USAID | LIBERIA
FROM THE AMERICAN PEOPLE

EGRA Plus: Liberia

Data Analytic Report: EGRA Plus: Liberia Baseline Assessment

Early Grade Reading Assessment (EGRA) Plus: Liberia
EdData II Task Number 6
Contract Number EHC-E-06-04-00004-00
Strategic Objective 3
April 30, 2009

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International and the Liberian Education Trust.

EGRA Plus: Liberia

Data Analytic Report:

EGRA Plus: Liberia Baseline Assessment

Contract EHC-E-06-04-00004-00
April 30, 2009

Prepared for
USAID/Liberia

Prepared by
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

Table of Contents

Figures

Tables

1. Executive Summary

In November 2008, RTI International and its subcontractor—Liberian Education Trust—collaborated with Liberian education officers to collect a nationally representative comprehensive baseline early grade reading assessment in grades 2 and 3. A total of 176 schools were assessed, including 60 control, 59 full treatment and 57 light treatment schools, for a total of 2957 students. Students were assessed in a full battery of early grade reading tasks, including letter naming fluency, phonemic awareness, familiar word fluency, unfamiliar word fluency, connected text oral reading fluency, reading comprehension, and listening comprehension. Analysis of the Early Grade Reading Assessment (EGRA) itself shows that the assessment is reliable and the various subtasks that assess different parts of the underlying early grade reading skills tie together well as a reliable test. In fact, the Cronbach's alpha results show reliability of 0.85, which is quite good.[1]

Descriptive and bivariate analyses of the EGRA reading subtasks show that Liberian children were capable of identifying the names of letters for the most part, with the average grade 2 child identifying 54.9 letters in a minute and the average grade 3 child identifying 66.4 letters. Student achievement on tasks that require phonemic awareness was not as strong, however. Grade 2 students were able to identify 6.1 familiar and 1.6 unfamiliar words, with grade 3 students at 12.7 and 3.0 words, respectively. International benchmarks for grade 2 unfamiliar word decoding, for what they are worth, require 30 and 50 words to be classified as emerging and established, respectively, both far above the Liberian average of 1.6.

Given the importance of oral reading fluency skills in future academic achievement and the ability to move from learning to read and reading to learn, much of this report focuses on oral reading fluency levels and the impact of various predictor variables on this construct. Grade 2 students read 14.5 words per minute correctly, while grade 3 students read 25.1 words. These averages mask significant variation, since approximately a third of both grade 2 and grade 3 students were unable to read any words correctly at all. Reading comprehension scores, as we would expect, depend heavily on the ability to accurately read the text the questions are based on. The average student was able to answer 1.25 of the 5 reading comprehension questions, and for listening comprehension, 1.68 of the three items.

The large sample size allows more precision in the estimation of differences between grades and gender. In all subtasks, grade 3 students scored statistically significantly higher than grade 2 students, with more than 10 additional words read correctly per minute on the oral reading fluency subtask. Likewise, boys outperformed girls on

[1] Cronbach's alpha is a measure of how well a set of variables (in this case, Early Grade Reading Assessment subtasks) measure an underlying construct (in this case, early grade reading skill). In short, it is a measure of test reliability.

nearly every subtask, suggesting the need to control for gender in further analyses.[2] Multiple regression analysis suggests that student and school-level predictors are capable of predicting oral reading fluency scores but the few teacher pedagogical variables were unable to do so. In summary, the early grade reading skills of the Liberian children assessed in grades 2 and 3 were relatively low. The research design allows for comparisons between control and two different treatment groups, both to estimate the impact of the program and to assess ways to improve the program design at midpoints during program implementation.

2. Introduction

During November 2008, RTI International, in collaboration with Liberian Education Trust and Liberian education officers, undertook a national baseline assessment of early grade literacy skills. This baseline was performed in 176 schools for a total of 2957 students, and the results are presented in this report. [3] This baseline data collection was part of the USAID-funded Early Grade Reading Assessment Plus program being implemented during 2008–2010, designed with an experimental intervention. The intervention was part of a joint collaboration among the Liberian Ministry of Education, the World Bank, and USAID/Liberia, and was initiated by a pilot assessment in June 2008.

This experiment, entitled EGRA Plus: Liberia, uses empirical data from early grade reading assessments to track progress toward improvements of quality of early grade reading instruction, with particular focus on phonics-based instruction. The research and intervention design allows for the comparison of three groups against each other. The first is a control group which will receive no program interventions. The second group, the "light" intervention, is a set of schools where parents and community members are provided data on the quality of their school's literacy instruction based on student achievement score. The final group, the "full" intervention, provides an intensive teacher-training based program targeting reading instructional strategies combined with information on student achievement provided to parents and communities.

The assessment was randomly implemented across a total of 176 schools, with a target of 60 in each control, light, and full intervention school type.[4] In each school, either 10 or 20 students were assessed, depending on the size of the school and

[2] Controlling for gender in further analyses, particularly in multiple regression, means that statistically you are able to account for the gender impacts, and the estimates of the other variables of interest can be interpreted without the interference of gender.

[3] The missing 4 schools were assessed in January and February 2009, but were not included into the data analysis.

[4] The sampling procedure used in this study and in the intervention is a means of identifying the true impact of the program. Without having a counterfactual, a comparison group, it is impossible to know whether any impacts we see are the result of program effects, typical growth over the course of the school year, or changes that apply to all students equally. Having a control group allows us to differentiate among those possibilities. In this case, there is one control group and two experimental groups—one having a full intervention and one a light intervention, basically meaning that parents are given information regarding the quality of their students' skills in reading.

number of teachers. The assessment itself has several components, which have been tested in a variety of other low-income countries, as well as the June 2008 pilot assessment in Liberia.

First, children were asked to demonstrate some basic reading awareness knowledge skills. Next, in an allotted amount of time, children were asked to quickly name a set of letters. Next, sampled children were assessed on their phonemic awareness skills, followed by an assessment of their ability to identify a set of commonly used words. Then, children were asked to identify a set of unfamiliar words using phonics-based skills, followed by an assessment of children's ability to read a short passage. This subtask was followed by a set of reading comprehension questions, and the full assessment was completed with an assessment of a child's listening comprehension skills. The baseline assessment was supplemented by a set of student-level questions investigating home and school backgrounds using predictors the literature suggests are related to literacy achievement. Additionally, teacher, director, and school level predictors were collected.

Analysis of the EGRA itself shows that the assessment is reliable and the various subtasks assess different parts of the underlying early grade reading skills as well as tying together well as a reliable test. In fact, as noted in Section 1, the Cronbach's alpha results show reliability of 0.85, which is quite good.

Beginning portions of the analytic report lay out the various subtasks of the assessment, and point out how they are related to important characteristics of early reading skills and proficiency. The analysis presented here focuses on a particular set of research questions designed to inform the early stages of the program intervention as well as to provide a baseline of early grade reading skills across Liberia. This analysis presents a full description of reading achievement and supplements the pilot findings presented to the World Bank in June 2008.[5] This analytic report is organized in the following way.

- First, descriptive statistics are presented for both predictor and outcome variables. Then, we compare these descriptive statistics across important characteristics, particularly student gender, treatment group, and grade level.
- Second, we assess the reliability of the assessment itself using a variety of statistical methods and follow this by presenting correlations of relevant variables. Finally, we present graphic depictions of student achievement across various metrics and present some models that predict student achievement in early reading.

3. Early Grade Intervention in Reading

Ensuring fidelity to the research and intervention experimental design, the EGRA Plus: Liberia intervention builds on the findings from the 2008 World Bank EGRA

[5] Crouch, L., & Korda, M. (2008). *EGRA Liberia: Baseline assessment of reading levels and associated factors.* Report prepared as part of a process of collaboration between USAID and the World Bank. Research Triangle Park, North Carolina: RTI International. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=158

report with a three-stage intervention process to improve the quality of student achievement. First, a baseline EGRA was implemented in a nationally representative set of Liberian primary schools. This assessment will serve as the baseline for the impact evaluations, but also will inform the intervention itself, taking student achievement evidence as the first step in assessing teacher training needs, and developing teacher professional development courses to respond to the critical learning areas for improving student achievement. As part of the EGRA Plus: Liberia program, this report serves as the initial analysis of early grade reading skills.

Second, RTI International in collaboration with Liberian Education Trust is implementing a teacher professional development model program that encompasses intensive, week-long capacity-building workshops using early grade teaching techniques, ongoing professional development, external support, and internal processes and procedures. The intervention is buttressed with activities designed to foster community action and stakeholder participation, particularly around the production and dissemination of EGRA findings reports at various stages in the EGRA Plus intervention, along with the fostering of interactive meetings between school managers and community members. This set of community action activities serves as the main intervention in light intervention schools, while full intervention schools also receive the professional development and supervision support for grade 2 and 3 teachers.

While there are other activities building capacity at all relevant levels of government, the third major intervention activity is an additional two rounds of EGRA in Liberia, allowing for a truly longitudinal research design. This design will allow researchers and the Ministry of Education to identify whether and how the interventions have had a significant impact on student achievement, as well as by what causal mechanisms the program was successful.

4. Research Design

As highlighted in the sections above, the research design has three groups: control, light intervention, and full intervention. This design, developed in consultation between the Ministry of Education, the World Bank, and USAID, allows for a sophisticated analysis of both the impact of the programs and the causal mechanisms at work. For example, while it is useful to know that a set of interventions focused on early grade achievement was successful, it is even more useful to be able to differentiate the relative impacts of fostering community involvement and direct teacher professional development inputs. The matching of the achievement data with a full set of family, teacher, principal, and community variables will allow for a significant amount of predictive ability regarding the mechanisms at work in improving student achievement. It is worth noting that EGRA Plus: Liberia will follow three groups of 60 schools, randomly assigned, over a two-year period. This will lend richness to the information gathered, relevant to local policy, as well as to the knowledge of the impacts of these types of reading interventions.

Table 1 shows the achieved sample assessed in this baseline report. Note that while 60 schools were targeted in all three groups (control, light and full intervention), 59 and 57 schools were actually assessed in the full intervention and control groups, respectively. Note that, given the fewer schools assessed in the control and full intervention schools, it is not surprising that fewer students were assessed in both groups. However, at the student level, nearly 3000 students were assessed (2970), allowing for a rich data set that allows some precision in the tests implemented.

Table 1. Pretest sampling

	TREATMENT			
	Light	Full	Control	TOTAL
Schools	60	59	57	176
Students	1030	980	951	2970

Table 2 shows that there were more boys sampled than girls, 54.99% to 45.01%. Assessors conducted random sampling of students and the difference in the numbers of boys and girls sampled can be attributed to random differences. Evidence suggests that it might simply be an artifact of Liberian primary schools, since education management information system (EMIS) enrollment data from across the nation's primary schools suggest that 53.08% of the population is boys. The EGRA sample is quite similar to the findings from the EMIS enrollment data. It might simply be that girls are more likely to drop out more quickly as they traverse primary school, or that girls' persistence rates are lower. In the sampled schools, for example, the difference between grade 2 and grade 3 enrollments was much higher for girls than for boys (16.3% rather than 1.8%). The cross-sectional nature of these data limits their usefulness, however, since we are unsure whether this is a characteristic of this particular cohort of children or of a systemic difference in persistence over time. Much more data are necessary to determine this, although the later EGRA assessments will be able to shed some light on these questions. Note that for analytic purposes, holding gender constant is more than sufficient for controlling for any gender dropout effect, which in any case, is unlikely to adversely affect our ability to make claims about the quality of reading skills.

Table 2. Pretest sample by gender and level

	Treatment				Level		
GENDER	Control	Full	Light	TOTAL	Grade 2	Grade 3	TOTAL
Boys	543	512	571	1626	816	801	1617
Girls	406	462	463	1331	720	603	1323
TOTAL	949	974	1034	2957	1536	1404	2940

5. EGRA Subtask Descriptions

In order to introduce the reader to the Early Grade Reading Assessment tool, this section briefly introduces the various subtasks, so that the analysis below will be meaningful. The EGRA tool consists of a variety of subtasks or "testlets," and they have been somewhat differentially applied in various countries in order to ensure context-specific relevance. The EGRA Plus: Liberia tool[6] assessed the following set of skills:

1. *Print orientation*: awareness of the direction of text, and the knowledge that a reader should read down the page.
2. *Letter naming fluency*: ability to read the letters of the alphabet without hesitation and naturally. This is a timed test that assesses automaticity and fluency of letter recognition. It is timed to one minute, which saves time and also prevents children from having to spend time on something they are having a hard time with.
3. *Phonemic awareness*: awareness of how sounds work with words. This is generally considered a pre-reading skill, and can be assessed in a variety of ways. In the case of Liberia this was assessed by asking the student which word, out of three, starts with a different sound (e.g., *ball*, in "mouse, ball, moon").
4. *Familiar word recognition and oral reading fluency*: ability to read high-frequency words. This assesses whether children can process words quickly. It is timed to one minute.
5. *Unfamiliar or nonsense words, oral reading fluency*: ability to process words that could exist in the language in question, but do not, or are likely to be very unfamiliar. The nonwords used for EGRA are truly made-up words. The subtask assesses the child's ability to "decode" words fluently. It is timed to one minute.
6. *Connected text oral reading fluency*: ability to read a passage, about 60 words long, that tells a story. It is timed to one minute.
7. *Comprehension in connected text*: ability to answer five questions based on the passage read.
8. *Listening comprehension*: being able to follow and understand a simple oral story. This assesses the child's ability to concentrate and focus to understand a very simple story of three sentences with simple noninferential (factual) questions. It is considered a pre-reading skill.

6. Descriptive Statistics

In this section, we briefly present the findings from our descriptive statistical analysis. ***Table 3*** below presents a simple set of statistics, including the number of observations, statistical mean, standard deviation, and minimum and maximum values. This table indicates that Liberian children were relatively successful in naming

[6] The baseline assessment instruments for students, teachers, and principals are available from the EdData II website, https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=159.

letters (mean=61.16), although slightly less successful than in the June 2008 assessment, which may be attributable to random difference or perhaps the possibility of some knowledge loss due to the summer holidays (Crouch & Korda, 2008). Children struggled with naming familiar words (mean=9.26) and struggled quite a bit with unfamiliar words (mean=2.24). The combination of these findings suggests that while students did know the names of letters, they were less capable of manipulating the sounds associated with the letters. This suggests that some targeting of phonemic awareness skills, particularly phonics patterns, might increase children's ability to adapt phonemic awareness skills to the decoding of unfamiliar words, which is a critical skill for successful reading. Average children were able to read nearly 20 words per minute (19.58), below international benchmarks but in line with what we found in the June 2008 pilot assessment, although slightly lower. Note that these descriptive statistics show a good balance between children in grade 2 and 3, slightly more boys than girls, and a close balance between control, light, and full intervention schools and children.

Table 3. Descriptive statistics for outcome and predictor variables (grade level, gender and treatment)

Item	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	2970	2.67	.76	0	3
Letter naming fluency	2971	61.16	25.30	0	180
Phonemic awareness	2971	3.49	2.29	0	10
Familiar word fluency	2946	9.26	13.90	0	76.67
Unfamiliar word fluency	2950	2.24	6.02	0	53.6
Connected text fluency	2952	19.58	20.03	0	96.58
Reading comprehension	2971	1.25	1.21	0	4
Listening comprehension	2971	1.68	1.02	0	3
Grade level	2953	2.48	.50	2	3
Gender	2957	1.45	.48	1	2
Treatment (control, light, full)	2970	2.03	.82	1	3

In ***Table 4*** below, we tabulate the descriptive statistics for the EGRA tasks by grade. In the sections that follow we present confidence intervals to determine whether the sample size was sufficient to claim a spread between grade 2 and grade 3 achievement. For now it is sufficient to note that there is a decent enough difference in achievement by grade to suggest that children progress in their reading skills between grades 2 and 3. For now, note that the average grade 2 and 3 student in this sample scored 55.35 and 67.50 letters named correctly respectively, which is less than

the June 2008 pilot where the scores were 64 and 74 letters correctly named. Grade 3 children were able to identify approximately twice as many familiar (12.7 vs. 6.1) and unfamiliar words (3.0 vs. 1.6) than grade 2 children. On the connected text passage, grade 3 children were able to read 25.1 rather than 14.5 words correctly in a minute, and answer 1.5 rather than 1.0 reading comprehension questions. In short, children in grade 3 had much more developed reading skills than grade 2 children, and our assessment was sensitive enough to be able to identify those differences with some precision.

Table 4. Tabulated descriptive statistics by grade

Item	Grade	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	Grade 2	1540	2.621	.821	0	3
	Grade 3	1413	2.721	.690	0	3
Letter naming fluency	Grade 2	1535	55.352	24.863	0	140.25
	Grade 3	1411	67.496	24.314	0	180
Phonemic awareness	Grade 2	1534	3.190	2.193	0	10
	Grade 3	1410	3.830	2.356	0	10
Familiar word fluency	Grade 2	1524	6.060	10.924	0	76.667
	Grade 3	1400	12.732	15.816	0	70.5
Unfamiliar word fluency	Grade 2	1530	1.596	5.196	0	53.6
	Grade 3	1398	2.965	6.736	0	52
Connected text fluency	Grade 2	1523	14.535	16.898	0	96.583
	Grade 3	1404	25.068	21.704	0	96.1
Reading comprehension	Grade 2	1534	.980	1.109	0	4
	Grade 3	1512	1.546	1.251	0	4
Listening comprehension	Grade 2	1534	1.523	1.025	0	3
	Grade 3	1412	1.862	.986	0	3

Table 5 below compares achievement between the genders. In almost all cases, boys scored higher than girls, with the differences in letters named (63.5 vs. 58.1), familiar words named (10.8 vs. 7.4), and oral reading fluency (21.8 and 16.9) most marked. Note that further statistical tests are necessary to determine whether these gender differences are statistically significant or if they are a result of the smaller number of grade 3 girls we identified above. We present these tests in a section below.

Table 5. Tabulated descriptive statistics by gender

Item	Gender	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	Boys	1626	2.664	.766	0	3
	Girls	1331	2.678	.756	0	3
Letter naming fluency	Boys	1622	63.519	25.040	0	100
	Girls	1328	58.114	25.351	0	100
Phonemic awareness	Boys	1620	3.596	2.309	0	10
	Girls	1328	3.368	2.266	0	10
Familiar word fluency	Boys	1609	10.805	14.729	0	74.067
	Girls	1319	7.397	12.619	0	76.667
Unfamiliar word fluency	Boys	1614	2.785	6.666	0	53.6
	Girls	1318	1.596	5.068	0	47.3
Connected text fluency	Boys	1610	21.754	20.285	0	95
	Girls	1321	16.870	19.411	0	96.583
Reading comprehension	Boys	1621	1.345	1.210	0	4
	Girls	1329	1.130	1.204	0	4
Listening comprehension	Boys	1621	1.748	1.008	0	3
	Girls	1329	1.602	1.031	0	3

Table 6 below, investigating differences among control, full, and light intervention schools on the achievement measures, is presented as a check to our random sampling. If the sample size was large enough, we would expect no differences at all among the three different types of schools, given that data were collected pre-intervention. However, as is the case in any random sampling procedure, there are differences, hopefully randomly distributed. In any case, statistical methods can account for any pre-intervention differences between randomly assigned schools.

Note that there were small differences in the letter naming frequency, with control schools (63.2) scoring higher than full (58.5) or light intervention (61.8) schools. Counterintuitively, children in light intervention (10.3) schools named more familiar words than did those in control (8.3) or full intervention (9.2) schools. The pattern continued with light intervention schools scoring higher on unfamiliar words (3.2) and connected text fluency (21.0). The differences were not particularly large, in any category, but we present further analyses below to assess whether the differences are statistically significant.

Table 6. Tabulated descriptive statistics by treatment group

Item	School type	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	Control	951	2.637	.810	0	3
	Full	980	2.653	.789	0	3
	Light	1039	2.715	.685	0	3
Letter naming fluency	Control	948	63.161	25.492	0	180
	Full	979	58.481	24.798	0	126.667
	Light	1036	61.813	25.423	0	140.25
Phonemic awareness	Control	946	3.432	2.339	0	10
	Full	978	3.422	2.162	0	10
	Light	1037	3.617	2.368	0	10
Familiar word fluency	Control	940	8.307	12.595	0	74.067
	Full	971	9.159	13.762	0	70.833
	Light	1030	10.253	15.080	0	76.667
Unfamiliar word fluency	Control	947	1.684	4.724	0	53.6
	Full	971	1.832	5.301	0	49
	Light	1027	3.159	7.455	0	52
Connected text fluency	Control	945	18.179	18.581	0	95
	Full	968	19.440	19.794	0	90.933
	Light	1031	21.041	21.445	0	96.583
Reading comprehension	Control	948	1.271	1.217	0	4
	Full	979	1.192	1.187	0	4
	Light	1036	1.289	1.230	0	4
Listening comprehension	Control	948	1.647	1.025	0	3
	Full	979	1.715	.983	0	3
	Light	1036	1.686	1.048	0	3

In short, analysis of the descriptive statistics shows relatively low achievement, and lower in this sample than that found in Crouch & Korda (2008). There were also some differences between gender on a variety of subtasks and certainly higher achievement

in grade 3 than grade 2. We present detailed descriptive statistic tables disaggregated by treatment type, gender, and grade level in Appendix 1 (1a is for control schools, 1b for full treatment, and 1c for light treatment).

7. Statistical Tests

In order to determine whether the apparent differences presented in the tables above are statistically significant, and not a result of random variation, this section presents the findings from several *t*-tests designed to test the equivalence of means in a variety of outcome variables across gender, treatment, and grade level. For the sake of space, we present only those tests that were statistically significant.

7.1 Statistical Tests for Treatment

The tables below present outcome variables for which there was a statistically significant difference between treatment and control schools. Note that the light and full intervention schools have been grouped together for these analyses since at the baseline, we are interested to know whether there are systematic differences between treatment and control groups that could bias further analyses. For the ease of presentation and analysis, we have combined the full and light intervention groups into one treatment group and compared them against the control schools. Note that this analysis is done in combination with the earlier descriptive statistics that showed few differences by treatment group. ***Table 7*** shows that students in control schools were statistically significantly more capable of identifying letters correctly (63.2 vs. 60.2) than did students in treatment schools. ***Table 8*** shows a statistically significant difference between the correct words read per minute, with students in treatment schools reading about two more than those in control schools (20.3 vs. 18.2). These findings are a bit counterintuitive, because one would expect that letter knowledge is a prerequisite for oral reading fluency, and more knowledge of letters would result in better oral reading fluency skills. Regardless, the differences are still substantively quite small, 2.5 letters and 2 words, respectively.

Table 7. Letters correct, by treatment

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Control	948	63.161	.828	25.492	61.536	64.786
Treatment	2015	60.194	.561	25.170	59.095	61.294
TOTAL	2963	61.143	.465	25.307	60.232	62.055
					T	Pr
					2.981	<.01

Table 8. Oral reading fluency, by treatment

	No. of observa-tions	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Control	945	18.179	.604	18.581	16.993	19.365
Treatment	1999	20.257	.462	20.673	19.359	21.801
TOTAL	2945	2.247	.111	6.021	2.030	2.465
					T	Pr
					-3.503	<.001

7.2 Statistical Tests for Gender

In ***Tables 9 through 15*** that follow, we show that there were several subtasks for which there were statistically significant differences by gender. Boys outperformed girls in letters correct, phonemic awareness, familiar words, unfamiliar words, oral reading fluency, reading comprehension, and listening comprehension. The differences were not always large, but they were always statistically significant. Notably, the difference between boys and girls in oral reading fluency by this assessment was nearly 5 words per minute, a significant difference both statistically and substantively. This suggests that programs will do well to target some of this gender difference in early grade reading achievement, with particular attention to the phonemic awareness skills that lead to oral reading fluency. It should be noted that the consistency of this gender effect is not found in many other countries where EGRA has been implemented (likely because the sample sizes in other countries were much smaller), and further research is necessary to determine what is different in the Liberian context to cause higher achievement for boys. Part of the solution is simply that this current baseline study has a significant sample size and therefore is able to detect statistical significance much more precisely than many of the other EGRAs in other countries in sub-Saharan Africa. On the other hand, given the consistency of the direction of the effect, it might be worth investigating further.

Table 9. Letters correct, by gender

	No. of observa-tions	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	162296	63.519	.622	25.040	62.300	64.739
Girls	1328	58.114	.696	25.351	56.750	59.479
TOTAL	2950	61.086	.466	25.319	60.172	62.000
					T	Pr
					5.800	<.001

Table 10. Phonics correct, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1620	3.596	.057	2.309	3.484	3.709
Girls	1328	3.368	.062	2.266	3.246	3.490
TOTAL	2948	3.494	.042	2.292	3.411	3.576
					T	Pr
					2.691	<.01

Table 11. Familiar words, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1609	10.805	.367	14.729	10.084	11.525
Girls	1319	7.397	.347	12.619	6.715	8.078
TOTAL	2928	9.269	.257	13.920	8.765	9.774
					T	Pr
					6.639	<.001

Table 12. Unfamiliar words, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1614	2.785	.166	6.667	2.459	3.110
Girls	1318	1.597	.140	5.068	1.327	1.870
TOTAL	2957	2.670	.014	.761	2.643	2.700
					T	Pr
					5.334	<.001

Table 13. Correct words per minute, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1610	21.754	.506	20.285	20.762	22.746
Girls	1321	16.870	.534	19.411	15.822	17.917
TOTAL	2931	19.553	.370	20.040	18.827	20.278
					T	Pr
					6.613	<.001

Table 14. Reading comprehension, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1621	1.345	.030	1.210	1.287	1.404
Girls	1329	1.130	.033	1.204	10.065	1.194
TOTAL	2950	1.248	.022	1.212	1.205	1.292
					T	Pr
					4.820	<.001

Table 15. Listening comprehension, by gender

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Boys	1621	1.748	.025	1.008	1.699	1.797
Girls	1329	1.602	.028	1.031	1.546	1.657
TOTAL	2950	1.682	.019	1.021	1.645	1.719
					T	Pr
					3.868	<.001

7.3. Statistical Tests for Grade

It is expected that students in grade 3 will outperform students in grade 2. This is particularly of interest in this research design given the sampling decisions that RTI International made were based on our ability to detect statistically significant differences between grades by achievement, taking into account confidence intervals. Fortunately, ***Tables 16 through 22*** below show not only statistically significant differences between achievement on the various subtasks and grade, but also a

sufficient spread between the confidence intervals bounding our estimates for grade 2 and grade 3 that we are relatively confident in the reliability of the test itself. It is of note that oral reading fluency scores were 14.5 for grade 2 and 25.1 for grade 3. This difference of 10.5 is a bit less than the average of 14 words per minute difference between grades found elsewhere in low-income countries in which EGRA has been implemented. However, this increase of a little more than 10 is almost exactly the same difference noted in the June 2008 assessment in Liberia, although once again the results were slightly higher in the June 2008 assessment than in our nationally representative sample.

Table 16. Letters correct, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1535	55.352	.635	24.864	54.107	56.597
Grade 3	1411	67.496	.647	24.314	66.227	68.766
TOTAL	2946	61.168	.467	25.335	60.253	62.084
					T	Pr
					-13.385	<.001

Table 17. Phonics correct, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1534	3.190	.056	2.193	3.081	3.300
Grade 3	1410	3.830	.063	2.356	3.707	3.954
TOTAL	2944	3.497	.043	2.294	3.414	3.580
					T	Pr
					-7.635	<.001

Table 18. Familiar words per minute, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1524	6.060	.280	10.924	5.511	6.609
Grade 3	1400	12.732	.423	15.816	11.903	13.561
TOTAL	2924	9.255	.257	13.893	8.751	9.758
					T	Pr
					-13.361	<.001

Table 19. Unfamiliar words per minute, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1530	1.596	.133	5.196	1.336	1.857
Grade 3	1398	2.965	.180	6.736	2.612	3.318
TOTAL	2928	2.250	.111	6.019	2.032	2.468
					T	Pr
					-6.186	<.001

Table 20. Correct passage words per minute, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1523	14.535	.433	16.898	13.686	15.384
Grade 3	1404	25.068	.579	21.704	23.932	26.204
TOTAL	2927	19.588	.371	20.052	18.861	20.314
					T	Pr
					-14.711	<.001

Table 21. Reading comprehension, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1534	.980	.028	1.110	.924	1.035
Grade 3	1512	1.546	.033	1.251	1.481	1.611
TOTAL	2946	1.251	.022	1.213	1.207	1.295
					T	Pr
					-13.019	<.001

Table 22. Listening comprehension, by grade

	No. of observations	Mean	Std. error	Std. dev.	95% confidence interval	
					Lower	Upper
Grade 2	1534	1.523	.026	1.025	1.471	1.574
Grade 3	1412	1.862	.026	.986	1.810	1.913
TOTAL	2946	1.685	.019	1.021	1.648	1.722
					T	Pr
					-9.134	<.001

8. Tool Reliability

In order to determine whether and how the various subtasks in the Early Grade Reading Assessment as implemented in Liberia are reliable, and whether they are testing an underlying skill, presumably early grade reading skills, the reliability tests presented below were performed. First, simple bivariate correlations between the various subtasks were performed and are presented in ***Table 23.*** Note that the correlations were lowest in the phonemic awareness and listening comprehension rows and columns, suggesting that they fit less well together with the rest of the exam on the one hand, and on the other, that they are assessing slightly different skills than the rest of the construct.

Table 23. Correlation matrix for EGRA Plus tasks

	Letter naming fluency	Phonemic awareness	Familiar word fluency	Unfamiliar word fluency	Connected text fluency	Reading comprehension	Listening comprehension
Letter naming fluency	1.00						
Phonemic awareness	0.32	1.00					
Familiar word fluency	0.53	0.33	1.00				
Unfamiliar word fluency	0.37	0.33	0.62	1.00			
Connected text fluency	0.57	0.32	0.84	0.52	1.00		
Reading comprehension	0.53	0.34	0.63	0.41	0.75	1.00	
Listening comprehension	0.31	0.26	0.28	0.20	0.34	0.39	1.00

Following the correlation matrix, we performed a Cronbach's alpha reliability test, the results of which are presented in ***Table 24*** below. Note that the lowest item-test correlations were found for the listening comprehension and phonemic awareness subtasks. Even with those slight issues, the entire test's Cronbach's alpha is 0.85,

which is more than acceptable for low-stakes assessments like the EGRA, and is likely sufficient for higher-stakes examinations as well.

Table 24. Reliability analysis of EGRA tool in Liberia

Item	Item-test correlation	Item-rest correlation	Average inter-item correlation	Alpha
Letter naming fluency	0.72	0.60	0.44	0.82
Phonemic awareness	0.58	0.42	0.49	0.85
Familiar word fluency	0.84	0.76	0.40	0.80
Unfamiliar word fluency	0.69	0.56	0.45	0.83
Connected text fluency	0.86	0.79	0.39	0.79
Reading comprehension	0.80	0.71	0.41	0.81
Listening comprehension	0.55	0.39	0.49	0.85
Overall test			0.44	0.85

Next, we performed a principal components analysis[7] to investigate whether there was an underlying construct that all of the EGRA tool subtasks were evaluating. The resultant principal component 1 loaded quite highly on all of the subtasks, and a little bit less so on phonemic awareness and listening comprehension, as the first column indicates. The second column shows that both of those subtasks (phonemic awareness and listening comprehension) added unique information to the total assessment, supporting their inclusion into the entire assessment.

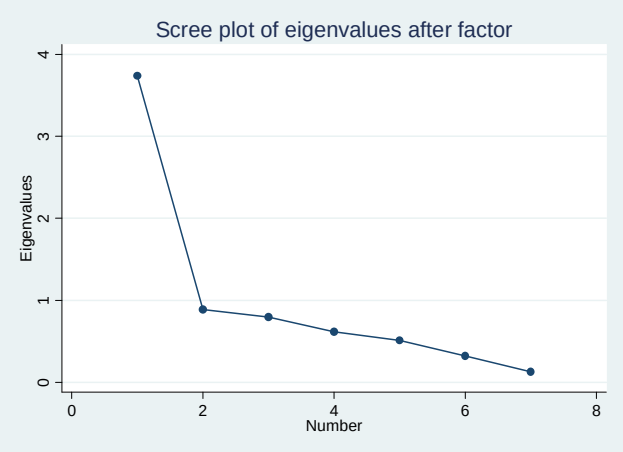
[7] Principal components analysis is another means of determining whether the components of this particular assessment measure an underlying skill, and how much of the variation in the subtask achievement this particular underlying component measures.

Table 25. Principal component of the Liberia EGRA

Principal component 1 loading		Uniqueness of each component	
Letter naming fluency	0.72	Letter naming fluency	0.48
Phonemic awareness	0.52	Phonemic awareness	0.73
Familiar word fluency	0.87	Familiar word fluency	0.24
Unfamiliar word fluency	0.68	Unfamiliar word fluency	0.53
Connected text fluency	0.89	Connected text fluency	0.20
Reading comprehension	0.82	Reading comprehension	0.33
Listening comprehension	0.50	Listening comprehension	0.75

We created a screeplot ***(Figure 1)*** to identify how much of the variation in the total EGRA was explained by the new principal component that we created. Figure 1 shows that the first component explains about 3.74 eigenvalues of variation. In other words, nearly half of the entire variation of all the subtasks is found within this one component, which we would argue is an early grade reading skill. This is good news for our ability to argue that this set of subtasks is a good estimate of what we are interested in understanding, the quality of students' early reading skills. The fact that the second principal component in Figure 1 below is so low, with less than 1.0 eigenvalue, suggests that the variation of this underlying construct in the first principal component is doing a reasonably good job of identifying the true underlying skill of early grade reading.

Figure 1. Screeplot of eigenvalues for principal component 1, "early grade reading skill," and other potential principal components, all of which have eigenvalues of less than 1



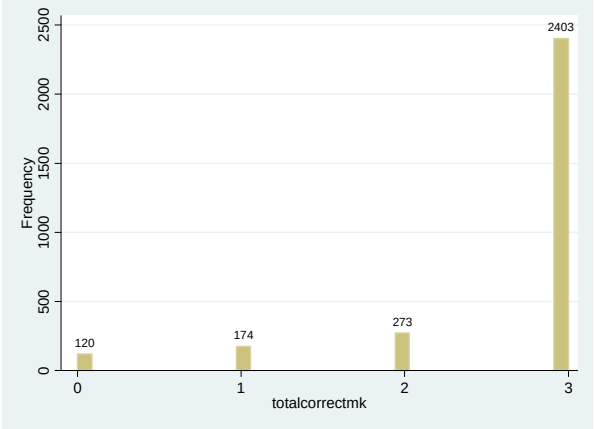
9. Assessment Results

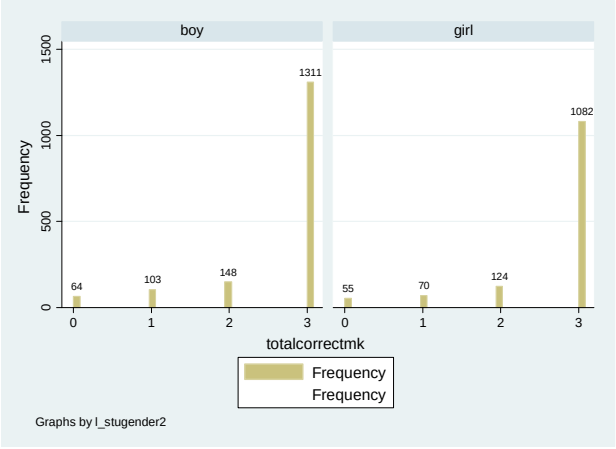
In the section that follows, figures created to illustrate the impact on achievement of the various subcategories are shown. This section is organized by subtask and looks at how particular variables are predictive of student achievement in particular subtasks.

9.1 Print Orientation

Figure 2 presents histograms of achievement on the initial print orientation subtask, which asked children to identify the direction in which text should be read and to place their finger on the correct place for the subtask. Note that the maximum score on this subtask was 3. Most children scored 3 correct, in fact, and the histogram on the right shows very little, if any, systemic differences by gender. It is important to note, however, that a significant and substantive number of children in grade 2 and 3 were unable to accurately show in which direction text should be read and where the text reading should start.

Figure 2. Histograms for print orientation and print orientation by gender

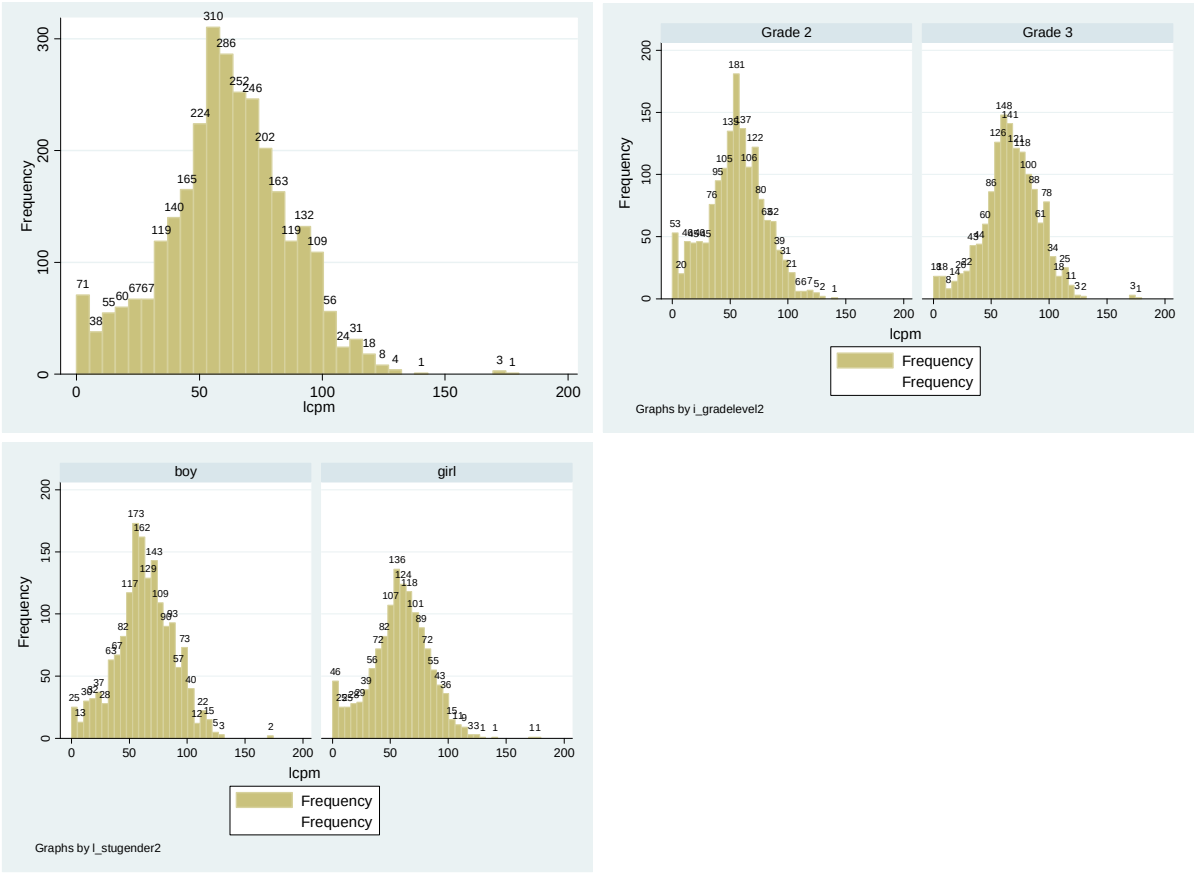




9.2 Letter Naming Fluency

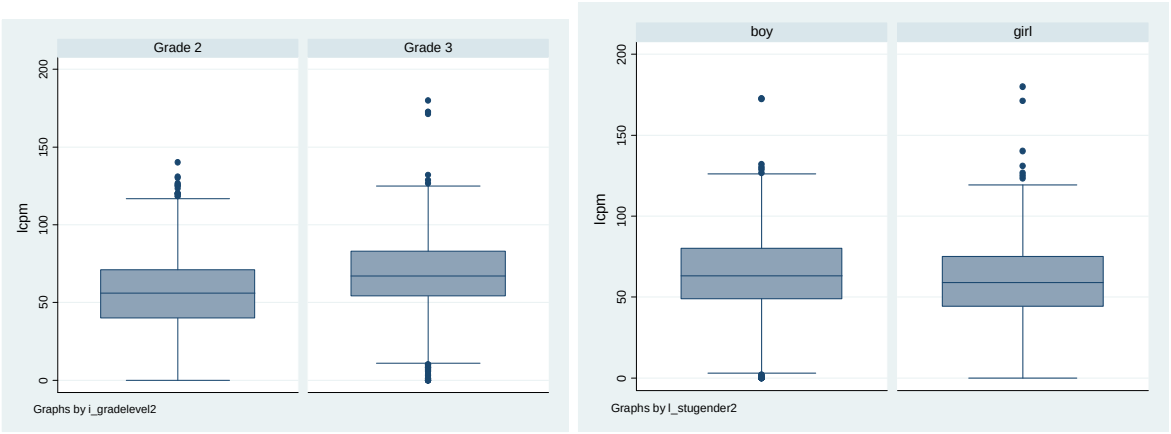
The three histograms in ***Figure 3*** below show achievement on the letter fluency subtask. The first histogram, with the full assessment totals, shows why letter fluency must be done using a per minute manipulation: 192 children scored a perfect 100 on this subtask, and the assessment, without a time correction, would not adequately estimate their skill in this subtask. Also note that the grade 3 chart to the right was pushed much more to the right than was the grade 2 chart, which looked almost like a normal distribution save the extra observations at 0. Finally, the third histogram provides evidence of the fact that boys performed better on this subtask than girls, on average.

Figure 3. Histograms for letter naming fluency and letter naming fluency by gender and grade



These findings are mirrored in ***Figure 4*** below. The box plot on the left compares achievement in letter naming fluency by grade; the one on the right is by gender. Note that the grade 2 box plot bottoms as well as tops out, while the grade 3 box plot has some outliers at the bottom of the distribution. There seems to be a real progression in this subtask by grade, with the 25th percentile score for grade 3 nearly being the same as the mean score for grade 2. Similarly, the box plot to the right shows that the means, 25th, and 75th percentile scores are much higher for boys than girls. Note that while the international assessment Dynamic Indicators of Basic Early Literacy Skills (DIBELS) does not have a benchmark for grades 2 and 3 on this subtask, the at-risk benchmark for kindergarten in the United States is 40. In this case, 529 children, or 17.7% of Liberian children sampled, would have been considered at risk for reading difficulties since they scored below 40 letters read on this subtask.

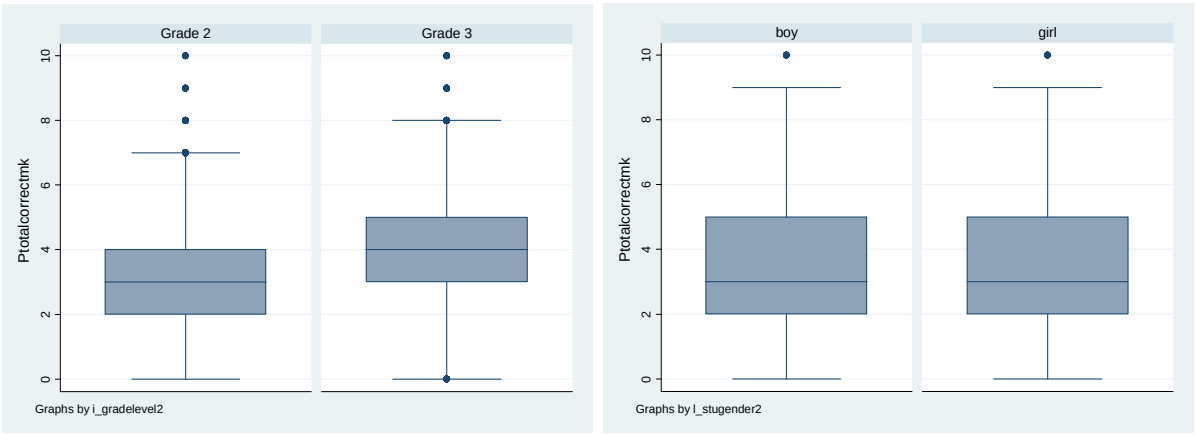
Figure 4. Box plots for letter naming fluency and letter naming fluency by gender



9.3 Phonemic awareness

The phonemic awareness subtask has 10 items, so the scores are spread across 10 markers in the ***Figure 5*** box plots below. Note the differentiation by grade; again, the 25th percentile for grade 3 is very close to the mean for grade 2. In the box plot to the right, note that there appear to be few differences in achievement in this subtask, and it is the large sample size that allows us to identify the statistically significant differences. It is worth noting that some children really struggled with basic phonemic awareness skills. A full 16.9% of children (504) were unable to correctly answer a single phonemic awareness question.

Figure 5. Box plots for phonemic awareness, by grade and by gender

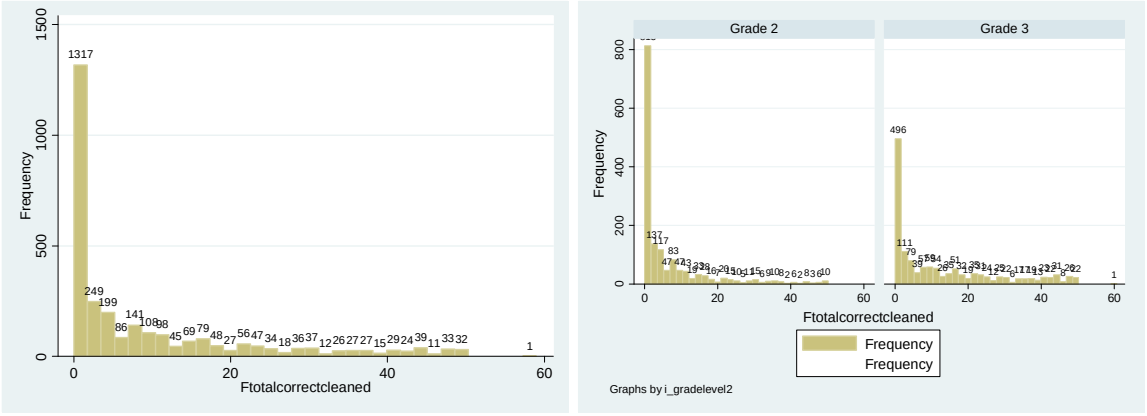


9.4 Familiar words

The histograms in ***Figure 6*** below show the quality of student responses to familiar word identification tasks. Note that while the average score on this subtask was 9.3, this masks significant variation in the quality of student achievement in this subtask, with 44.6% of children (1317) unable to correctly identify a single “familiar” word. The histogram to the right shows that this changes over time, with many more

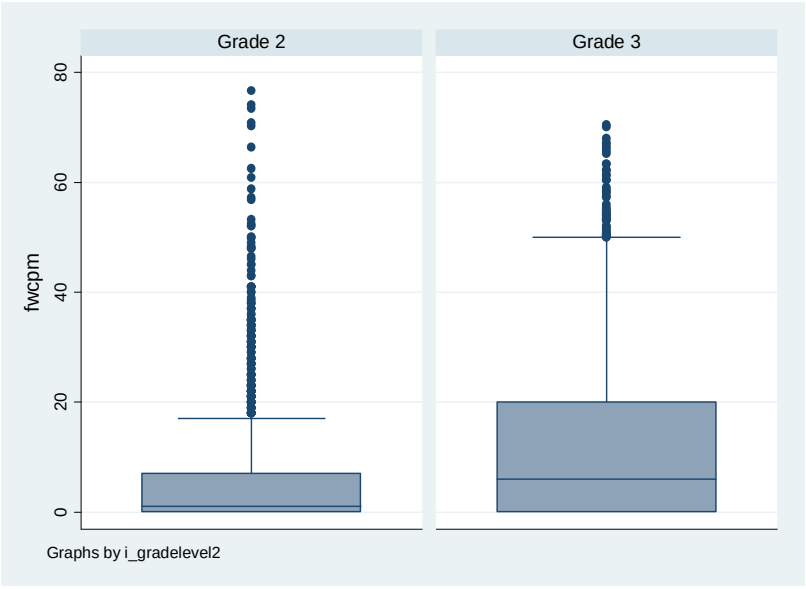
children able to answer many more questions, but that there remains a very large percentage of children who do not have the ability to accurately identify a single word.

Figure 6. Histograms for familiar word fluency and familiar word fluency by grade



The low performance on this subtask, and—potentially, although less likely—the inappropriateness of this subtask to adequately measure variation in Liberian student performance is evident in the ***Figure 7*** box plot below, which shows that students in grade 2 did very poorly on this subtask and grade 3 students struggled quite a bit. Note that the average achievement for grade 2 students was just above 0, and the average for grade 3 was not much higher. There were some outliers, and in grade 2 outliers that scored nearly 80 familiar words read per minute, which is an impressive score in any context. That said, the 90th percentile for grade 2 students was less than 20 familiar words read per minute.

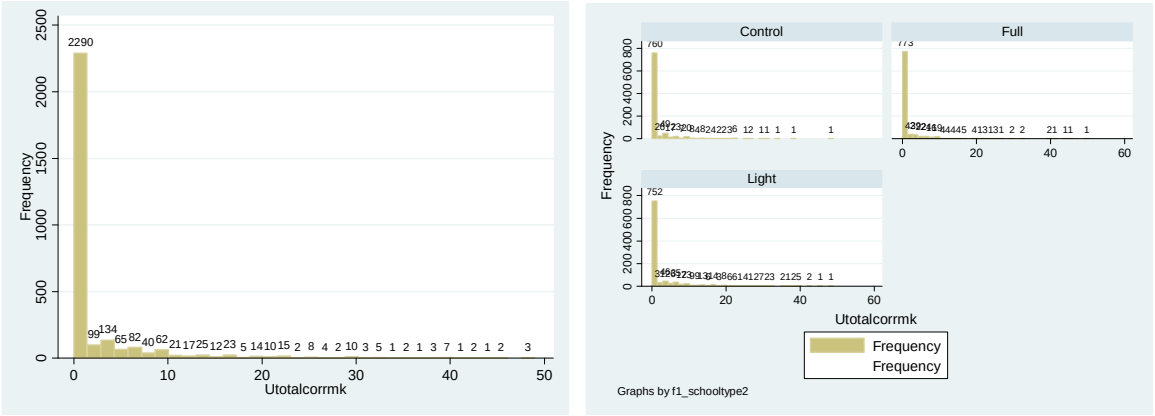
Figure 7. Box plot for familiar word fluency, by grade



9.5 Unfamiliar words

This subtask required that children use their decoding skills to identify nonsense words using the rules of decoding. The low achievement of children on this subtask is stark, with a full 59.7% of children scoring 0 on the subtask (1783 children). The average score, including those zeros, was only 2.24 words correctly identified. There seems to be little differentiation by the type of school, as we would expect, since most children struggled on this subtask. It should be asked, at this point, whether the skills taught in Liberian schools match well with the skills necessary to perform well on this subtask, particularly decoding and phonemic awareness. Given that Liberia uses English as a medium of instruction in these grades, however, it will be interesting to note whether children perform better on this subtask after teachers are provided significant professional development on these topics. If so, then the low achievement on this subtask at this point is an artifact of a lack of teaching by teachers of an important reading skill. If not, then more careful attention must be paid to the local Liberian context in determining which types of instruction are most appropriate. Another logical implication of this is that while students knew letter names, they were not being taught letter sounds sufficiently.

Figure 8. Histograms for unfamiliar word fluency and unfamiliar word fluency by type of school



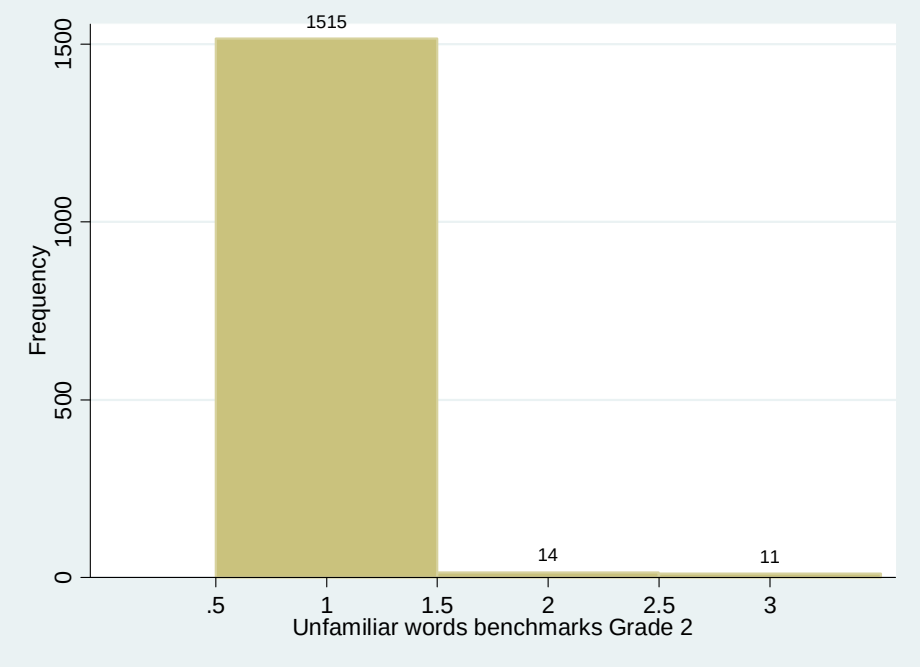
Given that DIBELS uses the unfamiliar word fluency subtask as a benchmark for assessing "at-risk"ness in U.S. children, it is useful for illustration purposes to identify where Liberian children would perform based on the U.S. benchmarks. It is worth noting at this point that Liberian reading instruction is different, the context is different, the resources available in schools are different, and the training of teachers is different. That said, given the lack of appropriate benchmarks for these types of skills in sub-Saharan Africa, we use the U.S. benchmarks illustratively, until a research base can be created for local and indigenous benchmarks. ***Figure 9*** below depicts the benchmarks for words read at the beginning of grade 2, for deficit, emerging, and established readers. Using those same benchmark levels against the Liberian EGRA unfamiliar word subtask, ***Figure 10*** is a histogram noting how many children would fall into the deficit, emerging, and established subcategories.

Figure 9. U.S. DIBELS benchmarks for the beginning of second grade, using unfamiliar words

SECOND GRADE	Beginning of Year Month 1 - 3	
DIBELS Measure	**Scores**	**Status**
Nonsense Word Fluency (NWF-CLS)	0 - 29	Deficit
	30 - 49	Emerging
	50 and above	Established

Figure 10 shows that for the grade 2 students assessed, 1515 (98.3%) would fall into the deficit category, 14 (1.0%) into the emerging category, and 11 (0.7%) into the established category. If ability to decode nonsense words is a critical part of the decoding repertoire of established readers, who are able to move from learning to read to reading to learn, then Liberian children's achievement is concerning. If this finding is related to discomfort with reading words with no meaning or if the decoding skills taught in Liberia are significantly different from those appropriate in the other countries were EGRA has been implemented, then there are plausible explanations with solutions related to EGRA assessment improvement. If not, and if this is representative of a lack of decoding skills among Liberian children, then intensive interventions in helping teachers better teach these skills would be very likely to have a significant impact, considering from where these children's achievement will be coming.

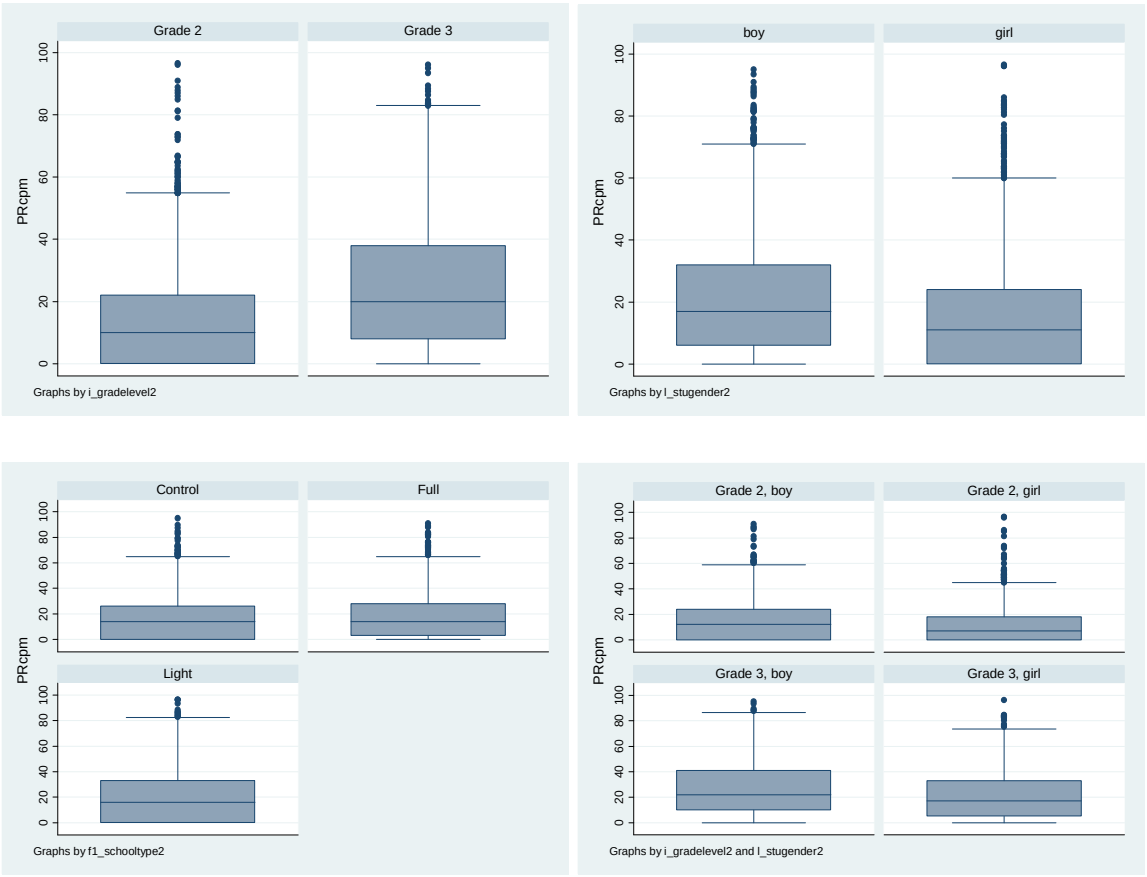
Figure 10. U.S. DIBELS benchmarks for the beginning of second grade using unfamiliar words, with Liberian achievement on the unfamiliar words subtask



9.6 Oral Reading Fluency

In this subtask, children were asked to read aloud a passage of local relevance within one minute. The variable we used was the correct number of words read within that one minute, adjusted for time for the fast readers. The four box plots in ***Figure 11*** below show the achievement of Liberian children in this sample on this subtask. Note that grade 3 students performed much better, with the 75th percentile again quite close to the mean score at grade 2. Similarly, boys scored much higher on this subtask, which throughout the literature is suggested to be most closely related to the types of skills necessary for long-term reading (and with it, schooling) success. Note that the bottom left box plot shows that achievement was slightly higher on this subtask in light intervention schools. Finally, the bottom right box plot shows both a grade and a gender effect for this subtask. Boys did better than girls within each grade and grade 3 children outperformed grade 2.

Figure 11. Box plots for oral reading fluency of connected text, by grade, gender, type of school, and grade level and gender

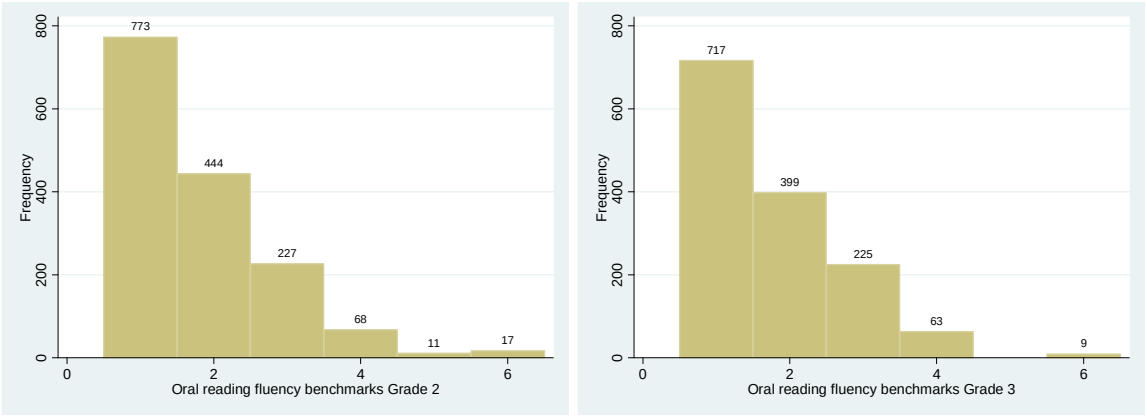


Again, for this oral reading fluency subtask, there are benchmarks used internationally that were published in 2006.[8] These benchmarks are presented in Appendix 2 of this report. We plotted the achievement of grade 2 and grade 3 Liberian students against

[8] Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher 59*(7): 636–644. See also Appendix 2.

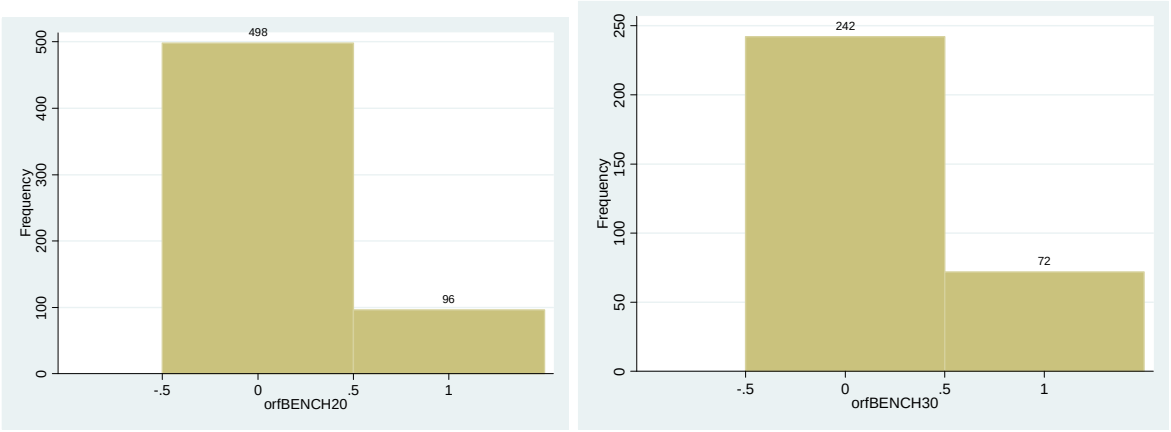
these benchmarks, for 10th, 25th, 50th, 75th and 90th percentiles, as shown in ***Figure 12.***

Figure 12. Histogram for oral reading fluency of connected text, using Hasbrouck & Tindal's benchmarks at 10th, 25th, 50th, 75th and 90th percentiles



If achievement was normally distributed across the international benchmarks, then we would see most Liberian children scoring between the 25th and 75th percentiles. On the contrary, for both grade 2 and 3 (and consistent for grade 2 and grade 3) about half of Liberian children scored less than 10th percentile, and almost all scored below the 50th percentile of the international benchmarks. ***Figure 13*** shows these same data another way. For both grade 2 and grade 3, it simply shows the number of children who scored an absolute 0 on the oral reading fluency subtask against the number of children who scored over 50th percentile. For both grade 2 and grade 3, approximately 5 times more children scored 0 than scored over 50th percentile on the oral reading fluency portion of this assessment.

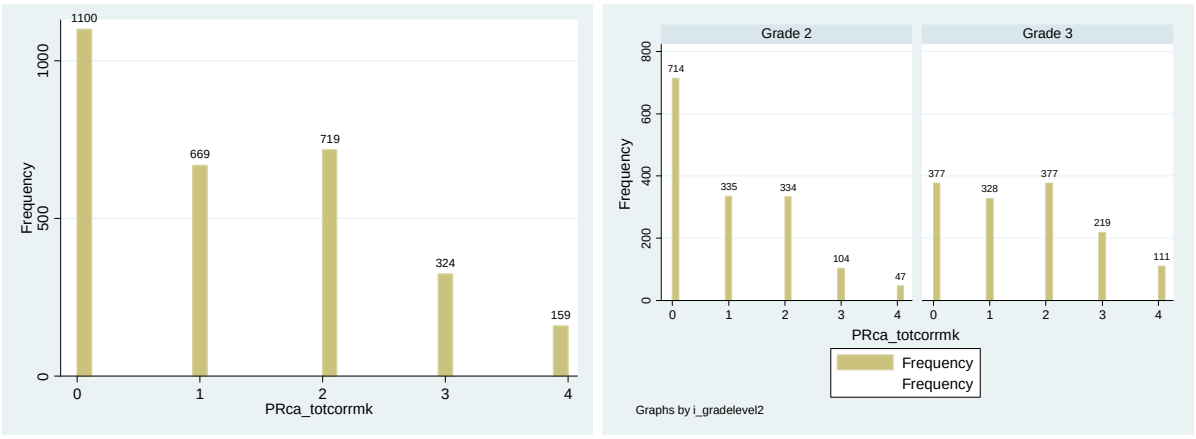
Figure 13. Histogram for oral reading fluency of connected text, using Hasbrouck & Tindal's benchmarks at 50th percentile and above, as compared to those who scored zero



9.7 Reading Comprehension

On this subtask, children were simply asked to answer five basic comprehension questions based on the connected text passage they just read. ***Figure 14*** provides histogram evidence of student achievement on this subtask. Note that 37.3% of children were unable to answer any questions, likely because few children were able to read the entire passage. Note also that achievement was higher for children in grade 3 than in grade 2.

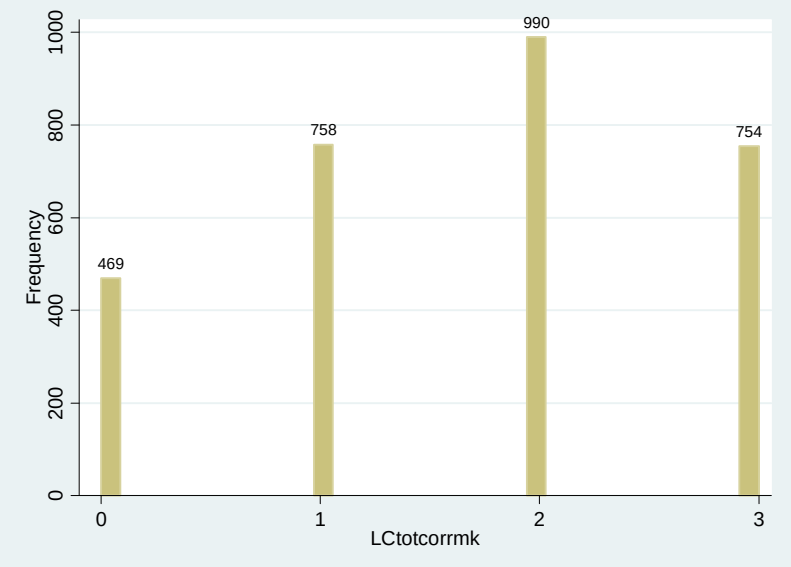
Figure 14. Reading comprehension achievement for full sample and by grade



9.8 Listening Comprehension

On this subtask, children listened to a short story about which they were asked some questions and then required to respond, testing the listening portion of reading skills and phonemic awareness. ***Figure 15*** below is a simple histogram showing achievement on this subtask. Note that the modal score was two answers correct out of three.

Figure 15. Histogram for listening comprehension achievement, for full sample



10. Multiple Regression Analyses

Both researchers and policy makers are interested in predicting achievement outcomes. Researchers are interested because such models are useful for contributing to the literature about a certain topic, and policy makers because they can use such information to create appropriate interventions that will have an impact on particular outcomes. This particular EGRA administration provides a rich data set whereby predictions can be made about family and home practices that have an impact on student achievement. In fact, the data set is so rich that for the purposes of this particular analytic report, we focus only on the plausible relationships among student, family, and school-level predictors and save some of the more complex analyses using teacher and principal predictors for the full project analyses to come. These analyses will use multiple regression techniques to estimate whether there are predictive relationships between teacher-level and principal-level predictors on student outcomes.

That said, the student questionnaire attached to the EGRA assessment itself provides a rich set of data from which some claims can be made about the impact of a variety of predictor variables. A research decision was made to present the findings with oral reading fluency—rather than other alternative subtasks— as the outcome variable, since this outcome variable is most predictive of future success and seems most closely tied to the underlying construct early reading skill, based on the principal components and Cronbach's alpha analysis above.

The process by which the analyses presented here were determined is simple. First, pairwise correlations between predictor variables from the student questionnaire and the outcome variable, oral reading fluency, were performed. For space reasons, that full correlational analysis is not presented here; however, this analysis identified several promising predictive variables for further inspection. It might be interesting, however, to note some of the predictor variables that did not have a statistically significant correlation with student achievement on oral reading fluency. For example, home language of both father and mother had no relationship with oral reading fluency, possibly due to a lack of local differentiation on that variable.

Similarly, having books at home and having someone read those books to a child was not correlated with that child's achievement on this measure. Neither was a teacher practicing sounds with the child in the classroom, interestingly enough. None of the typical proxies for household wealth had an impact on achievement, including having a television, refrigerator, motorbike, car, or radio. Neither did teacher practices, such as sounding out words, silent reading, reading aloud, telling the meaning of new words, retelling stories, or giving reading assignments. This is, of course, not to argue that these practices or wealth proxies are not important, but simply that other variables were associated with variation in student achievement on this oral fluency assessment. A handful of variables were statistically significantly related to oral reading fluency of connected text, which we present in ***Table 26*** below.

Note that the first set of regression findings below comes from single regression analyses, using one predictor in the model to predict oral reading fluency scores.

Similar to the June 2008 analysis, the difference between oral reading fluency scores in grade 2 and 3 was 10.5 words per minute. The difference between male and female oral reading fluency scores was similarly large, at 4.88 words per minute, favoring boys. Children who attended nursery school scored 1.15 words less than those who attended the first year of kindergarten (K1), and 1.35 words less than those who attended the second year of public kindergarten (K2). It appears that the quality of early preparation matters quite a bit; however, this variable might simply be a proxy for family wealth. As expected, children who repeated a grade did not do as well as those who did not, by 6.54 words on average. For every day that children reported studying after school, they scored 1.78 words higher on the oral reading fluency subtask. Likewise, for every day that a child missed school in a week, the child scored 1.12 words lower, although this might be a proxy for either distance from school or child health, or both. Finally, at the parental level, students who reported that their parents did nothing when they learned of their failure on a test scored 4.35 points lower. This lack of involvement showed a significantly large relationship with student achievement. Encouraging success (1.68 words) and helping children who struggle (2.51 words) are parental practices that are associated with higher achievement. Another potential proxy for wealth and/or health, eating before school, was associated with 1.58 words higher on the oral reading fluency subtask.

The next set of regression results tested the impact of predictors separately for grade 2 and grade 3. We found that attending preschool (not kindergarten) was related to oral reading fluency for grade 2 students, but not for grade 3. One interpretation of this is that any impact of preschool quality was already felt by grade 3, or that those who suffered from low-quality preschool were more likely to have dropped out of school. On the other hand, studying after school only had a relationship with grade 3, rather than grade 2, achievement. Likewise, children whose parents encouraged them after learning of a good test result had 2.40 points words more on the oral reading fluency measure in grade 2, but no difference in grade 3. This is potentially indicative of the notion that the selective processes in Liberian schools in the first two grades mean that many of the variables that serve as a proxy for parental wealth and involvement have already had an impact on schooling and persistence decisions by grade 3. Note that for the rest of the regression models that we fit, there was no difference in the results by grade 2 and grade 3.

The final regression results showed that a model with both gender and grade had parameter estimates that were significant for both grade and gender at a magnitude similar to the models above. The final two models, with gender and age, fit separately for grade 2 and grade 3 children, show that age had little relationship with student achievement in grade 2, but by grade 3, overage children scored a bit lower on oral reading fluency measures. This might be a proxy for repetition, since those children who have repeated at least one time by grade 3 will be overage and it is logical that their achievement would be lower. In short, these regression findings show that student level variables, family wealth proxies, parental behaviors, and student activities have a predictive relationship with oral reading fluency, while several other teacher practice variables do not. This will be important to revisit after the implementation of the

program, particularly with respect to whether teacher level characteristics have a statistically significant relationship with student-level achievement.

Finally, the three columns on the right of Table 26 show the overall fit of the regression models. Note that while many of the models that we present here are statistically significant, the R-squared scores are quite low, with most models explaining less than 5% of the variation in oral reading fluency. More sophisticated models might do better. However, this suggests that while we do know some of the barriers to oral reading fluency, and by proxy, early reading skills in Liberia, student achievement is generally low. For policy, it also suggests that targeting the general quality of early reading instruction is likely to be a more appropriate strategy than only targeting particular groups of children, teachers, or schools.

Table 26. Results from regression analyses

Outcome	Predictor	Coeff.	Std. Error	*T*	Sig.	Confidence interval		*F*	Sig.	R^2
						Lower	Upper			
Oral reading fluency (ORF)	Grade	10.53	.71	14.71	<.001	9.13	11.94	216.41	<.001	.069
ORF	Gender (girl)	-4.88	.74	-6.61	<.001	-6.33	-3.44	43.73	<.001	.015
ORF	Attended nursery (not kindergarten)	-2.70	.74	-3.66	<.001	-4.14	-1.25	13.38	<.001	.005
ORF	Repeated a grade	-6.54	.83	-7.89	<.001	-8.16	-4.91	62.19	<.001	.021
ORF	Reading aloud	.117	.05	2.43	.01	.02	.21	5.92	.02	.002
ORF	Days study after school	1.783	.17	10.73	<.001	1.46	2.11	115.20	<.001	.038
ORF	Missed days	-0.222	.070	-3.17	<.00	-.35	-.08	10.03	<.01	.004
ORF	How many days missed	-1.12	.38	-2.95	<.01	-1.86	-.37	8.71	<.01	.003
ORF	Parents learned of failure, did nothing	-4.352	.95	-4.56	<.001	-6.22	-2.48	20.77	<.001	.007
ORF	Parents learned of failure, helped child	2.51	.91	2.75	<.01	.72	4.30	7.59	<.01	.003
ORF	Ate before school today	1.58	.77	2.06	.04	.08	3.09	4.25	.04	.002
ORF (Grade 2)	Attended some preschool	-.09	.04	-2.24	.03	-.17	-.01	5.01	.03	.003
ORF (Grade 3)	Attended some preschool	-.02	.08	-0.32	.75	-.18	.13	.10	.75	.000
ORF (Grade 2)	Study after school	-.03	.05	-0.51	.61	-.12	.07	.26	.61	.000
ORF (Grade 3)	Study after school	-.20	.09	-2.10	.04	-.38	-.01	4.43	.04	.003
ORF (Grade 2)	Parents learned of success, encouraged	2.40	.82	2.91	<.01	.78	4.01	8.49	<.01	.006
ORF (Grade 3)	Parents learned of success, encouraged	.17	1.08	0.16	.87	-1.95	2.29	.03	.87	.000
ORF	Gender (girl)	-4.46	.72	-6.23	<.001	-5.86	-3.06			
	Grade	10.39	.71	14.56	<.001	8.99	11.79	129.07	<.001	.082
ORF (Grade 2)	Gender (girl)	-4.38	.87	-5.03	<.001	-6.08	-2.67			
	Age	-.09	.09	-1.04	.30	-.26	.08	13.09	<.001	.017
ORF (Grade 3)	Gender (girl)	-5.04	1.19	-4.24	<.001	-7.37	-2.71			
	Age	-.90	.24	-3.76	<.001	-1.37	-.43	14.27	<.001	.021

11. Conclusion

Access to this rich baseline data set provided a great deal of knowledge of the current status of Liberian early grade reading skills. The findings sections above provide quite a few suggestions for how a successful early grade reading intervention might well be designed to improve the quality of student learning in reading. For example, here are some suggestions for how the findings from this baseline survey could be integrated with the current EGRA Plus reading intervention:

- Both light and full intervention schools and communities would further benefit from knowledge as to the current status of reading achievement in Liberian schools. Particularly, attention should be paid to the relatively low levels of student achievement at grades 2 and 3, as well as to the comparisons between nonreaders and readers, with respect to percentage scores as well as knowledge of phonemic awareness.
- It appears that many children have knowledge of letters, but they struggle much more when it comes to phonemic awareness and the ability to decode familiar and unfamiliar words. Therefore, teacher strategies for developing phonemic awareness and decoding skills should be designed.
- The gender gap is quite pronounced in this data set, such that girls are disadvantaged. The achievement of girls should be a particular target of this program, although it is likely that providing strong teacher professional development for teachers will result in increased achievement of children across the spectrum, girls included.
- During the second assessment, care should be taken that the familiar word and unfamiliar word fluency tasks are locally appropriate. It appears that, given the pilot and baseline experience, the problem comes from low achievement, but effort should be made to ensure that this is the case in the next EGRA administration.

The design of this intervention, surrounded by three phases of data collection, lends itself to detailed analysis of program impacts. To ensure that causal claims are able to be made, the following design issues should be kept under consideration:

- The schools used in the baseline survey should be retained throughout the intervention.
- Given the finding in a Kenya EGRA that program leakage influenced the results of a similar experiment, care should be made to include only treatment schools in intervention programs.
- The three phases of data collection will allow for a rich longitudinal design. In order to support this, care must be taken to ensure that teacher, student, principal, and school codes are matched carefully.

Appendix 1a. Achievement by gender and grade for control schools on reading subtasks

Item	Gender	Grade	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	Boy	2	287	2.620	.814	0	3
		3	256	2.652	.797	0	3
	Girl	2	212	2.585	.880	0	3
		3	193	2.710	.721	0	3
	All	2	499	2.605	.842	0	3
		3	451	2.678	.764	0	3
Letter naming fluency	Boy	2	285	59.818	23.883	0	130.35
		3	256	70.829	25.546	0	172.55
	Girl	2	211	55.318	23.020	0	110
		3	193	66.521	27.120	0	180
	All	2	496	57.903	23.602	0	130.35
		3	451	68.993	26.251	0	180
Phonemic awareness	Boy	2	285	3.428	2.330	0	10
		3	255	3.761	2.256	0	10
	Girl	2	210	3.086	2.305	0	9
		3	193	3.378	2.434	0	9
	All	2	495	3.283	2.323	0	10
		3	450	3.598	2.349	0	10
Familiar word fluency	Boy	2	285	7.728	11.495	0	74.067
		3	253	12.604	15.570	0	67.2
	Girl	2	210	4.275	8.217	0	58.8
		3	189	8.017	12.178	0	58.4
	All	2	495	6.264	10.365	0	74.067
		3	444	10.64	14.364	0	67.2
Unfamiliar word fluency	Boy	2	286	1.733	5.203	0	53.6
		3	254	2.685	5.809	0	38
	Girl	2	211	.834	2.816	0	25
		3	193	1.249	3.776	0	24.160
	All	2	497	1.351	4.372	0	53.6
		3	449	2.056	5.069	0	38
Connected text fluency	Boy	2	285	16.407	16.226	0	79
		3	255	24.811	20.763	0	95
	Girl	2	210	11.785	14.388	0	72
		3	192	19.165	20.057	0	83.733
	All	2	495	14.446	15.626	0	79
		3	449	22.335	20.611	0	95
Reading comprehension	Boy	2	285	1.151	1.176	0	4
		3	256	1.645	1.272	0	4
	Girl	2	211	.962	1.108	0	4
		3	193	1.301	1.200	0	4
	All	2	496	1.071	1.150	0	4
		3	451	1.494	1.251	0	4
Listening comprehension	Boy	2	285	1.551	1.004	0	3
		3	256	1.875	.986	0	3
	Girl	2	211	1.521	1.034	0	3
		3	193	1.622	1.049	0	3
	All	2	496	1.538	1.016	0	3
		3	451	1.769	1.020	0	3

Appendix 1b. Achievement by gender and grade for full intervention schools on early grade reading subtasks

Item	Gender	Grade	*N*	Mean	Standard deviation	Minimum	Maximum
Print orientation	Boy	2	236	2.568	.894	0	3
		3	272	2.676	.748	0	3
	Girl	2	256	2.621	.27	0	3
		3	203	2.768	.630	0	3
	All	2	494	2.589	.866	0	3
		3	479	2.714	.702	0	3
Letter naming fluency	Boy	2	236	56.162	24.655	0	126.667
		3	271	65.449	22.835	0	117.15
	Girl	2	256	48.482	25.380	0	113.167
		3	203	63.805	22.741	0	126.667
	All	2	494	52.271	25.305	0	126.667
		3	478	64.831	22.733	0	126.667
Phonemic awareness	Boy	2	236	2.987	1.923	0	9
		3	271	3.841	2.300	0	9
	Girl	2	255	3.122	2.036	0	10
		3	203	3.754	2.286	0	10
	All	2	493	3.059	1.978	0	10
		3	478	3.793	2.290	0	10
Familiar word fluency	Boy	2	233	5.789	10.737	0	70.833
		3	268	14.969	16.007	0	70
	Girl	2	254	4.170	8.139	0	57.2
		3	203	11.350	15.543	0	70.5
	All	2	489	4.949	9.485	0	70.833
		3	475	13.336	15.858	0	70.5
Unfamiliar word fluency	Boy	2	236	1.648	5.514	0	49
		3	268	2.776	6.014	0	44
	Girl	2	254	.846	2.933	0	32
		3	200	2.005	5.976	0	45
	All	2	492	1.228	4.377	0	49
		3	472	2.430	5.981	0	45
Connected text fluency	Boy	2	234	15.326	16.980	0	90.933
		3	267	27.659	21.283	0	83.583
	Girl	2	252	10.688	14.289	0	81.2
		3	202	23.764	21.160	0	83.583
	All	2	488	12.966	15.814	0	90.933
		3	473	26.001	21.270	0	83.583
Reading comprehension	Boy	2	236	.915	1.015	0	4
		3	271	1.638	1.215	0	4
	Girl	2	256	.766	1.036	0	4
		3	203	1.419	1.234	0	4
	All	2	494	.840	1.027	0	4
		3	478	1.548	1.229	0	4
Listening comprehension	Boy	2	236	1.606	1.011	0	3
		3	271	1.956	.934	0	3
	Girl	2	256	1.445	.977	0	3
		3	203	1.882	.931	0	3
	All	2	494	1.522	.994	0	3
		3	478	1.918	.931	0	3

Appendix 1c. Achievement by gender and grade for light schools on early grade reading subtasks

Item	Gender	Grade	*N*	Mean	Standard deviation	Minimum	Maximum
Page orientation	Boy	2	293	2.686	.729	0	3
		3	273	2.762	.605	0	3
	Girl	2	252	2.639	.789	0	3
		3	207	2.773	.593	0	3
	All	2	547	2.665	.756	0	3
		3	483	2.768	.597	0	3
Letter naming fluency	Boy	2	292	58.023	25.460	0	120
		3	273	70.795	24.200	0	119.633
	Girl	2	251	53.093	24.827	0	140.25
		3	206	66.001	23.009	0	125
	All	2	545	55.822	25.324	0	140.25
		3	482	68.739	23.770	0	125
Phonemic awareness	Boy	2	292	3.483	2.307	0	10
		3	272	4.059	2.527	0	10
	Girl	2	252	2.933	2.151	0	10
		3	207	4.101	2.248	0	10
	All	2	546	3.225	2.253	0	10
		3	482	4.085	2.407	0	10
Familiar word fluency	Boy	2	288	7.612	12.950	0	73.5
		3	273	15.728	17.148	0	68
	Girl	2	250	5.978	11.823	0	76.667
		3	205	11.851	16.310	0	58.75
	All	2	540	6.879	12.468	0	76.667
		3	481	14.101	16.849	0	6
Unfamiliar word fluency	Boy	2	290	2.532	6.463	0	46
		3	271	5.244	9.355	0	52
	Girl	2	249	1.716	6.336	0	47.3
		3	203	3.163	6.803	0	39
	All	2	541	2.156	6.402	0	47.3
		3	477	4.352	8.400	0	52
Connected text fluency	Boy	2	288	17.572	18.360	0	87
		3	272	28.525	23.029	0	93.533
	Girl	2	250	14.080	18.971	0	96.583
		3	207	24.101	22.561	0	96.1
	All	2	540	16.034	18.763	0	96.583
		3	482	26.699	22.888	0	96.1
Reading comprehension	Boy	2	291	1.076	1.105	0	4
		3	273	1.637	1.232	0	4
	Girl	2	251	.952	1.151	0	4
		3	207	1.527	1.329	0	4
	All	2	544	1.024	1.133	0	4
		3	483	1.592	1.273	0	4
Listening comprehension	Boy	2	291	1.625	1.038	0	3
		3	273	1.890	1.012	0	3
	Girl	2	251	1.367	1.074	0	3
		3	207	1.889	.996	0	3
	All	2	544	1.509	1.062	0	3
		3	483	1.892	1.002	0	3

Appendix 2. Oral reading fluency norms for the United States, grades 1–8

TABLE 1
Oral reading fluency norms, grades 1-8

Grade	Percentile	Fall WCPM	Winter WCPM	Spring WCPM
1	90		81	111
	75		47	82
	50		23	53
	25		12	28
	10		6	15
	SD		32	39
	Count		16,950	19,434
2	90	106	125	142
	75	79	100	117
	50	51	72	89
	25	25	42	61
	10	11	18	31
	SD	37	41	42
	Count	15,896	18,229	20,128
3	90	128	146	162
	75	99	120	137
	50	71	92	107
	25	44	62	78
	10	21	36	48
	SD	40	43	44
	Count	16,988	17,383	18,372
4	90	145	166	180
	75	119	139	152
	50	94	112	123
	25	68	87	98
	10	45	61	72
	SD	40	41	43
	Count	16,523	14,572	16,269
5	90	166	182	194
	75	139	156	168
	50	110	127	139
	25	85	99	109
	10	61	74	83
	SD	45	44	45
	Count	16,212	13,331	15,292
6	90	177	195	204
	75	153	167	177
	50	127	140	150
	25	98	111	122
	10	68	82	93
	SD	42	45	44
	Count	10,520	9,218	11,290
7	90	180	192	202
	75	156	165	177
	50	128	136	150
	25	102	109	123
	10	79	88	98
	SD	40	43	41
	Count	6,482	4,058	5,998
8	90	185	199	199
	75	161	173	177
	50	133	146	151
	25	106	115	124
	10	77	84	97
	SD	43	45	41
	Count	5,546	3,496	5,335

WCPM: Words correct per minute
SD: Standard deviation
Count: Number of student scores

Source: Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher 59*(7): 636–644. Texts are designed to be appropriate for each grade level.