Proposing Benchmarks for Early Grade Reading Skills in Liberia

BACKGROUND

From 2008, beginning with EGRA+ and continuing under the Liberia Teacher Training Program (LTTP), USAID has supported improved early grade reading in Liberia. Since its inception, the LTTP supported intervention to improve early grade reading is being implemented in 1,200 schools. The most recent analysis of that program's impact (based on a reading assessment conducted towards the end of the 2012-13 school year) showed that, on average, children in grades 1 through 3 in participating schools significantly improved their achievement in all the reading skills tested. From 2011 to 2013, grade 1 students on average almost tripled their ability to read familiar words, more than tripled their decoding skills, and were able to read text more than one-and-a-half times more fluently. Grade 2 and 3 students on average were 10 times better at decoding and had almost three times higher oral reading fluency than before the program. On average, students in all grades improved their listening and reading comprehension.

However, even grade 3 students were still not on average attaining a level of reading skill adequate to ensure full comprehension of what they read. Liberian student performance in reading remains far below U.S. or other developed country standards. This begs the question, "What is an acceptable level of reading achievement for Liberian students in the early grades of primary school?"

With additional assistance from USAID a workshop to address this question took place during the first week of March, 2014.

Ministry of Education officials, district education officers, and a cross section of stakeholders attended this workshop over two days to begin Liberia's first ever effort to define standards for student performance in key areas of reading skill development in grades 1, 2 and 3.

LIBERIA IS UNIQUELY POSITIONED TO UNDERTAKE AN INFORMED PROCESS OF SETTING READING BENCHMARKS

Only a handful of developing countries have taken on the challenge of setting benchmarks for reading skills in early grades. Mexico did so several years ago. And more recently Kenya and Egypt have defined benchmarks, with Kenya officially adopting a standard for oral reading fluency in both English and Kiswahili. Liberia has the distinct advantage of having a bounty of data to inform the setting of reading skill benchmarks.

Grade 2 Performance and Improvement Over Time in EGRA Plus And LTTP Schools
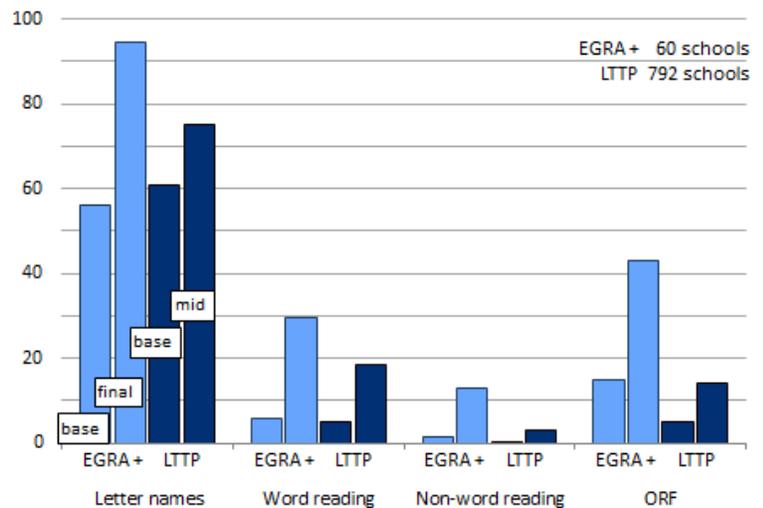
As shown in the graph above, the EGRA+ and LTTP programs have baseline and subsequent measures of reading performance that show not just how well students perform in different skill areas, but also how much improvement can be achieved through a targeted instructional intervention. This provides a realistic foundation from which to discuss what benchmarks may be most appropriate for the current Liberian context.

THE BENCHMARK SETTING WORKSHOP

A two day workshop on March 6 and 7, 2014 brought together 60 MOE officials, district education officers, donor agency representatives, NGOs active in the education sector, LTTP project staff, and outside experts, to begin a process of defining benchmarks for specific skill areas of early grade reading. The objectives of the workshop were to:

- Share the most recent assessment results from LTTP's reading intervention

- Orient and engage a cross section of Liberian stakeholders in a participatory process of setting reading benchmarks for grades 1, 2 and 3.

During the first morning of the workshop, data from the LTTP midterm assessment and from the EGRA+ end line assessment were shared and discussed. Data from international (PIRLS 2011) and from U.S. assessments of reading were also shared. In addition to providing points of comparison, these data helped illustrate that reducing the percentages of students scoring at the lowest levels is a key strategy for improving a country's overall performance.

Following the presentation, participants were engaged in a discussion of benchmarks – what they are and how to set them by combining empirical data both from other countries and from Liberia, working knowledge of Liberia's education sector, and common sense. Overall, the ratio of presentation to participatory work was about 1 to 2.5.

Small working groups took on the challenge of analyzing the available information, discussing and debating what seemed possible, and then defining an initial set of benchmarks for grade 3.

Those results were shared and discussed, prior to moving on to setting benchmarks for the other grades.

At the end of the workshop, all the groups' points of view were recorded and areas of convergence and divergence in recommended benchmarks were identified and discussed so as to generate further convergence. This report shares those results, showing the full range of points of view advocated by the participants, and concluding with the recommendations of the authors as to what may be useful "converged" benchmarks for early grade reading in Liberia at the present time.

THE READING SUBTASKS

The policy workshop helped define benchmarks for three reading subtasks evaluated using the Early Grade Reading Assessment (EGRA) in grades 1, 2 and 3. The three EGRA subtasks include:

- **Non-word fluency.** This subtask evaluates a student's ability to decode unfamiliar words. Short combinations of three letters (often consonant, vowel, consonant) that do not form words (e.g., "nak") are used so that the assessment can distinguish the skill of decoding from the skill of whole word reading. The subtask is timed, so the resulting measure is the number of non-words decoded correctly per minute.

- **Oral reading fluency.** This subtask evaluates how well a child reads out loud a coherent, short passage of text. It is also timed, and therefore produces a measure that is the number of words of text correctly read per minute.

- **Reading comprehension.** Students are asked five questions relating to the text which they would have read aloud for the oral reading fluency portion of the assessment. The resulting measure is a number or percent of correct responses out of five.

THE BENCHMARK SETTING PROCESS

Working in seven separate small groups, participants received data tables showing how students performed on these subtasks at the start and end of EGRA+ (2008 and 2010) and at the start and midpoint of LTTP (2011 and 2013).

Data that expressed the relationship between these subtasks were also shared. For example, a scatter plot of oral reading fluency and comprehension showed that students who demonstrated comprehension at 80% or better (answering 4 out of 5 questions correctly) were for the most part

reading with oral fluency of between 45 and 65 words per minute. Similar data were used to demonstrate the relationship between students' decoding abilities (as measured by non-word reading) and their levels of oral reading fluency. In addition, some international data were shared. These data helped participants to use a benchmark set in one area – say comprehension – to define the benchmarks in the other skill areas.

Each group, armed with these data and their own vast working knowledge of the education system in Liberia, was asked to:

- For each subtask, define three aspects:
  o The benchmark value for the indicator for that subtask,
  o The percentage of students that would be meeting that benchmark in five years, and
  o The percentage of students who would be scoring zero on that indicator in five years.
- Define the above values for grade 3
- For grade 3, define first the values for reading comprehension (benchmark, percentage meeting the benchmark and percentage scoring zero) and then use that to inform the "needed" values for the two other skill areas.
- Having completed this work for all three subtasks (comprehension, oral reading fluency and non-word reading) for grade 3, the groups reconvened in plenary to compare, discuss and arbitrate among their responses.
- Following the plenary discussion, the groups were charged with first revisiting what they had proposed for grade 3, then, using the grade 3 values they decided on, define the levels that would propose for each of the three subtasks for grade 2 and then grade 1.
- Group work was interspersed with short plenaries to clarify concepts and document convergence.

THE BENCHMARKING RESULTS

The table below summarizes the results of the group work defining benchmarks for all three subtasks for grade 3. Each subtask skill is a shown in a column in the table. In addition to the output from the groups,

relevant data from the LTTP midterm is shown as a point of comparison to the standards proposed by the workshop participants.

Grade 3	Reading Comprehension[1]	Oral Reading Fluency[2]	Non-word Reading / decoding[3]
Average performance at LTTP midterm	*22%*	*20.2*	*4.5*
Proposed benchmark	**60-80%**	**45-50**	**10-20**
% of students meeting proposed benchmark (LTTP midterm)	*8%*	*17%*	*18%*
Proposed % of students at benchmark in 5 years	**45-50%**	**40-50%**	**30-40%**
% of students scoring zero (LTTP midterm)	*44%*	*10%*	*52%*
Proposed % of students scoring zero in 5 years	**20-30%**	**5-15%**	**25-35%**

[1] % of questions answered correctly; [2] correct words per minute; [3] correct non-words per minute

Concerning reading comprehension, the workshop participants discussed a benchmark target reading score of between 60% and 80% correct. This compares to the average performance at LTTP midterm of 22%. The group in general wanted to set a standard well above what grade 3 students are scoring now, reasoning that the standard should reflect a decent level of comprehension of grade appropriate text by the end of grade 3. They were also informed by what EGRA+ was able to achieve, while conscious that a pilot project is quite different from a national scale-up.

While setting the benchmark somewhat high, the group was more modest in their estimation of the percentage of students who would be at that benchmark in 5 years' time. The groups converged around 40-50% of students being able to meet the benchmark of 60 to 80 percent comprehension. Compared to LTTP midterm –only 8% of grade 3 students achieved 60% or better – the target of 40-50% would represent significant improvement. The same could be said for zero scores – the groups all proposed a 5 year target of fewer students scoring zero in reading comprehension than did so on the LTTP midterm.

Results for grades 2 and 1 are presented below.

Grade 2	Reading Compre-hension[1]	Oral Reading Fluency[2]	Non-word Reading / decoding[3]
Average performance at LTTP midterm	*14%*	*14.2*	*3.2*
Proposed benchmark	**40-60%**	**35-40**	**10-15**
% of students meeting proposed benchmark (LTTP midterm)	*10%*	*19%*	*23%*
Proposed % of students at benchmark in 5 years	**40-50%**	**40-50%**	**25-40%**
% of students scoring zero (LTTP midterm)	*59%*	*27%*	*64%*
Proposed % of students scoring zero in 5 years	**25-30%**	**15-20%**	**25-40%**

[1] % of questions answered correctly; [2] correct words per minute; [3] correct non-words per minute

Grade 1	Reading Compre-hension[1]	Oral Reading Fluency[2]	Non-word Reading / decoding[3]
Average performance at LTTP midterm	*22%*	*20.2*	*4.5*
Proposed benchmark	**40-60%**	**30-40**	**5-10**
% of students meeting proposed benchmark (LTTP midterm)	*3%*	*6%*	*7%*
Proposed % of students at benchmark in 5 years	**30-45%**	**30-50%**	**25-40%**
% of students scoring zero (LTTP midterm)	*44%*	*10%*	*52%*
Proposed % of students scoring zero in 5 years	**30-50%**	**15-35%**	**25-50%**

[1] % of questions answered correctly; [2] correct words per minute; [3] correct non-words per minute

As was the case for grade 3, the workshop participants proposed benchmark targets that surpass the levels of performance seen on the LTTP midterm. The groups set benchmarks above the LTTP midterm averages in each skill area and proposed more students meeting those benchmark levels of performance than had previously did so. Also, the standards reflect a progression of increasing levels of achievement from grade 1, to grade 2, to grade 3.

During the benchmarking process, there was much lively discussion and debate about how much improvement over the present levels of performance one could expect to see. Groups vacillated between being ambitious and setting standards well above current levels of achievement and being realistic, if not pessimistic.

Often the question was raised as to what the groups should assume the MOE and its partners would be doing during the next five years to improve reading instruction. This led to considerable debate about whether the education sector in Liberia had sufficient resources, capacity and know-how to bring about dramatic improvements in reading performance. The encouraging fact was that the LTTP project (and before it EGRA+) had demonstrated that it is possible to improve reading outcomes in Liberia. Whether concerted effort can be continued and in fact broadened to address the needs in all schools across the country is the paramount concern. The proposed benchmarks have to assume that concerted effort will be made—otherwise, in a sense, there is no point in setting benchmarks. In fact, non-achievement of the benchmarks would be a tell-tale sign that not enough resources and effort are being mobilized

CONCLUSIONS AND RECOMMENDATIONS
After careful consideration of the work produced by the participants in the Benchmarking Workshop, the following table summarizes what we would recommend to the MOE, its partners and stakeholders as standards for reading performance in the three skill areas across the three grades.

Recommended Benchmarks	Reading Compre-hension[1]	Oral Reading Fluency[2]	Non-word Reading / decoding[3]
Grade 3 Benchmark	75%	50	15
Grade 2 Benchmark	50%	35	10
Grade 1 Benchmark	40%	30	8
Grade 3, % at benchmark in 5 years	50%	50%	50%
Grade 2, % at benchmark in 5 years	50%	50%	50%
Grade 1, % at benchmark in 5 years	50%	50%	50%
Grade 3, % scoring zero	15%	5%	15%
Grade 2, % scoring zero	20%	15%	20%
Grade1, % scoring zero	20%	15%	25%

[1] % of questions answered correctly; [2] correct words per minute; [3] correct non-words per minute

Note that all benchmarks and indicators are proposed based on continuing to use a single assessment aligned to grade 2 in all three grades. The approach of using a common assessment across the three grades allows us to easily evaluate differences in reading performance from grade 1 to grade 2 to grade 3. This is what we recommend continue to be the approach in Liberia for monitoring of progress, fully recognizing that at the classroom level the expectation would be that teachers and students are working with grade appropriate materials and that teachers would evaluate their students accordingly. It is at a system level that it makes sense to monitor progress (for the time being) against a fix, single grade level of material.

The recommended benchmark levels of performance for reading comprehension, oral reading fluency and non-word reading for grade 3 are at the lower end of the ranges proposed by the working group participants. We are recommending the less ambitious benchmark because current achievement is so low and because we prefer being more ambitious regarding the other two indicators – the percentage of students meeting the benchmark and the percent scoring zero in five years.

Beginning with comprehension, we reasoned that by the end of grade 3, students should be attaining a reasonable level of comprehension. Therefore they should be getting at least 75% of comprehension questions correct. The level of oral reading fluency that is associated with 75% comprehension in the Liberia data is 45 to 65 words per minute, so a standard of 50 wpm seems appropriate for assuring the desired level of comprehension. In a similar manner, the benchmark for non-word reading can be defined - the level of decoding skill students need to be reading with fluency approaching 50 words per minute.

Where the recommendations diverge from what was put forth in the workshop is in the standards for the percentage of students meeting the benchmark and in the percent scoring zero. We recommend a somewhat higher percentage of students meeting the benchmark levels and foresee that as being consistent across all subtasks and grades. The reasoning being that the system should strive to have at least half the students meeting benchmark performance in all skill areas.

In relation to zero scores, we recommend targets for the percentage of students scoring zero that are at the low ends of the ranges proposed in the workshop because we think this is where the system should target its improvement efforts. Overall performance is best raised by improving the achievement of students at the lowest ends of the distribution. Both EGRA+ and LTTP (and similar interventions in other countries) have been successful at reducing zero scores so we recommend a slightly more ambitious approach to this key indicator.