EGRA Liberia: Baseline Assessment of Reading Levels and Associated Factors

Report prepared under contract to the World Bank[1]
Luis Crouch
Medina Korda
RTI International
November 2008

Table of Contents

List of Tables

List of Figures

[1] Prepared as part of a process of collaboration between USAID and World Bank. This report prepared baseline knowledge for a USAID-funded experimental intervention.

Executive summary and layout

This document reports on the results of an Early Grade Reading Assessment (EGRA) study done in Liberia in June of 2008 as part of process of collaboration between the Ministry of Education, the World Bank, and USAID, with technical assistance provided under contract by RTI and its consultants and collaborators.

This executive summary reproduced the logic of the entire report and is thus also a guide to the contents and layout of the report.

This document first lays out the nature and content of the assessment instrument. The EGRA tool tests skills that are frequently needed in the early grades if children are to be provided with a solid basis upon which to progress with their reading in the later grades. The assessment approach was based on two foundations: a) a well-vetted default instrument that has received input from leading international reading experts at various workshops convened by USAID, the World Bank, and RTI, and b) input from Liberian experts at a workshop carried out in June. The assessment, in the end, had components on a) orientation to print, b) phonological awareness, c) letter naming fluency, d) familiar word fluency, e) non-familiar word fluency, f) fluency in reading connected text, g) comprehension based on read text, and h) a listening comprehension test.

The internal cohesion and reliability of the assessment is checked using various statistical procedures, and the reliability is found to be good, certainly in the range of other similar assessments used in both developed and developing countries. Depending on the exact version, for example, the alpha coefficient of reliability is above 0.8, which is a good benchmark (0.7 being considered an absolute minimum).

The role of one specific test-let, namely oral fluency in reading connected text, is prominent. This test-let or skill is seen to be the one that is most highly correlated with all other skills, and is the one most correlated with the simple measure of comprehension used in this EGRA application. For that reason we are justified in using as the most important benchmark skill for assessing various correlates. And, for progress assessments, in cases where there is only time or complexity-tolerance for assessing one single skill, it makes sense to focus on oral reading fluency in connected text.

Next, average levels of skill in the various test-lets are presented. A comparison with USA benchmarks is presented, not because there is any assertion that Liberia should aim for the same level as the USA, but just to provide a sense of comparison. Liberian children are found to be doing reasonably well on some skills, such as letter-naming, but oral reading fluency is very low, at around 18 correct words per minute at the end of grade 2. (Children reading less than 70 would be considered at risk in the USA.)

Comparisons of reading fluency are then made to see whether three key child-level variables, namely the grade, the age, and the gender of the child, affect the fluency. Grade of studies does affect fluency, which one would expect: the gain in fluency between grade 2 and grade 3 is 10 correct words per minute. Age, when holding grade

constant, does not have an important effect, which suggests ALP children do not do much better or worse than others. Boys do seem to read a little worse than girls, in grade 3 but not in grade 2, which may or may not signify a trend. It would be interesting to research whether this trend continues, as it would be worrisome if it does.

Finally, it is possible to compare various home and teacher practices for their apparent impact on fluency. Many of the practices were ascertained at the child level, others at the teacher level. Starting with the child level practices or conditions at home, the results indicate that many of the factors causing poor reading in the early grades in Liberia are factors that, in most developed countries (and for the middle classes in poor countries) are already provided for at home. In Liberia, for example, children who come from homes where mainly English is spoken, who have books at home, who read aloud at home, who have someone read aloud to them at home, etc., do much better than children who do not have those conditions. Similarly, children who are in schools where some of the home deficiencies are quickly corrected, such as getting onto letter naming quickly, and getting assigned practice to take home, seem to do better. In a sense, the mission of the "average" school in the early grades, then, might be to make available many of the most basic early skills that children otherwise would not get, and children from middle-class and upper-middle-class homes do get. These practices, taken together, if performed early, could boost learning by at least one grade level, perhaps more. At the teacher level, a positive finding is that having specific training in reading, and in-service training, are both associated positively with children's reading skills. Pre-service training is not as strongly positively associated as in-service training, and much less strongly associated as in-service training specifically oriented at reading. In that sense, the policy prescriptions somewhat write themselves, if one minds the findings: teacher training matters; specific teacher training in reading matters even more; and that training needs to result in teachers getting on with making sure that average and poor children get the same very basic advantages that many richer children already come to school with.

The rest of the report makes all of the above points in more detail.

1. Sampling approach, field work

The sampling approach started with a simple random sample of schools, proportional to enrolment, all over Liberia, and then went on to randomly pick 10 students per grade, for grades 2 and 3, in each selected school, from as many classes there were for any given grade. In each class with grade 2 and grade 3 children, the teacher was interviewed. In total 47 schools where sampled. Some were too small to have all the needed children, so a total of 429 and 407 children from grades 2 and 3, respectively, were sampled. A total of 110 teachers were interviewed, and all the principals.

2. Design effect considerations

Because schools were the primary sampling unit, and thus acted as clusters of children, there are considerations with regard to under-estimation of variance, or design effects.

Using the most important proxy variable, fluency in reading connected text, the following data give key indicators on the design effect.

Table 1. Design Effect Considerations

	ICC	Cluster size (children per grade)	DEFF	DEFT
Grade 2	0.271	9.32	3.25	1.80
Grade 3	0.295	8.83	3.31	1.82

Source: calculated by the authors from survey data

The DEFF and the DEFT (square root of DEFF) were calculated using the standard formula

$$DEFF = 1 + (clustersize - 1)\, ICC ,$$

but were also confirmed using a linearized calculation of the standard error of the mean, using standard Stata commands.

3. Performance of components of the EGRA tool as applied in Liberia

Components of the tool

The EGRA tool consists of a variety of sub-tools or test-lets, and they have been somewhat differentially applied in various countries. The specific Liberia tool assesses a set of skills as follows:

1. *Phonological 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").
2. *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 non-inferential (factual) questions. It is considered a pre-reading skill.
3. *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 1 minute, which saves time and also prevents the child from having to spend time on something they are having a hard time with.
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 1 minute.
5. *Non-familiar or non-sense word 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. It assesses the child's ability to "decode" words fluently. It is timed to 1 minute.

6. *Connected text oral reading fluency*: ability to read a passage, about 60 words long, that tells a story. It is timed to 1 minute.
7. *Comprehension in connected text*: ability to answer five questions based on the passage read.

The tool as a whole, and these components, were discussed with Liberian experts in the EGRA workshop help in Monrovia on June 16-19, 2008.

Does EGRA perform as an integrated tool?

A first issue to seek clarity on is whether the EGRA tool performs well as an integrated tool, in that all the components or test-lets "hang together" well. This can be assessed in various ways. An initial step is to look at the internal correlations between the seven various test-lets, as shown in Table 2.

Table 2. Correlation matrix for EGRA tasks

	Letter naming fluency	Phonological awareness	Familiar word fluency	Unfamiliar word fluency	Connected text fluency	Reading comp	Listening comp
Letter naming fluency	1.00						
Phonological awareness	0.27	1.00					
Familiar word fluency	0.53	0.31	1.00				
Unfamiliar word fluency	0.24	0.23	0.46	1.00			
Connected text fluency	0.53	0.29	0.78	0.44	1.00		
Reading comp	0.43	0.29	0.61	0.36	0.70	1.00	
Listening comp	0.22	0.25	0.24	0.14	0.24	0.27	1.00

Source: calculated by the authors from survey data

The correlations for phonological awareness and listening comprehension with other variables are poor across all other variables. This suggests that these two test-lets do not hang together as well as one might hope with the other parts of the EGRA tool. One may ask oneself why these correlations are so low. Perhaps all the students are "topping out" (all doing very well) or "bottoming out" (all doing very badly) on these tasks, which would then explain why these tasks are unrelated with the other tasks. But such is not the case. For phonological awareness, students managed to complete about 4.3 out of 10 tasks successfully, on average. Not stellar, but hardly bottoming out, and certainly not topping out. It could also be that the variable is very fixed at 4.3, students not showing much variation between them, but that is also not the case, since the task has a standard deviation of 2.4. The same goes for listening comprehension. Students got about 2

questions out of 3 right, with a standard deviation of 0.9. Thus, the problem is simply a low correlation.

It is important to note that the tightest internal correlation amongst all the items is the correlation between oral reading fluency in connected text and reading comprehension (0.78).

An analysis of the correlation between each test-let and the other test-lets, and the tool as a whole is also useful, and leads one to an analysis of the internal consistency of the tool. If one includes all the items, the following is the result, shown in Table 3.

Table 3. Reliability analysis of EGRA tool in Liberia

Item	Item-test correlation	Item-rest Correlation	Average inter-item correlation	Alpha
Letter naming fluency	0.67	0.53	0.37	0.78
Phonological awareness	0.55	0.38	0.41	0.81
Familiar words fluency	0.82	0.73	0.33	0.74
Unfamiliar words fluency	0.60	0.44	0.39	0.80
Connected text fluency	0.84	0.75	0.32	0.74
Reading comp	0.77	0.66	0.34	0.76
Listening comp	0.50	0.32	0.43	0.82
Overall test			0.37	0.80

Source: calculated by the authors from survey data

Clearly, phonological awareness and listening comprehension have lower correlations with the rest of the test than any other items. Nonetheless, the overall reliability of the overall tool is 0.80. The literature typically recommends a reliability of, at minimum, 0.70 for relatively low stakes issues (such as EGRA) and 0.80 for higher-stakes issues. So there seems to be a very reasonable margin of reliability here. It is interesting to note that removing phonological awareness and listening comprehension from the test would add to the overall reliability of the test, since the "Alpha" coefficient without those items would be 0.81 and 0.82 respectively. Finally, note that as the literature suggests, the item that most solidly connects with all others is fluency in reading connected text. If one draws a relatively arbitrary line at 0.60, one is tempted to see what happens to the overall reliability. The results are shown in Table 4.

Table 4. Reliability with two weakest elements removed

Item	Item-test correlation	Item-rest Correlation	Average inter-item correlation	Alpha without
Letter naming fluency	0.70	0.52	0.56	0.83
Familiar words fluency	0.86	0.77	0.45	0.76
Unfamiliar words fluency	0.64	0.45	0.59	0.85
Connected text fluency	0.89	0.80	0.43	0.75
Reading comp	0.80	0.66	0.49	0.80
Overall test			0.50	0.84

Source: calculated by the authors from survey data

Reliability goes up to 0.84, and connected text fluency is still the most important variable, in the sense that its correlation with the other variables is the highest.

Finally, a sense of the overall cohesiveness of the EGRA instrument can be derived by performing a principal components analysis. The point of this analysis is to assess whether there is a single underlying construct or ability ("component") that the instrument seems to pick up. The results are shown in Table 5. How the technique works is not particularly relevant. The point is that there is indeed one single "principal" component that "explains" or "accounts for" 62% (0.62) of the variation "contained" in the entire tool.[2] The next-highest component explains only 15% of the variation. One normally would like to see a proportion higher than 70% for the main component, but 62% is reasonable, particularly if the next component is much lower: this justifies the conclusion that the test does indeed have one "intrinsic" important component. Furthermore, in the second panel of the table, one can see the relative importance of each measured skill in constituting this underlying "intrinsic" measure of reading skill. All the test-lets have fairly high and even weights, which suggests that the principal "intrinsic" measure of reading skill is simply some sort of average of all the individual skills. Note that once again fluency in reading connected text has the highest "loading."

Table 5. Principal components of the EGRA Liberia results

Proportion of total variance explained by each "intrinsic" component		Weight of each reading skill on Comp 1	
Comp1	0.62	Letter naming fluency	0.39
Comp2	0.15	Familiar words fluency	0.50
Comp3	0.11	Unfamiliar words fluency	0.34
Comp4	0.07	Connected text fluency	0.51
Comp5	0.04	Reading comp	0.46

Source: calculated by the authors from survey data

[2] Please note that the two panels of the table are completely unrelated, and one cannot read across the rows. The data are presented in this manner only for the sake of convenience and space-saving.

We can conclude that the test "hands together" fairly well, and does seem to measure an intrinsic skill one could call "early reading" ability. One can also conclude that to a useful degree, and at least for analytical but not for pedagogical purposes, fluency in reading connected text is a useful proxy for that skill. This justifies using this fluency in reading connected text as a useful indicator which can be correlated with other various factors.

4. Sampling error and hypothesis testing ability of EGRA

Aside from the question of whether the components "hang together" and are reliable in that sense, it is important to ascertain whether the estimates for the key components are statistically reliable. Given the importance of oral reading fluency in connected text, as a variable, we focus our analysis on that key variable. It would be possible, but tedious, to report the same for all the variables.

Now, a random sample survey, such as EGRA, cannot tell us what the underlying average value for, say, oral connected text reading fluency is, for the underlying population of 2nd or 3rd graders. A survey takes a given random sample of children out of the population; a slightly different sample, with different children, out of the same underlying population, would have yielded a slightly different average. The statistical reliability of our measure, and our sample, is a way of judging how likely it is that different samples drawn from the same population would give us very different estimates of, say, oral reading fluency in connected text. One obviously seeks, in studies, to make sure that sample results do not vary from each other. This can easily be achieved by making the sample very big. Evidently, the bigger the sample, the smaller the probability of getting "odd" persons in the sample, and therefore the less the sample averages will vary from each other and from the underlying population average. However, making samples large is very costly. Researchers therefore try to balance out cost against reliability. The EGRA application in Liberia seems to have done this quite well.

Table 6. Statistical reliability of key EGRA variable: oral reading fluency in connected text

	Grade 2	Grade 3
Sample size	429	407
Mean ORF-connected text (in correct words per minute)	17.7	27.8
Assuming simple random sampling		
Std Dev	18.7	21.9
Std Err	0.9	1.1
Lower bound 95% CI	15.9	25.6
Upper bound 95% CI	19.4	29.9
Assuming children are clustered into schools		
Std Dev	33.8	41.4
Std Err	1.6	2.1
Lower bound 95% CI	14.4	23.6
Upper bound 95% CI	20.9	31.9

Source: calculated by the authors from survey data

A few facts stand out that one should keep in mind. First, the inter-grade difference is 10 correct words per minute. This is a little lower than in other countries where EGRA has been tried out, where the inter-grade difference is averaging out to about 14 correct words per minute. Second, the estimates are fairly narrow if one looks at the data assuming a simple random sample. For grade 2, one can state that the sample mean was 17.7, and one can state a "confidence interval" by saying "we are 95% sure that the underlying population mean is somewhere between 15.9 and 19.4." That is what "Lower bound 95% CI" means: it is the lower bound of a confidence interval for the mean. This, then, is a total interval width of only 3.5 points—a rather good result. Third, however, as noted above in section 2, there is a problem in that the assumption of simple random sample does not hold. A simple random sample would pick children completely at random from the schools. This is not only not feasible from a practical point of view, since no one has a list of all children at any central location from which one can sample, but is also very expensive, because visiting just one or two children per school implies enormous travel costs. Instead, EGRA applications first pick schools, and then sample about 10 children per grade—a clustered approach. However, because children within given schools vary less from each other than children at different schools, this process results in an under-estimate of the true variability of the population, and thus an over-estimate of the reliability of averages based on repeated sampling. One has to make certain corrections, which assume that the children are clustered within schools. This is shown in the second panel of Table 6. The standard deviations and standard errors are now larger, and the confidence intervals broader. Yet, even then, the total confidence interval width is only 5.5 points, still a very good result. Thus, even with this adjustment, we are still at least 95% sure that the oral reading fluency in connected text in grade 2 is lower than 20.9, and in grade 3 it is higher than 23.6.[3] That is a good separation between the grades, and allows us to be very certain that there is progression going on between the grades. Our estimate that there are about 10 words per minute of progression between the grades is quite safe. How safe exactly? This is hard to tell from the above table. However, a simple regression analysis, where one can estimate how much oral fluency in connected text varies as a function of grade, is possible. This analysis tells us that even taking into account the clustering, there is less than one chance in a million that we would have observed a difference of 10 words per minute between grades 2 and 3 if there were none. So, one can confirm that the sample size is adequate to justify these conclusions. The results are shown in Table 7.

Table 7. Influence of key factors on reading level

	Model 1, Only grade as a factor	Model 2, Grade, age, and gender (female=0, male=1) as factors		
	Grade	Grade	Age	Gender
Influence of grade, age, or gender on oral reading fluency	10.3	11.4	-0.5	-3.4
Std error	1.6	1.8	0.4	1.5
t value	6.3	6.4	-1.1	-2.3

[3] This is slightly inaccurate as to make these statements one would need one-sided confidence intervals, but it is close enough, as one-sided confidence intervals are, in any case, slightly less conservative.

p value	0.000	0.000	0.272	0.026
Lower bound CI	7.0	7.8	-1.4	-6.4
Upper bound CI	13.6	15.0	0.4	-0.4

Source: calculated by authors from survey data

The table confirms that going through a grade (more specifically, going from grade 2 to grade 3) adds 10.3 words per minute to the oral reading fluency, in the model that has only the factor "grade" as an influence on reading, and that this is statistically very significant (this can be read from the p value: there is less than a 1 in 1000 chance of observing an effect as large as 10.3, in a sample of the size we have used, if there is no effect). In the model that allows for the influence of age and gender, we can see that age has a small negative influence (older children read less well) but *this is not statistically significant*. In this sense, ALP children are not reading either better or worse than other children, but note that this is not a major consolation as this is likely due to the fact that all are reading with low fluency. Being a boy does affect reading in a negative fashion, holding grade and age constant, though the magnitude of the effect is not large (3.4 words per minute).

5. Performance on other variables and benchmarks

It is worthwhile to look at performance on other variables against some possible benchmarks, without necessarily delving into the details of statistical reliability. We have anchored the remarks on statistical reliability on the most important variable, oral fluency in connected text. For the other variables we will just compare against at least one set of benchmarks. Table 8 shows the results for Liberia for all the key variables, and the US benchmark. *There is no pretense that the US benchmarks are appropriate for Liberian children*. We present these benchmarks at the risk of being misunderstood and misquoted, so we must emphasize that *it is not appropriate to suppose that Liberian children should be reading as well as US children.* Liberia is a much poorer country, with a recent history of conflict. Children have very few materials at school. On the other hand, teaching reading is not something beyond even poor societies, if enough political will and training is devoted to it. So in some sense the benchmarks do give some general guideline.

Table 8. Base values for Liberia and US benchmarks

	Grade 2 Liberia	Grade 2 USA at-risk benchmark	Grade 3 Liberia	Grade 3 USA at-risk benchmark
Letter naming fluency	64	NA for Grade 2, only established for Kindergraten, greater than 40 no risk	73	NA for Grade 2, only established for Kindergraten, greater than 40 no risk
Familiar word fluency	14	NA	20	NA
Non-familiar word fluency	2	NA for end of Grade 2, Less than 30 considered weak in Grade 1	4	NA for end of Grade 3, Less than 30 considered weak in Grade 1
Connected text fluency	18	Less than 70 considered at risk	28	Less than 80 considered at risk

Source: calculated by the authors using survey data in the case of Liberia, DIBELS benchmarks and goals

for the US.

6. Factors associated with reading achievement – child and teacher level

In most applications thus far, the EGRA application has involved simply an assessment of reading ability. At the request of the top leadership of the Liberian Ministry of Education, the Liberia assessment looked a good bit into factors associated with children's reading skills. The term "factors associated" is used with deliberation. It is impossible, using a survey and correlational sorts of analyses, to claim unequivocally that the associated factors "produce" or "cause" reading skills. To prove this claim beyond doubt would require experimental trials beyond what has been done thus far in Liberia. Instead, we rely on a) correlation and an intuitive, b) a theoretical sense of factors that are likely to "cause" reading, and c) research from other countries. It is well known from other societies and experiments, for example, that if children spend a good bit of time reading, they do learn to read better—hardly a surprising finding, but one often ignored in classroom and homework practice. Thus, if we find, in the case of Liberia, a good correlation between children spending time reading, and their reading results, it is a safe bet that there is some sort of causality at work. Nevertheless, it is good not to over-reach and claim that simply getting children to spend more time reading both at school and at home will produce, by itself, and as a sort of miracle, the desired effect. Other factors need to be brought in. This section looks at all of these factors together.

The first set of factors is derived from child-by-child analysis. Table 9 shows the influence of these various factors. The table is somewhat complicated and thus bears a little explanation. The column entitled "average reading level" simply refers only to the "Memo items" that recall for us the average reading fluency (in connected text) in Grades 2 and 3. We are thus reminded that the average difference between Grades 2 and 3 is about 10 correct words per minute: from 18 to 28. *Columns two and three, respectively, show the average reading fluency of children who lack the given factor, and children who have it*. Column four shows the difference between the two (apparent disagreements are caused by rounding.) Thus, reading at the bottom of the table, for example, children who read aloud to someone at home read 10 words per minute more fluently than children who do not. That is about one grade level difference, i.e., equivalent to the difference between grade 2 and grade 3. The last column shows the percentage of children who are currently exposed to the positive factor, as a way of suggesting where there might be a good impact. If a factor is positively associated with good reading, but is already apparently widely practiced, it might not make a good deal of policy sense to focus on that factor. Similarly, if the factor cannot really be associated by educational policy and practice, such as English being spoken at home, and whether the home has a TV (indicators of socioeconomic status and/or precursors of school-related skills), a note is made that the factor cannot be directly affected by policy. However, it is important to at least note these factors, because they affect what schools should do: in some sense the mission of schools, for the poor and the lower-middle-class, is to quickly try to get children to catch up, and acquire some basic skills and practices that upper-middle class children may already have.

Table 9. Factors taken singly and reported by child: impact on measured reading in grades 2 and 3

	Average reading level	1 Children without the factor (average fluency in correct words/min)	2 Children with the factor (average fluency in correct words/min	2 – 1: Difference children with and without the factor	Percent with the factor
Memo item					
Average reading level in Grade 2	18				
Average reading level in Grade 3	28				
Factors with apparently negative influence					
Have failed a grade		24	20	-4	
Missed school days previous week		23	21	-2	
Factors with apparently (relatively) weak influence					
Teacher practice sounds		22	24	2	
Have radio at home		22	23	2	
Ate breakfast day of assessment		22	24	2	
Had lunch during school break		21	24	4	
Does homework		19	24	5	70%
Mother reads/writes in English		21	26	5	No policy to change this
Have library at school		22	27	5	12%
Father reads/writes in English		19	24	5	No policy to change this
Attended some form of pre-school		18	23	6	89%
Factors each of which is associated with difference nearly equal to one grade level or statistically significant					
Reads aloud to own class		18	24	6	75%
Have TV at home		21	28	7	No policy to change this
Someone reads aloud to child at home		19	27	8	51%
Speak English at home		20	28	8	No policy to change this
Homework 4 or 5 days / week (versus less)		21	29	8	16%
Practice silent reading at school		17	26	9	62%
Have reading books at home		20	30	9	26%
Teacher reads aloud to child		14	23	10	75%
Reads aloud to someone at home		17	28	10	51%

These factors are all taken singly, or one at a time. It is possible to do a form of statistical analysis that takes factors into account all at the same time. If one does this kind of analysis, one can provide a list of the factors that are jointly associated with better reading. This is presented in the following list:

- Whether child reads aloud at home
- Whether English spoken at home

- Whether children have TV at home
- Whether children have books to read at home
- Whether children read on their own at school
- Whether children practice letter sounds at school

These results essentially confirm the results found when taking the factors one at a time. One slight but important difference is that the importance of letter practice shows up much more strongly when all the other factors are considered together with it. That is, taking everything else as fixed, practicing letter sounds makes a difference.

A similar analysis was done with teacher-related factors. The results are not quite as rich, but they are still quite illuminating. The three most important factors associated with better children's reading were:

- Whether teacher received in service training last year
- Whether teacher has been specifically trained in teaching reading
- Whether teacher holds understanding as a key goal of reading

What is one to make of all this? The data strongly suggest that very basic skills, possessions, and practices, such as reading aloud, having access to books at home (or, presumably, that can be taken home), and so on, make a significant difference. Upper-middle class children can take these for granted. Poor and lower-middle class children cannot. The notion then suggests itself that a productive duty of the school should be to make up for these differences in basic and early habits, possessions (e.g., reading materials that can be taken home), and practices. In addition, the results from the teachers show that training matters, and that training specifically in how to teach reading matters even more. Putting the two results together in a fairly basic recommendation: improve, increase, and intensify training; make it as specific to reading as possible; and make the training focus on supplying the basic skills that the upper-middle class children can take for granted but the lower- and lower-middle class children are not getting.

7. Factors associated with reading achievement – school level

School level factors were assessed via principal interviews. Unfortunately, none turned out to be of much interest, at least with the sample size chosen. (It is important to recall that the sample size chosen was not meant to elicit statistically reliable information at the school level, so school effects either have to be taken heuristically or, if statistically significant, as a sort of bonus. The sample size was calibrated to allow statistical inference about children's reading, not about the school-level correlation between school factors and the levels of reading.) Table 10 shows some of the correlations.

Table 10. School and Principal Factors and Reading

Variable	Correlation coeff.
Whether principal observes teachers	-0.20
Whether inspection visits were received last year	-0.24
Whether feeding program exists	-0.36**
Whether dress code is enforced	0.04
Whether have sufficient resource materials	0.00
Whether have lock box for books	0.06
Whether have library room	0.14
Whether have regular PTA meetings	-0.09
Attendance at PTA meetings	0.10
How often teachers go for in-service training	-0.20
Years exp as principal	0.04
Principal gender	0.04
Principal has C certif. or higher	0.04

Source: calculated by the authors from survey data
Note: "reading" variable is fluency in connected text
** statistically significant at .05 level

Some of these factors might turn out to be statistically significantly different from zero if a larger sample were to be used. But the more interesting aspect of all this is that few of the correlations are substantively of interest (recall that a bivariate correlation is the same thing as an effect size), or have a sign opposite to what one would expect. The only variable that is statistically significantly correlated to reading levels is whether a feeding program exists at the school, and the correlation is negative. This is a useful reminder, of course, of the usual problem of "correlation versus causality:" most likely feeding programs are targeted to the poorer, and the poorer do not read as well. If one were truly interested in the effect of feeding programs one would have to carry out an experimental program, or at the very least control for other variables using multivariate techniques. With a sample size of only 46 principals, multivariate analysis is unlikely to be of much interest.

It is important to note that some of the correlations shown in Table 10 are low because there is essentially no variability in the variable in question. For example, almost all teachers (close to 100%) are said by the principals to go to in-service training more than every five years.

An interesting factor which is unfortunately also not statistically significantly correlated is the grade at which the best teacher in the school teaches, according to the principal. The grade where the best teacher is found is correlated with the reading fluency in connected text and in familiar unconnected words. The results are shown in Table 11. The table bears a little explanation. Each cell shows the correlation between the level of reading of children in grades 2 and 3, and a dichotomous variable for whether the best teacher teaches the grade named in the row. The table thus shows that the peak correlation between the reading levels of children in grades 2 and 3 and the location variable is when the location variable refers to grade 1.

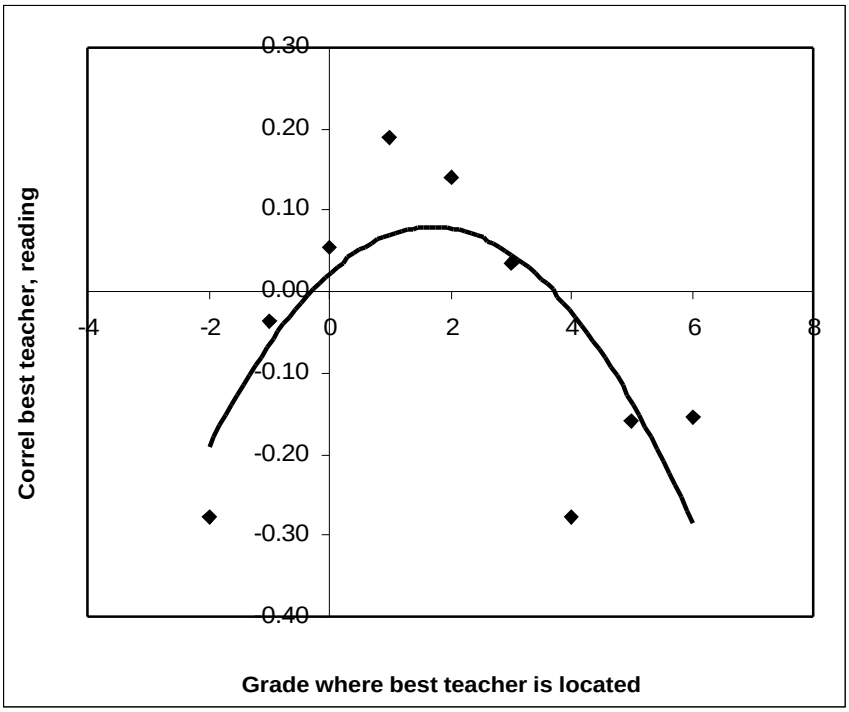
Table 11. Correlations Between Reading Levels and Location of Best Teacher

Best teacher in	Connected text fluency	Familiar word fluency
Pre-K	-0.28	-0.28
K1	-0.04	-0.03
K2	0.05	0.11
Grade 1	0.19	0.19
Grade 2	0.14	0.20
Grade 3	0.03	0.13
Grade 4	-0.28	-0.23
Grade 5	-0.16	-0.14
Grade 6	-0.16	-0.15

Source: calculated by the authors from survey data

Unfortunately none of these individual correlations are statistically significantly different from zero, but the pattern is what one would expect: schools in which the test teachers are in grades 1 or 2 generally have the best reading levels (as measured by reading in grades 2 and 3). The pattern is shown in Figure 1.

Figure 1. Correlations: Grade of Best Teacher, Reading Levels



While, as noted, not statistically significant, the pattern reaffirms the correlation between teacher quality and reading levels, and suggests that teacher quality in the early grades is important to early grade reading (and, noting that early reading lays the groundwork for all reading, probably later reading as well).

8. Conclusion: summary of factors affecting design of intervention

The assessment described in this note was not itself a rigorous study of the impact of interventions. But it was used to underwrite, with local data, suggestions as to how to improve reading that could then be tested with an intervention project which has an experimental nature. The study underwrites the focus of the intervention by calling our attention to the following factors, all of which will be emphasized in the intervention:

1. A focus on the classroom teacher as the point of the intervention
2. A focus on in-service, not pre-service training
3. A focus on reading as the target of the intervention
4. A focus on teachers in the early grades
5. Focus on materials that would underpin reading activities in the classroom and reading at home; provision of both "exercise" reading books (decodable books for class use) as well as libraries for extra and home reading
6. Reporting to parents and communities as to levels of reading achieved, as a way to elicit involvement
7. Asking parents to vouch for the fact that children are reading aloud at home in a formal manner

These factors will be strengthened in the intervention schools in the USAID-funded EGRA Plus project. The project will randomly allocate schools to three groups: a control group, a "Light" treatment group which focuses on factor 6 in the list above, and a "Full" treatment group which attempts to implement all of the factors.