

Comparison of Canadian Language Benchmarks Reading Assessment and Measure Up Tasks- Field Test 2

Prepared by Christa Bruns
May 13, 2004

Comparison of Canadian Language Benchmarks Reading Assessment and Measure Up Tasks- Field Test 2

The purpose of this field test is to continue to examine the relationship between Measure Up tasks and the Canadian Language Benchmarks Assessment. During Field Test 1, 27 participants from Academic English programs at SAIT and Chinook College were tested with both the Measure UP (Self Assessment 4 and 6) and the CLBA (Stage 2 Reading and Writing and the Listening and Speaking interview). Following the recommendations of Field Test 1, Field Test 2 differs in two important ways. Firstly, the participants were recruited from a lower proficiency group (LINC 3 and 4). Secondly, this field test focused exclusively on Reading and did not include the listening/speaking interview, nor the writing component of the CLBA.

1. Narrative of activities

Recruitment- For this study, 2 classes of students from the Calgary YMCA were invited to participate. 26 students took part in the study; 13 each from LINC 3 and LINC 4. One participant did not complete Stage 2 although he qualified. Therefore the total number of participants was 25.

Prescribed procedure for administration of CLBA Reading.

In administering the CLBA Reading assessment, the procedure is to begin with Stage 1 and allow 45 minutes. If the client qualifies by receiving a composite score of 11 or 12 out of a total of 12; he or she is eligible to attempt Stage 2. For Stage 2, the client is allowed up to 1 hour to complete the test.

Testing Procedure On May 6th, 2004 during the participants' regularly scheduled ESL class, they took both the Measure Up tasks and the Stage 1 CLBA Reading. Some participants in LINC 4 completed Stage 1 so quickly that they were able to go on to Stage 2 immediately. The reason that they were asked to write Stage 2 without waiting to mark Stage 1 is that while they were writing, their teacher showed the researcher her students' previous teacher assigned benchmarks. Most of their reading benchmarks were above 4. However, in the LINC 3 class most of the students' reading benchmarks were 4 and below.

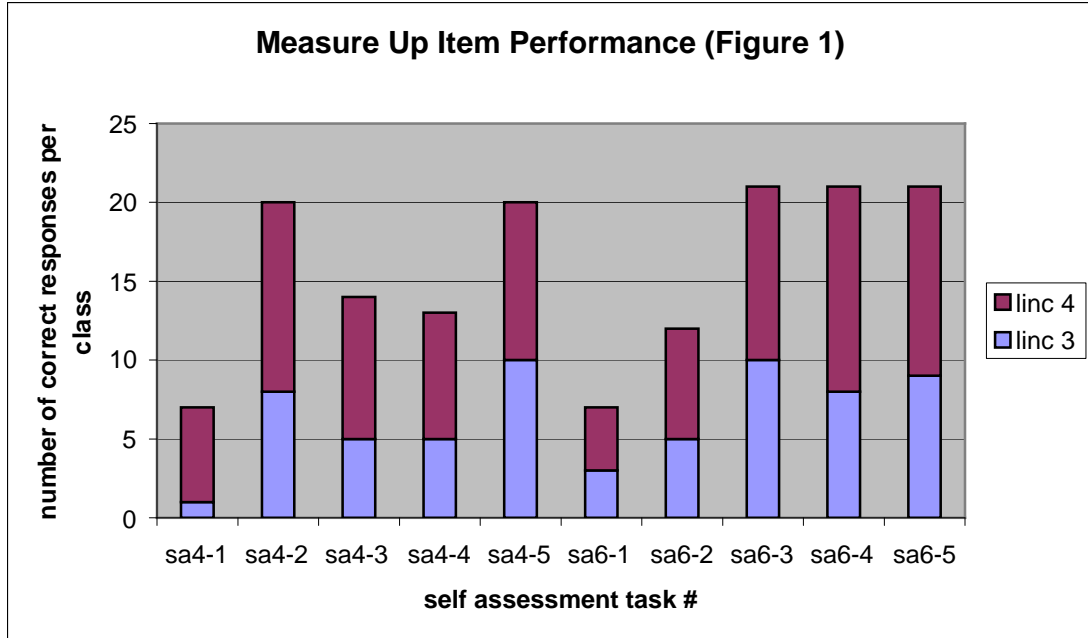
On May 10th, the researcher returned to the classes to administer Stage 2 Reading to 6 participants who were found to qualify based on their performance on the Stage 1 Reading assessment. 5 of those people were present that day.

The participants completed the *Measure Up* tasks in a group setting. The time required for participants to complete Measure Up was measured and the participants were informed that the time would be recorded. Meanwhile, the time required for CLBA was not measured.

Upon completing the Measure Up tasks, the participants were asked to write their comments about the test. The comments they wrote were largely off topic and vague-unusable. They were being asked to essentially evaluate an unfamiliar test and record their results in a meaningful way. Upon reflection, this was too difficult a task to request of that level of learners. In Field Test 1, the participants were asked orally one-on-one about their impressions. That was much more effective.

The participants had done either the CLBA or CLBPT previous to this study. Those are the two tests used to place them in Language Instruction for Newcomers (LINC) classes. On the other hand they were not familiar with the *Measure Up* tasks.

2. Analysis of Measure Up Test Items



By comparing the results of self assessment 4 and self assessment 6 of the Measure Up test, we can see that the participants' performance on self assessment 6 was better than on self assessment 4. Refer to the graph above showing performance of classes on each item as well as the table below showing the class average for each part of the test.

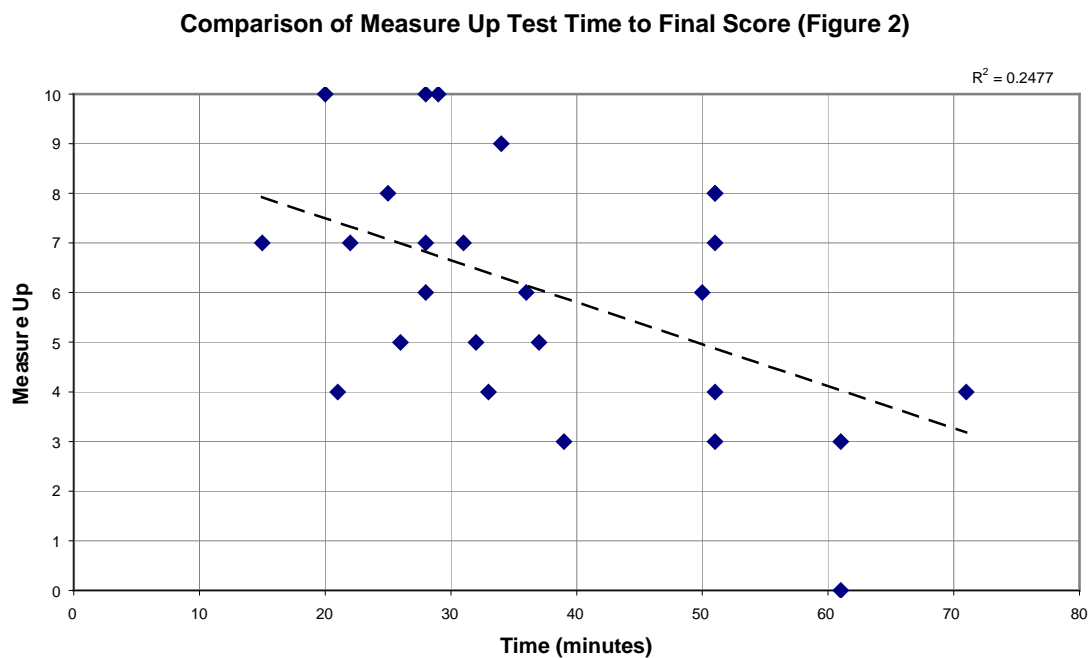
Class	Self Assessment 4	Self Assessment 6
LINC 3	45%	56%
LINC 4	71%	72%
All participants	57%	64%

Note: We acknowledge that dividing the group according to their class is not particularly valid in view of the fact that participants were not placed in their classes based on their CLBA Reading results. Indeed there were 3 individuals in LINC 4 with benchmark 3, and there were 3 individuals in LINC 3 with CLBA benchmarks over 3. However they did take the tests in these groupings.

3. Analysis of Test Times

We would expect better readers to both read faster and answer more questions correctly. Because we measured how long students spent on the Measure Up test, we

were able to compare time in minutes to results on the measure up test in Figure 2. As you can see, there is a negative correlation (the line slants to downward to the right). However, the R^2 value is only 0.2477. This suggests that there were other strong factors that impacted the length of time participants spent on the test. These could include factors such as various test taking strategies, various reading strategies, and perceived pressure to finish at the same time as other participants for example. Generally, the higher benchmark participants were in LINC 4 and the lower benchmark students were in LINC 3. When we sorted the data according to benchmark rather than class, the outliers (in terms of amount of time taken) were from the other class; emphasizing the impact of group dynamics on length of time taken.



Again, if we look at each class, the results are clear that indeed the more proficient readers did spend less time and get better scores. Refer to the table below.

Class	Time in Minutes	Average Score on Measure Up	Average Benchmark
LINC 3	48.75	4.75	3.83
LINC 4	26.62	7.08	6.08

4. Comparisons of CLBA Reading Benchmarks to Measure Up Results

This comparison of CLBA reading benchmark to Measure Up total score is the most important part of this report. As such, we start with the overall comparison in table form. In the following table, we see the breakdown of the group by benchmark.

Number of participants	CLBA Reading Benchmark	Average Measure Up Total	Average Time in Minutes for Measure Up
1	2	4	71
10	3	4	41
0	4		
3	5	6	38
3	6	7	36
4	7	9	28
4	8	7.5	22

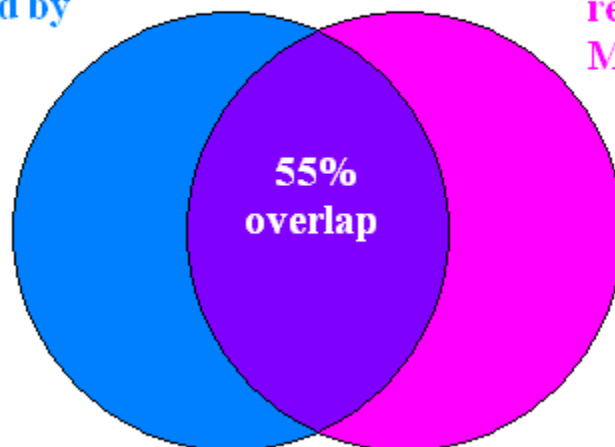
Although the number of participants in each benchmark group is small, we do see emerging patterns. As the benchmark increases, the average score on Measure Up increases. The upset in benchmarks 7 and 8 is probably due to individual differences in the small number of participants. Also, as the benchmark increases the time required for the Measure Up tests decreases.

You will notice the absence of benchmark 4 participants. This is due to the test design, 4 being on the cusp of Stage 1 and Stage 2. This issue was discussed at length in the report of Field Test 1.

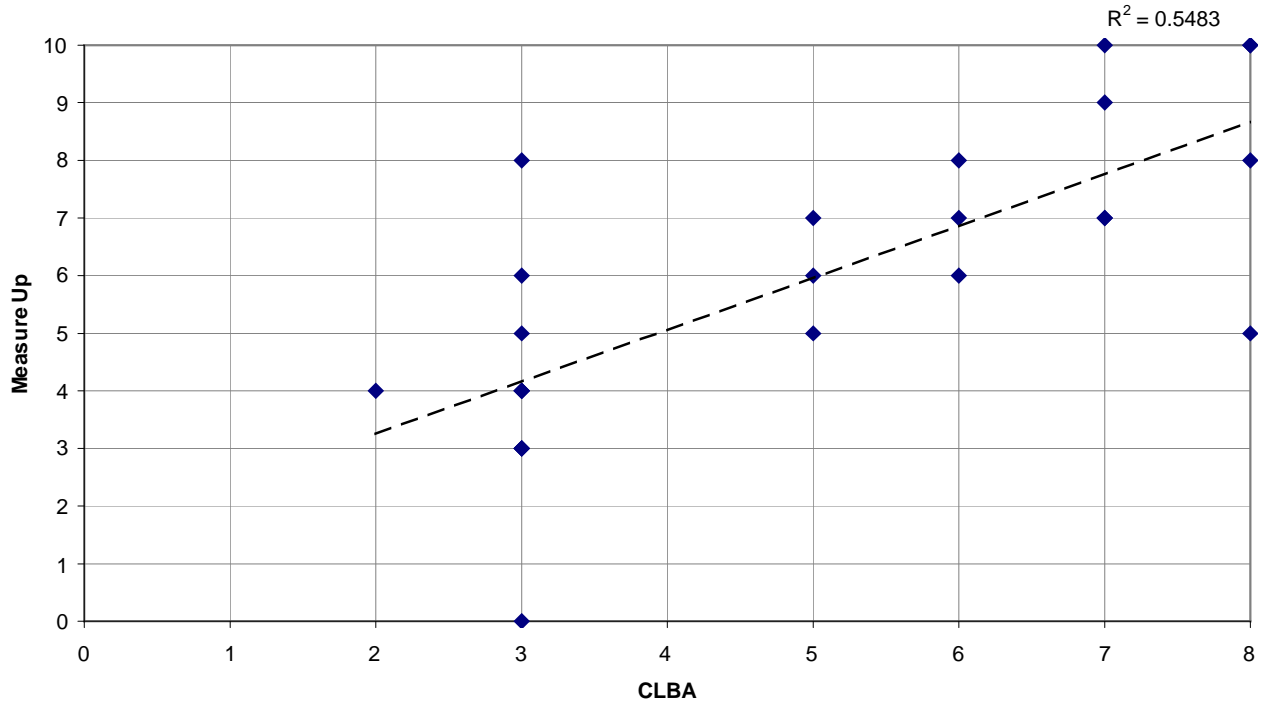
In Figure 3, we compare the CLBA Reading benchmark to the total on Measure Up. The R^2 value of 0.5483 indicates that 55% of what Measure Up tests is equal to 55% of what CLBA Reading tests.

Skills and
knowledge
required by
CLBA

Skills and
knowledge
required by
Measure Up



Comparison of CLBA to Measure Up (Figure 3)

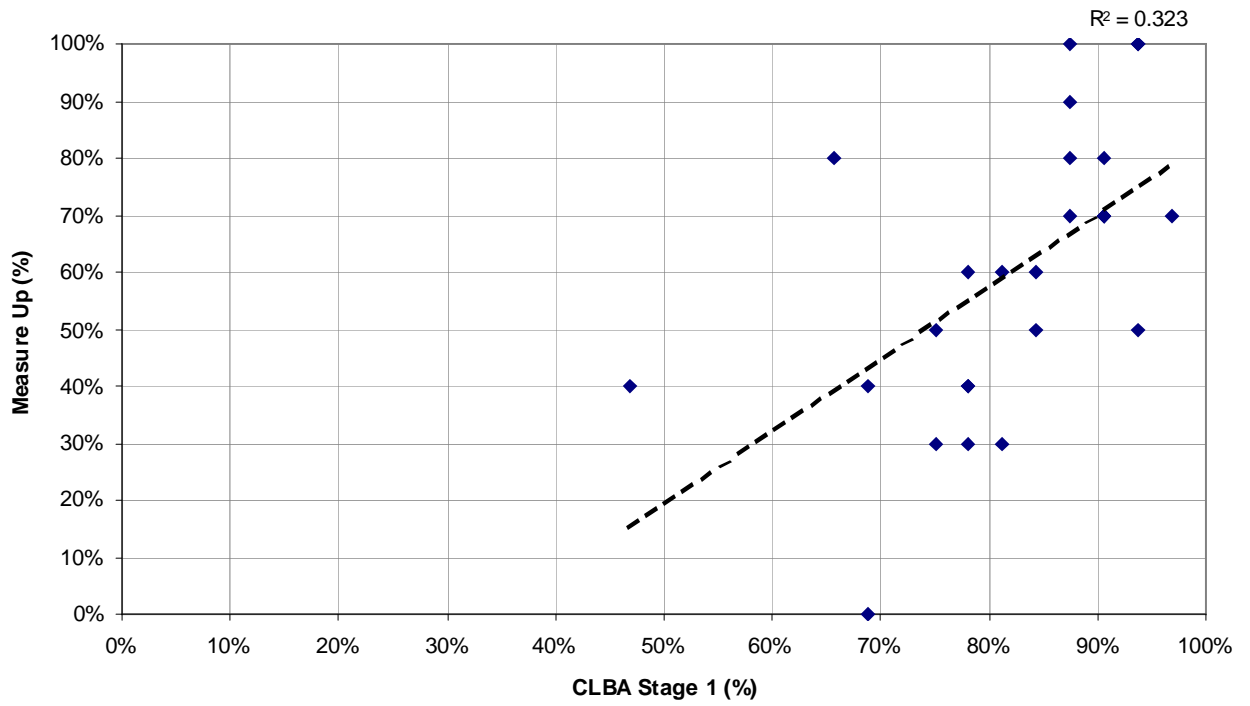


5. Comparison of Parts of CLBA reading (Stage 1 and Stage 2) to Measure Up Total

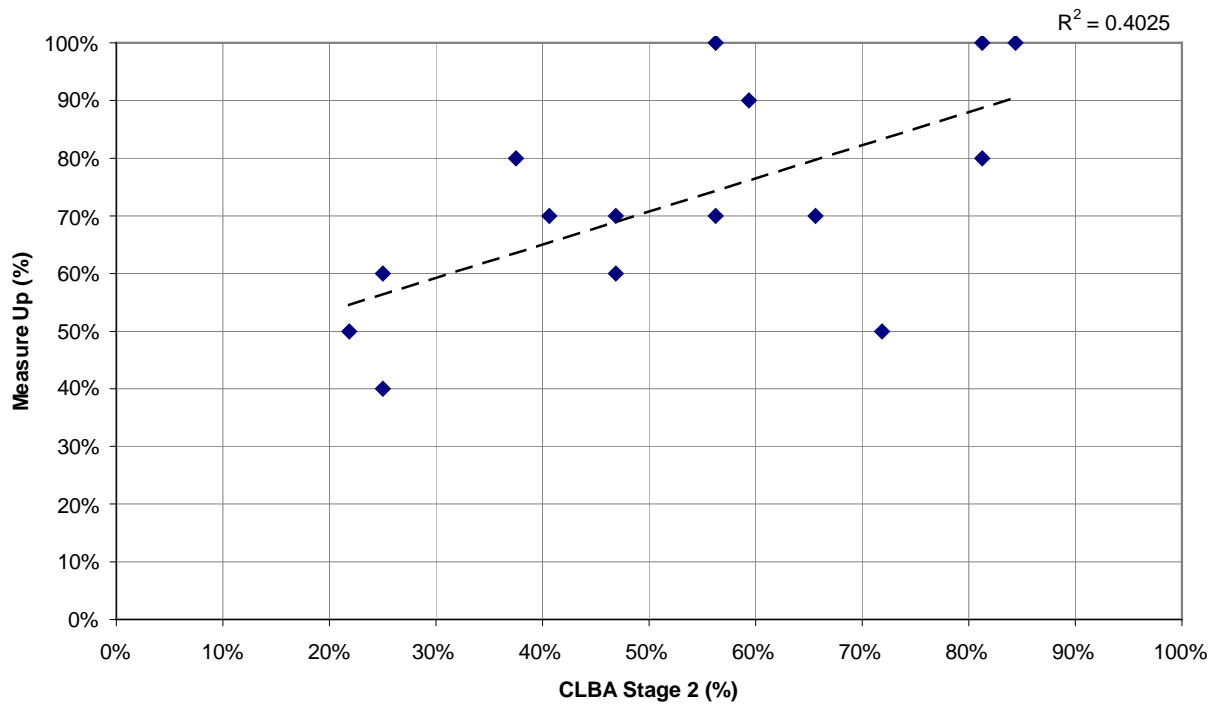
In order to see more of a spread within the CLBA, we also looked closely at participant performance on each Stage of the CLBA and compared their percentages (from raw scores) to percentages on the Measure Up. In Figure 4, you can see that most participants got very high scores on Stage 1. This is understandable, as it is used to qualify them for Stage 2. From the rising slope of the line, we can understand that participants who performed well on CLBA Stage 1 performed better on Measure Up.

We were able to use the results of those participants who completed Stage 2 to make a comparison with Stage 2 percentage and Measure Up percentage in Figure 5.

Comparison of CLBA Stage 1 to Measure Up (Figure 4)



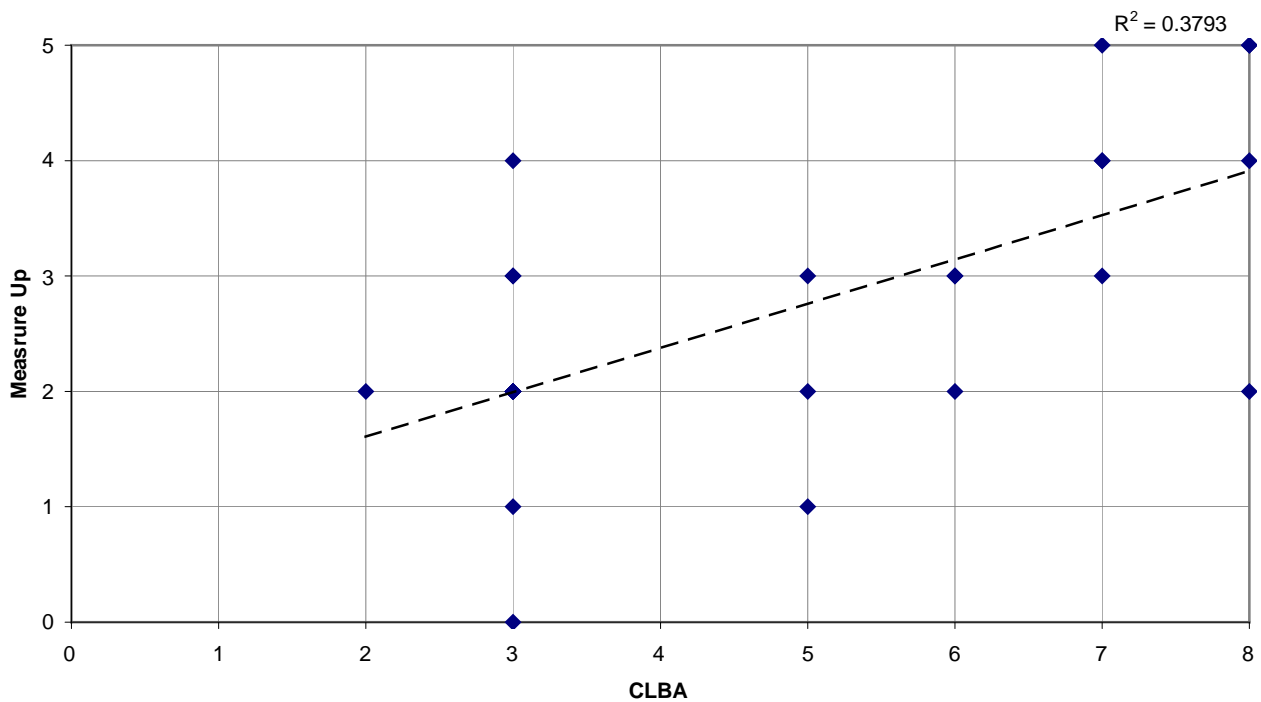
Comparison of CLBA Stage 2 to Measure Up (Figure 5)



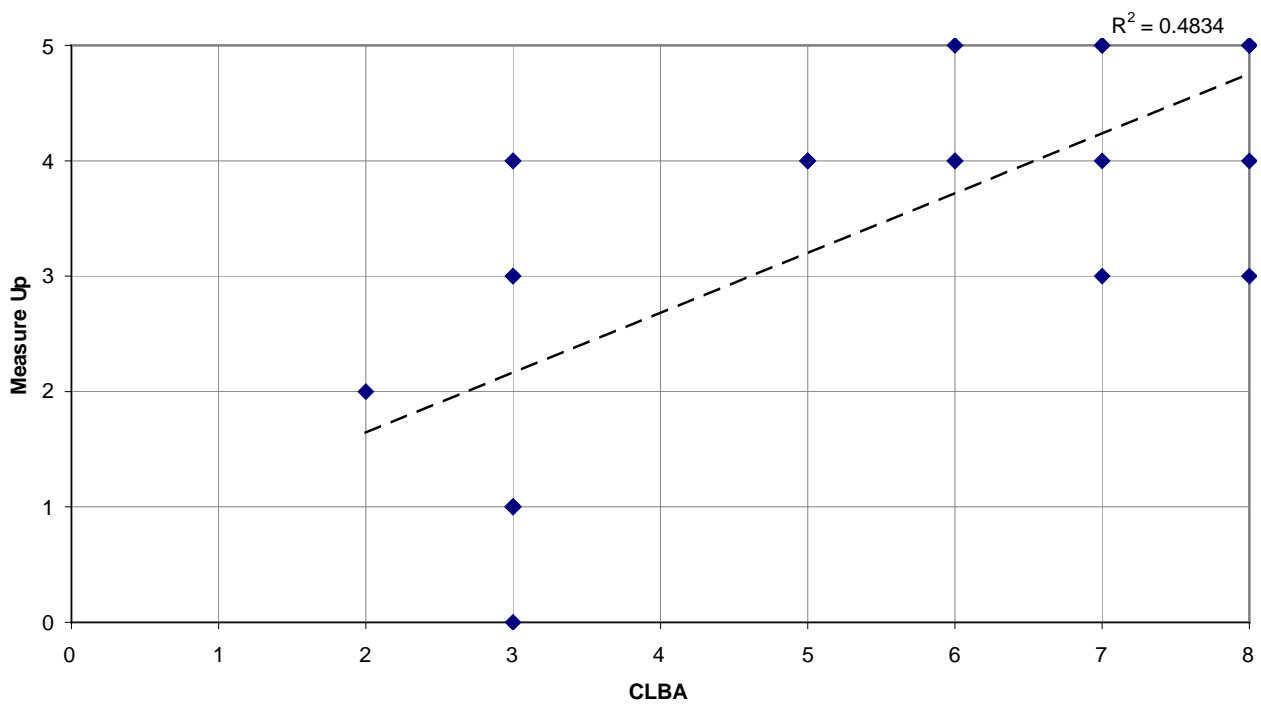
6. Comparison of Parts of Measure Up (Self Assessments 4 and 6) to CLBA Reading Benchmark

In Figures 6, we compare self assessment 4 with CLBA Reading Benchmark. You can see that the trend is for them to increase together. However, the correlation only has an R^2 value of 0.3793. Similarly for self assessment 6, in Figure 7, we find an R^2 value of 0.4834.

Comparison of CLBA to Measure Up Self-Assessment 4



Comparison of CLBA to Measure Up Self-Assessment 6



Although there is a relationship between these two tests, it is general. We can say that poor readers do poorly on both tests and good readers do well on both tests. The relationship is not strong enough to predict the benchmark of a person on the CLBA based on their score on the Measure Up test nor the reverse.

7. Recommendations

If the amount of time required to complete the test is considered an important part of the research, the CLBA could also be timed (keeping the maximum time allowed by the testing procedure). Many participants did complete the test in less than the maximum time allowed.

The data from Field Tests 1 and 2 could be analyzed together to see a larger picture—more participants and a greater range of benchmarks.

If it is important to gather participants' comments about the Measure Up test, oral interview or marking a scale would be more effective than having them write comments for low-level learners. This is especially important considering that their writing skills are usually behind their reading skills.

8. Attachments

You will find all the data in the attached excel file entitled “field test 2”.