviour gamein Dutch primary schools

Drummond, M.F. *et al.* (2005). Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press.

Evans, D (2016). That zero effect may not mean what you think it means, and other lessons from recent educational research. Impact Evaluations, 1/21/2016 (http://blogs.worldbank.org/impactevaluations)

Fiedler *et al.* (2008). An activity-based cost analysis of the Honduras community-based, integrated child care (AIN-C) programme. Health policy and planning, 23:408–427.

Fuchs D, Fuchs LS, Burish P. Peer-assisted learning strategies: An evidence-based practice to promote reading achievement. Learning Disabilities Research and Practice. 2000;15:85–91.

Gavin, S. (March, 2011). Literacy boost: Mozambique baseline report. Retrieved from http://resourcecentre.savethechildren.se/library/literacy-boost-Mozambiquebaseline-report. Accessed 2013 June 20.

Glewwe, Paul (2004) "An Investigation of the Determinants of School Progress and Academic Achievement in Vietnam," in P. Glewwe, D. Dollar and N Agrawal, eds., Economic Growth, Poverty, and Household Welfare in Vietnam. The World Bank. Washington, DC.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin (2002) "Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya," Development Research Group, World Bank, Washington, DC.

Glewwe, Paul and Karthik Muralidharan (2015). Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. RISE (Research on Improving Systems of Education) Working Paper October 2015

Global Education Digest 2012. Opportunities Lost: The impact of grade repetition and early school leaving. UNESCO Institute of Statistics. Montreal, Canada, www.uis.unesco.org

Gold *et al.* (1996). Cost-effectiveness in health and medicine. New York: Oxford University Press.

Gove, A. and P. Cvelich (2011). Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute www.eddataglobal.org

Johnson K, Hayes C, Center H, Daley C. Building capacity and sustainable prevention innovations: A sustainability planning model. Evaluation and Program Planning. 2004;27:135–149.

Levin and McEwan (2001). Cost-effectiveness analysis: methods and applications. 2nd ed. Thousand Oaks, CA: Sage.

McEwan, Patrick J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. Journal of Development Effectiveness 4:189

Michaelowa, K. (2001). Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations, World Development, 2001

Mongoi, D. et al. (2010). "Endline Report of Early Literacy among pre-school and primary school children in Mozambique." Save the Children.

Racine DP. Reliable effectiveness: A theory on sustaining and replicating worthwhile innovations. Administrative Policy in Mental Health and Mental Health Services Research. 2006;33:356–387.

Raupp, M., Newman, B. and Revés, L. (2013). *Impact Evaluation for the USAID/Aprender a Ler Project: Baseline Report*.

Raupp, M., Newman, B. and Lauchande, C. (2014) *Impact Evaluation for the USAID/Aprender a Ler Project: Midline Report.*

Rogers EM. Diffusion of innovations. 5. New York: Free Press; 2003.

Rubenstein M, Patrikakou E, Weissberg R, Armstrong M. Enhancing school–family partnerships: A teacher's guide. Chicago: Department of Psychology, University of Illinois at Chicago; 2000.

Scheirer MA. Is sustainability possible? A review and commentary on empirical studies on program sustainability. American Journal of Evaluation. 2005;26:320–347

Spaull, N. and Taylor, S. (2015) "Access to What? Creating a Composite Measure of Educational Quantity and Educational Quality for 11 African Countries," *Comparative Education Review* 59:133-165

Tatto, M.T., Nielsen, H.D., Cummings, W.C., Kularatna, N.G., & Dharmadasa, D.H. (1991). Comparing the effects and costs of different approaches for educating primary school teachers: The case of Sri Lanka. BRIDGES Research Report Series (10), Cambridge, Massachusetts: Harvard Institute for International Development.

UN Special Envoy for Global Education (April 2013). *Accelerating progress to 2015: Mozambique*. Working paper; World Bank at http://data.worldbank.org/indicator/SE.SEC.NENR?order=wbapi_data_value_2012+wbapi_data_value&sort=asc

World Bank (2015). Mozambique Service Delivery Indicators: Education

Yin RK. Life histories of innovations: How new practices become routinized. Public Administration

Annex J. Conflict of Interest Statements

Disclosure of Conflict of Interest for USAID Evaluation Team Members

Name	Magdala Raupp
Title	TEAM LEADER
Organization	IBTCI
Evaluation Position	(X) Team Leader (X) Team member
Evaluation Award Number *(contract or other instrument)*	AID-656-C-12-00002
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	Impact Evaluation of the USAID Aprender a Ler (ApaL) Project in Mozambique, AID-656-C-12-00002
I have real or potential conflicts of interest to disclose.	( ) Yes (X) No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* *1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* *2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* *3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* *4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* *5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* *6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	4/7/2015

ANNEX . CONFLICT OF INTEREST DISCLOSURE

Name	[illegible]
Title	[illegible]
Organization	IBTCI
Evaluation Position?	☐ Team Leader ☒ Team member
Evaluation Award Number *(contract or other instrument)*	RAN-I-00-09-00016/AID-656-TO-12-00002 AID-656-C-15-00002
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID Mozambique – Aprender a Ler World Education, Inc. AID-656-C-12-00001
I have real or potential conflicts of interest to disclose.	☐ Yes ☒ No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to* *1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated* *2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* *3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* *4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* *5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* *6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

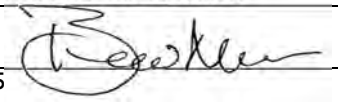
...ity (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	[signature]
Date	[illegible] 2016

Disclosure of Conflict of Interest for USAID Evaluation Team Members

Name	Bruce Newman
Title	Statistician/Senior Data Analyst
Organization	IBTCI
Evaluation Position	( ) Team Leader (X) Team member
Evaluation Award Number *(contract or other instrument)*	RAN-I-00-09-00016/AID-656-TO-12-00002 AID-656-C-15-00002
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	Impact Evaluation of the USAID Aprender a Ler (ApaL) Project in Mozambique, AID-656-C-12-00001
I have real or potential conflicts of interest to disclose.	( ) Yes (X) No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* *1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* *2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* *3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* *4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* *5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* *6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	4/7/2015

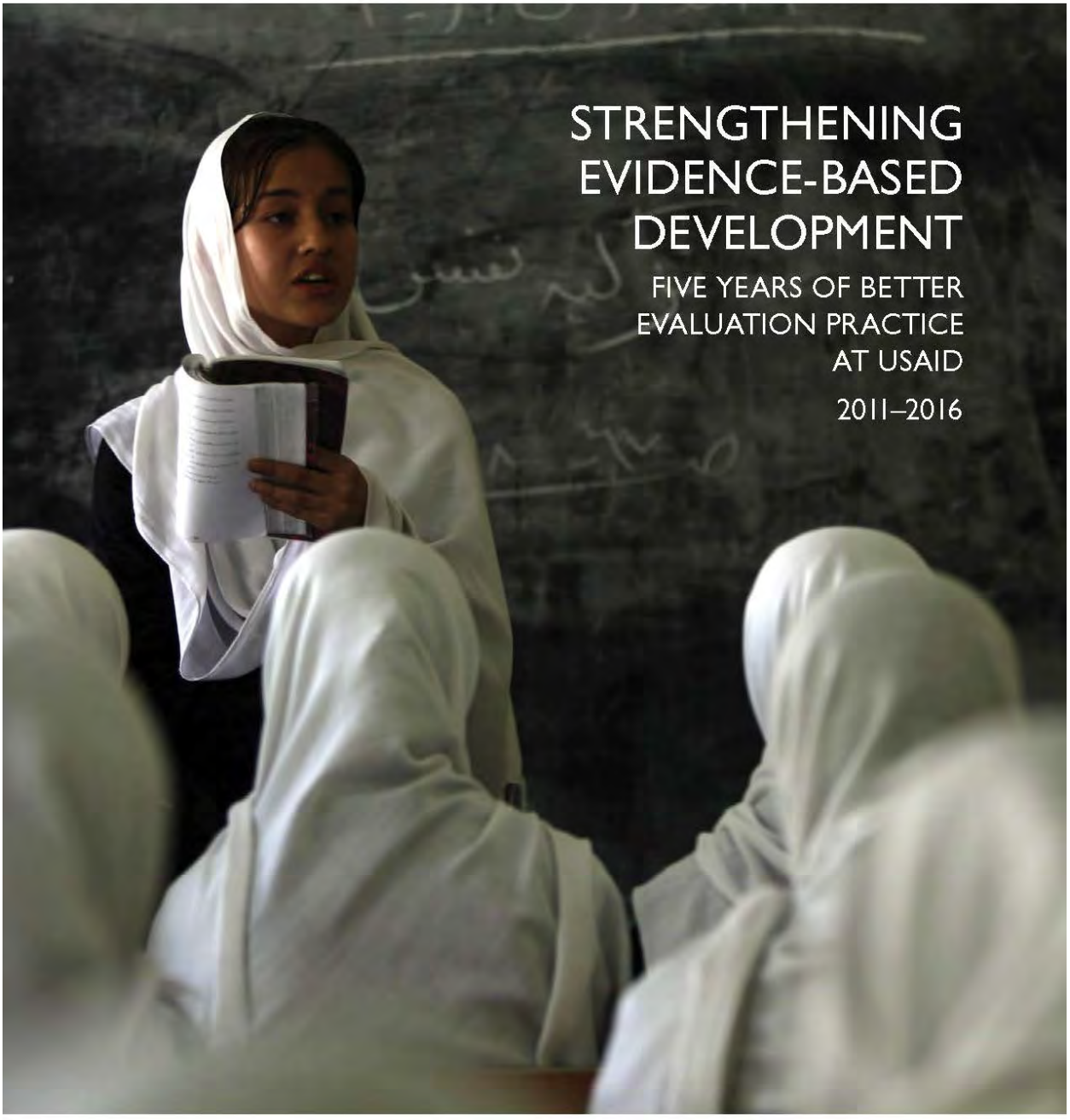
ANNEX . CONFLICT OF INTEREST DISCLOSURE

Name	Carlos Saavedra
Title	NA.
Organization	IBTCI
Evaluation Position?	☐ Team Leader ☒ Team member
Evaluation Award Number *(contract or other instrument)*	RAN-I-00-09-00016/AID-656-TO-12-00002 AID-656-C-15-00002
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	USAID Mozambique – Aprender a Ler World Education, Inc. AID-656-C-12-00001
I have real or potential conflicts of interest to disclose.	☐ Yes ☒ No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* *1. Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* *2. Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* *3. Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* *4. Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* *5. Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* *6. Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	Saavedra
Date	04/01/2016

Annex K. USAID Perspectives on the Impact Evaluation



USAID
FROM THE AMERICAN PEOPLE

USING EVIDENCE TO ADAPT EDUCATION PROGRAMMING

USAID/Mozambique: Helping Children to Read

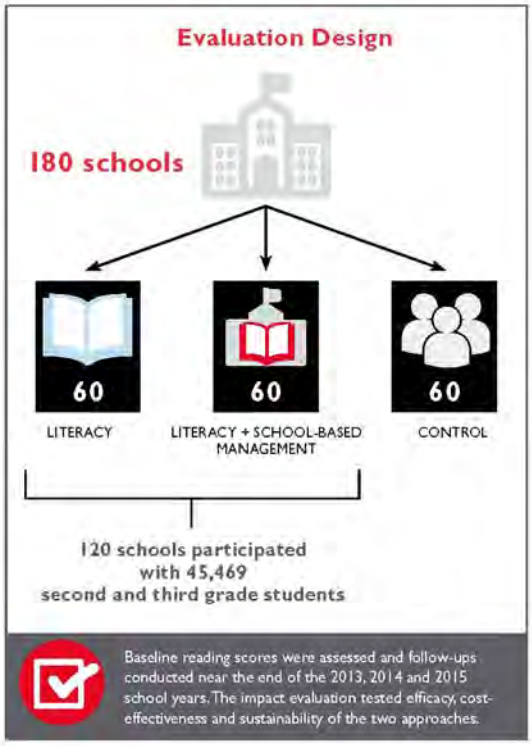
Program Description

In 2012, the USAID mission in Mozambique and the Government of Mozambique collaborated to launch Aprender a Ler (Learn to Read), an education project primarily focused on improving reading outcomes for second and third grade students in more than 1,000 urban and rural schools in the Nampula and Zambézia provinces of Mozambique. The program sought to achieve two outcomes: an increase in the quality of reading instruction through in-school coaching, teacher training and distribution of reading materials, and an increase in the quantity of reading instruction in school by strengthening school management practices to maximize the value of instruction time in school.

At the start of the program, USAID commissioned an impact evaluation to collect data on the reading skills of second and third grade students. The main purpose of the impact evaluation was to assess the effectiveness of the intervention and the cost-effectiveness and sustainability of two intervention levels. The two levels included medium treatment, which focused on improved student learning, and full treatment, which added components for more effective school management.

Results

After the first school year, students from all three groups improved their reading skills. Students who received assistance doubled their average word-per-minute reading. Students in schools with school management increased reading by 287 percent versus 221 percent for the schools that received reading instruction only. Contributing to this were significant improvements in attendance by both students and teachers. One year after USAID support ended, students continued to show improvements in critical aspects of reading. Additionally, the full treatment group, incorporating both literacy and school-based management, was also the more cost-effective approach.



Action Taken

Because USAID incorporated the impact evaluation into the program design, the implementing partner could quickly adjust its programming based on the results and recommendations. At the request of the Government of Mozambique, USAID expanded the full program to an additional 538 schools with 109,021 students and 2,002 teachers.

USAID-2016/03/30

THE BROOKINGS INSTITUTION

SAUL/ZILKHA ROOM

ASSESSING THE IMPACT OF FOREIGN ASSISTANCE:
THE ROLE OF EVALUATION

A CONVERSATION WITH USAID ADMINISTRATOR GAYLE SMITH

Washington, D.C.

Wednesday, March 30, 2016

PARTICIPANTS:

Introduction and Moderator:

GEORGE INGRAM
Senior Fellow, Global Economy and Development
The Brookings Institution

Keynote Speaker:

GAYLE SMITH
Administrator
USAID

Panelists:

RUTH LEVINE
Director, Global Development and Population Program
Hewlett Foundation

WADE WARREN
Assistant to the Administrator for the Bureau
for Policy, Planning, and Learning (PPL)
USAID

* * * * *

ANDERSON COURT REPORTING
706 Duke Street, Suite 100
Alexandria, VA 22314
Phone (703) 519-7180 Fax (703) 519-7190

few months. But, quite frankly, it's irreversible.

So there will be a plan to do that. We've blocked out what we want to get done over the next 10 months, this is one of the blocks. And Wade is a very kind gentleman. He and his team know that this is a priority, which means we're going to do it, so stay tuned.

MR. COONROD: Thank you. John Coonrod with The Hunger Project. I have a question about the difference because it's very obvious the difference that it's made to USAID as a leading development practitioner in the world. How has the increased evaluations impacted the program countries, the governments, the policymakers? Has there been a visible difference made through these evaluations with the formation of national policies?

MS. SMITH: I believe there has. I'm more anecdotal than systematic on that, and Wade may want to speak more to that on the panel. But if you look at Mozambique, if you look at Ethiopia, if you look at Kenya, there are a number of cases where on the basis of an evaluation or evaluations and, quite frankly, a good relationship, our teams in the field have been able to sit down with ministers or other counterparts and say, look, here's what we're finding. Therefore, we suggest a change in this or that direction. And often that change is made.

Sometimes at the project level, a couple of times already -- and I say "already" because five years is both a long time and a short time -- we've seen changes to governments' national policies, so that it may extend beyond the reach of our particular assistance or program.

MR. INGRAM: One more question, Gayle?

MS. SMITH: Sure.

MR. INGRAM: Right here.

that's when you can use sometimes impact evaluations, essentially have social experiments, that are conducted in a way that permits you to understand what the net impact is of a particular set of interventions.

So it is, in fact, more scientific. It's only appropriate in some cases. It needs to be thought of from the very beginning and the whole project has to be designed that way. So, obviously, that constrains the use to a pretty limited set of questions, but they are super important questions to be asking.

And it's also, as I said earlier, a way for this agency to use its vast resources and knowledge to contribute to better development practice among many others, not just USAID.

MR. WARREN: I was thinking maybe I could give you an example, George, of when we did an impact evaluation. We tried two different approaches to a problem and found out that one of them worked better than the other one, and what we did about it.

So in Mozambique, we had an education sector, a childhood reading program. We were working with second and third graders trying to increase their reading levels. And there were two ways the project was implemented.

The first one was just with the interventions around the reading materials and working with the teachers on instructing the children. The second one that was done -- and also we worked with the school administration on broader management issues in the school itself. We did a midterm impact evaluation. We found that the reading outcomes were much, much stronger in the second case where we were working both with the teachers and children, but also with the school administration.

So we then consulted with the government of Mozambique. I think there was a question about how do our evaluations are used by the countries where we work.

We consulted with the government of Mozambique and based on that consultation we expanded the project in two ways, from not just second and third graders, but also first graders, and then scaled it up from 120 schools to more than 1,000 schools in Mozambique. So that was a way that we learned on the fly from a project and changed the way we were implementing across a broader number of sites.

MR. INGRAM: Good, thanks. One more question before I turn to the audience and ask you all to take over the microphone. Learning, obviously we all think that it's very important for evaluations to be used for learning. Gayle noted the tension between accountability and learning.

Wade, talk to us about what AID -- how they're instituting feedback loops, what you're doing to take the knowledge and information from evaluations and then sharing it within the agency and beyond the agency. And, Ruth, you can second guess him by then describing what an ideal learning system is. What are the different pieces that you would have in a learning system in order to maximize the value of the learning.

MR. WARREN: Okay. So I guess first I should say that I'm glad we're talking about learning. And for those of you that know the history of the agency well, PPL, the Bureau for Policy, Planning, and Learning, is a successor bureau for a previous bureau that was called Policy, Planning, and Coordination. And I love the fact that the name was intentionally changed in its new incarnation to have "learning" as part of the title of the bureau because it is so important to what we do.

That having been said, I think I alluded earlier and I think Gayle also talked about it is probably the area where we need to do the most work, to make sure that the findings from evaluations are understood by people who are making decisions about project design and strategic planning and use in an appropriate way.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523