

Comparison of Canadian Language Benchmarks Assessment and Measure Up Tasks- Field Test 1

Prepared by Christa Bruns, certified CLBA Assessor
Southern Alberta Institute of Technology

March 31, 2004

Comparison of Canadian Language Benchmarks Assessment and Measure Up Tasks- Field Test 1

1. Introduction

The purpose of this study is to compare two assessments: The Canadian Language Benchmarks Assessment and the Measure Up Tasks. A field test was completed at SAIT in which 27 participants were tested with both the selected *Measure Up* tasks from levels 1 and 2 and CLBA. Throughout the report the following abbreviations are used: CLBS- Canadian Language Benchmarks Speaking Assessment, CLBR- Canadian Language Benchmarks Stage 2 Reading Assessment, and CLBW- Canadian Language Benchmarks Stage 2 Writing Assessment.

Information about the CLBR, the results of the field test, participant comments about the *Measure Up* tasks and recommendations for further field tests make up this report.

2. Description of CLBA Reading

a. Purpose

The CLBA Reading Stage 2 assessment is used to place newcomers in an appropriate level class for English as a Second Language instruction, as an entrance requirement to some post-secondary institutions and work experience programs; and to measure outcomes of language learning.

b. Content

The range of task types in stage 2 CLBA Reading include: Reading complex instructional texts, reading complex formatted texts, reading complex unformatted texts and reading complex informational texts.

c. Structure

Tasks A,C, and D are multiple choice type items, with a choice between 4 answers. Task B is short answer.

d. Scoring Issues around CLBA stage 2 Reading

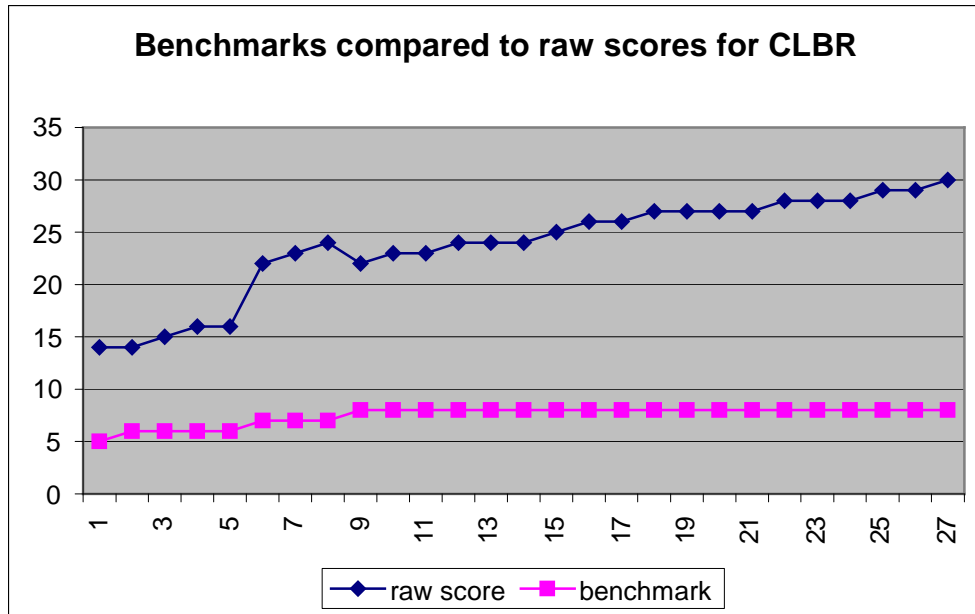
Those using CLBA Reading test this tool for research need to be aware of the characteristics of the tool.

Composite scores The benchmark score are composite. For each of the four tasks, the raw score is reduced to a score out of three. These four performance indicators are then added together and again reduced to a single benchmark from 4 to 8.

Where are the fours? The lowest possible benchmark on the Stage 2 test is Benchmark 4. In reality, it is highly unlikely that a student will get a benchmark four, because that would relate to a score of 0 to 9 out of a possible 32.

Why so many eights? Depending on how the tasks scores are distributed, a client could have as few as 21/32 and get a benchmark 8. Also, one could have as many as 27/32 and get a benchmark 7.

It may be more interesting to compare the raw scores out of 32 on the CLBA Reading to the Measure Up tasks. This way we can differentiate between the 8's. In the graph below, you can see that there is indeed a spread in the raw scores among



benchmark 8 tests. Note also that the benchmark 7 tests had raw scores equal to some of the benchmark 8 tests. The range of benchmark 7 raw scores was 22-24. The range of benchmark 8 raw scores was 22-30.

The data relating CLBA Reading raw scores and benchmarks is shown in the table below.

participant	raw score	Clbr benchmark
17	14	5
28	14	6
15	15	6
7	16	6
11	16	6
19	22	7
16	23	7
3	24	7
9	22	8
1	23	8
18	23	8
2	24	8
8	24	8
25	24	8
4	25	8
20	26	8
22	26	8
6	27	8
10	27	8
13	27	8
26	27	8
5	28	8
12	28	8
21	28	8
14	29	8
23	29	8
27	30	8

The range of CLB The Canadian Language Benchmarks range from 1 to 12, while the CLBA only tests up to benchmark 8. This means that a score of 8 indicates an 8 or higher. Considering that many of the participants learn largely through reading and have experience reading complex and lengthy texts; many of them probably do have reading benchmarks higher than 8. 18/27 participants had a benchmark 8 in reading. This shows the ceiling effect of the test.

3. Narrative of activities

Recruitment- For this study, students from the SAIT English Language Foundations Program (levels 2,3 4 and recent graduates) and Chinook College (Academic Preparation level 2) course were invited to participate. Twenty seven applicants were accepted to participate on a first come, first serve basis. During the CLBA speaking assessment, each participant reported his/her years of education. The range was from 12

to 17 years. The average was 14.9 years. Most participants were professionals in their fields before coming to Canada.

Testing Procedure The participants completed the *Measure Up* tasks in a group setting. The time required for participants to complete the tasks ranged from 13 to 28 minutes. The participants were informed that the time would be recorded.

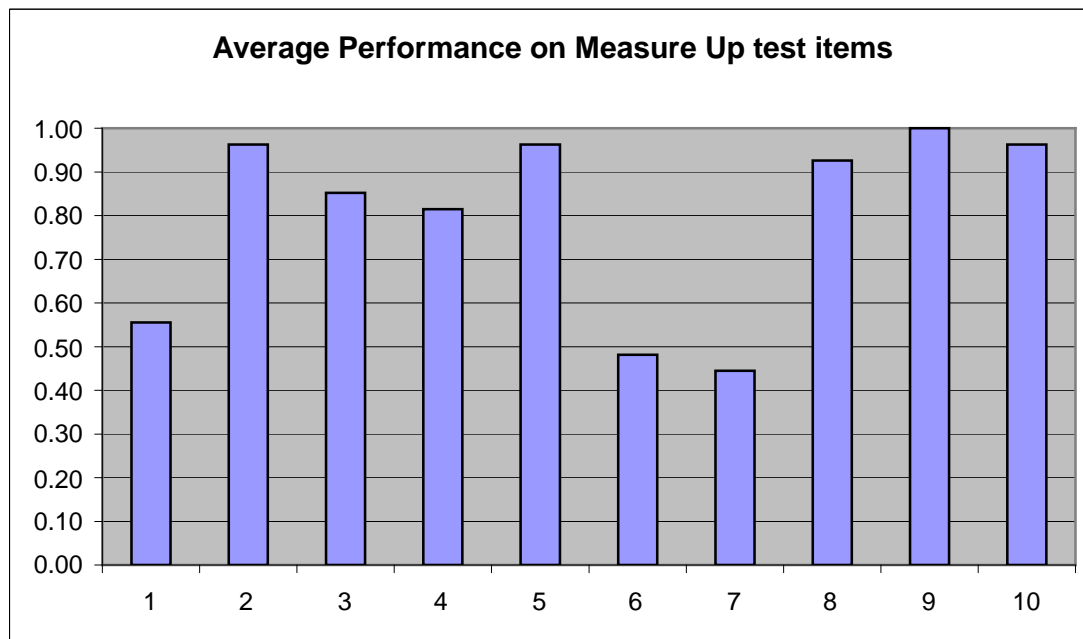
CLBA Speaking and Listening interviews were conducted individually according to a set schedule. Form 1 was used for all participants.

CLBA Reading and Writing Tests were completed in a supervised invigilation centre at the SAIT campus. Student were able to choose the day and time of these tests within limits. Various forms of the tests were used.

The participants had done the CLBA previous to this study. They were not familiar with the *Measure Up* tasks.

Scoring issues around *Measure Up* Tasks While scoring a short answer test, it is important to be consistent and decisive about what makes up an acceptable response. Therefore the same criteria applied to CLBA reading short answer test items were used for *Measure Up* tasks. Namely, incomplete answers were marked incorrect and answers including irrelevant information were also marked incorrect. For example, when the correct answer was “30 meters in front of and behind the bus”, an answer of “30 meters” was marked incorrect. As long as spelling and grammar errors did not leave doubt as to the participant’s understanding of the text and question, these errors were not marked incorrect. The *Measure Up* tasks were designed to be self-assessments. As such, the test-takers may be inclined to score their own responses more leniently.

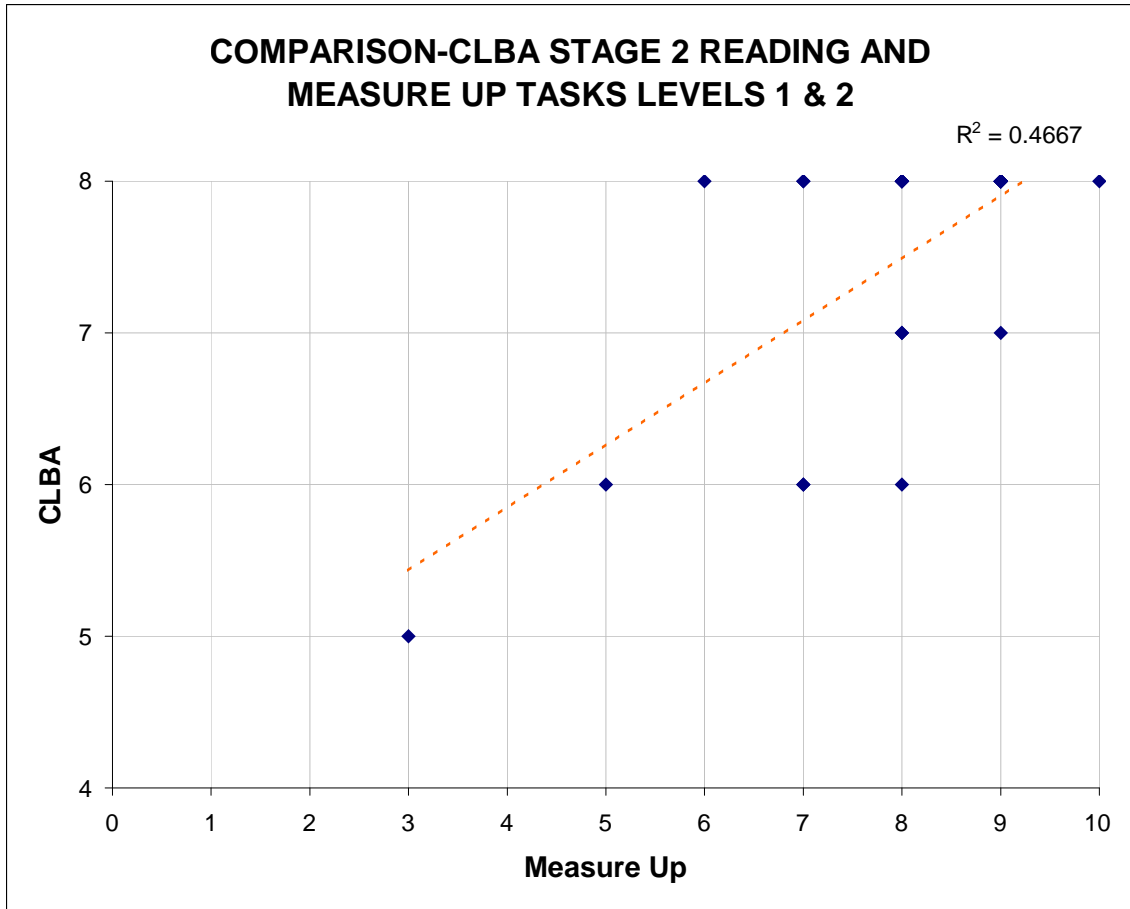
Some items on the *Measure Up* Tasks were easier than others as shown by the graph below. 1 means that all the participants answered correctly.



4. Comparison of Test Results- Record of Data

Measure Up itemized results and CLBA Reading results

participant	Measure Up Self Assessment 4					Measure Up Self Assessment 6					total/10	CLBR benchmark /8	CLBR raw score /32
	1	2	3	4	5	1	2	3	4	5			
1.	1	1	1	1	1	0	1	1	1	1	9	8	19
2.	1	1	1	1	1	1	0	1	1	1	9	8	19
3.	1	1	1	0	1	0	1	1	1	1	8	7	17
4.	0	1	1	1	0	1	1	1	1	1	8	8	18
5.	1	1	1	1	1	1	0	1	1	1	9	8	19
6.	1	1	1	1	1	0	1	1	1	1	9	8	19
7.	0	0	1	0	1	0	0	1	1	1	5	6	13
8.	1	1	1	0	1	1	0	1	1	1	8	8	18
9.	0	1	0	0	1	0	1	1	1	1	6	8	16
10.	1	1	1	1	1	1	1	1	1	1	10	8	20
11.	1	1	1	1	1	0	0	1	1	1	8	6	16
12.	1	1	0	1	1	1	1	1	1	1	9	8	19
13.	0	1	1	1	1	0	0	1	1	1	7	8	17
14.	0	1	1	1	1	1	1	1	1	1	9	8	19
15.	1	1	1	1	1	0	0	0	1	1	7	6	15
16.	1	1	1	1	1	0	1	1	1	1	9	7	18
17.	0	1	0	0	1	0	0	0	1	0	3	5	9
18.	0	1	1	1	1	1	0	1	1	1	8	8	18
19.	1	1	1	1	1	0	0	1	1	1	8	7	17
20.	0	1	1	1	1	0	1	1	1	1	8	8	18
21.	1	1	1	1	1	1	0	1	1	1	9	8	19
22.	1	1	1	1	1	1	0	1	1	1	9	8	19
23.	1	1	1	1	1	1	0	1	1	1	9	8	19
25.	0	1	1	1	1	0	0	1	1	1	7	8	17
26.	0	1	1	1	1	1	1	1	1	1	9	8	19
27.	0	1	0	1	1	1	1	1	1	1	8	8	18
28.	0	1	1	1	1	0	0	1	1	1	7	6	15
average	0.56	0.96	0.85	0.81	0.96	0.48	0.44	0.93	1.00	0.96	7.963	7.481	17.41



Basically interpreted, the r value of .68 reflects a substantial correlation between the two tests.

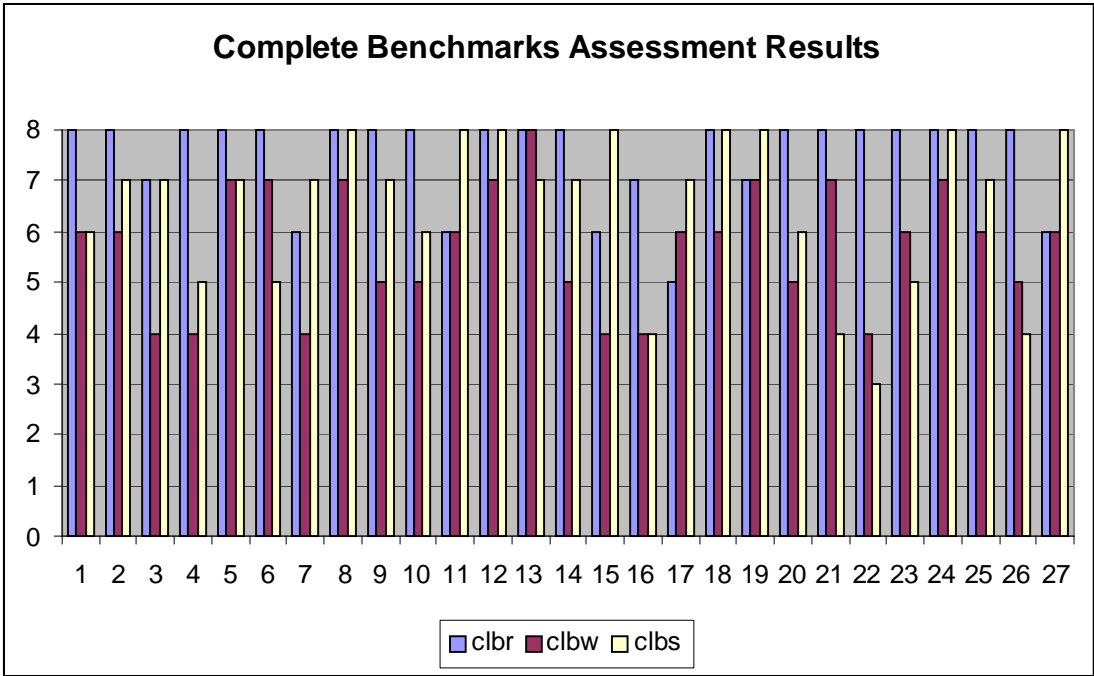
The R^2 value indicates that only 46% of the variance in Measure Up achievement is predictable from the variance in CLBA Reading achievement. Therefore, 64% of the variance is attributable to other factors that may include differences in test content, structure and scoring methods.

The participants were assessed not only on their reading skills, but on their listening/speaking and writing as well. The following table and graph represent their full benchmarks.

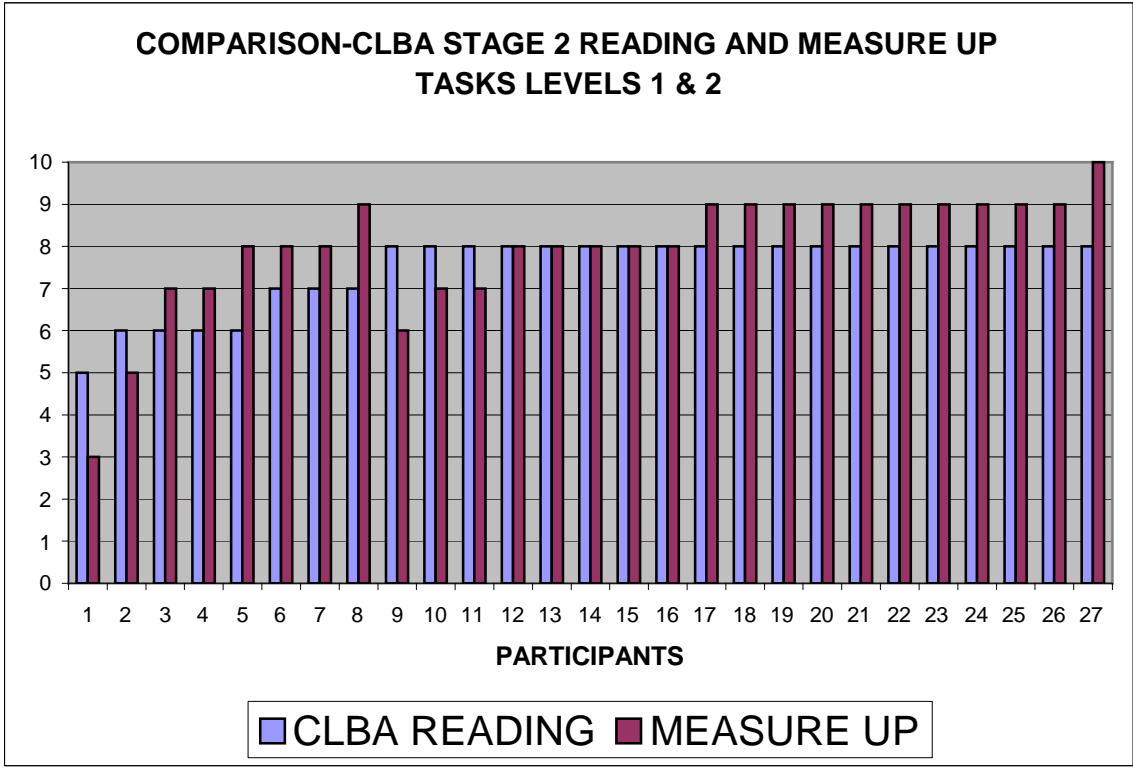
CLBA Benchmarks

participant clbr clbw clbs

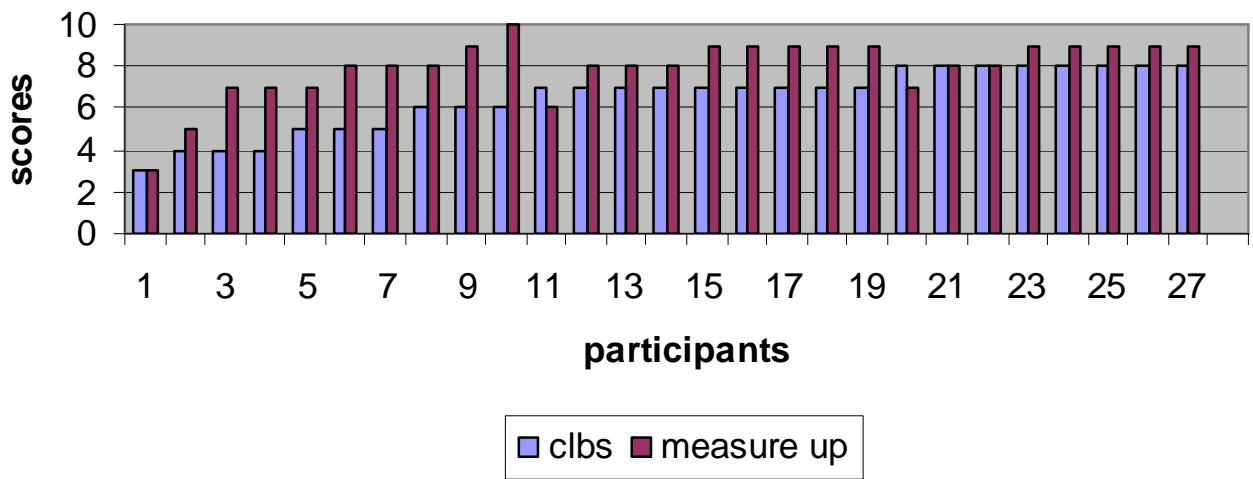
1	8	6	6
2	8	6	7
3	7	4	7
4	8	4	5
5	8	7	7
6	8	7	5
7	6	4	7
8	8	7	8
9	8	5	7
10	8	5	6
11	6	6	8
12	8	7	8
13	8	8	7
14	8	5	7
15	6	4	8
16	7	4	4
17	5	6	7
18	8	6	8
19	7	7	8
20	8	5	6
21	8	7	4
22	8	4	3
23	8	6	5
25	8	7	8
26	8	6	7
27	8	5	4
28	6	6	8



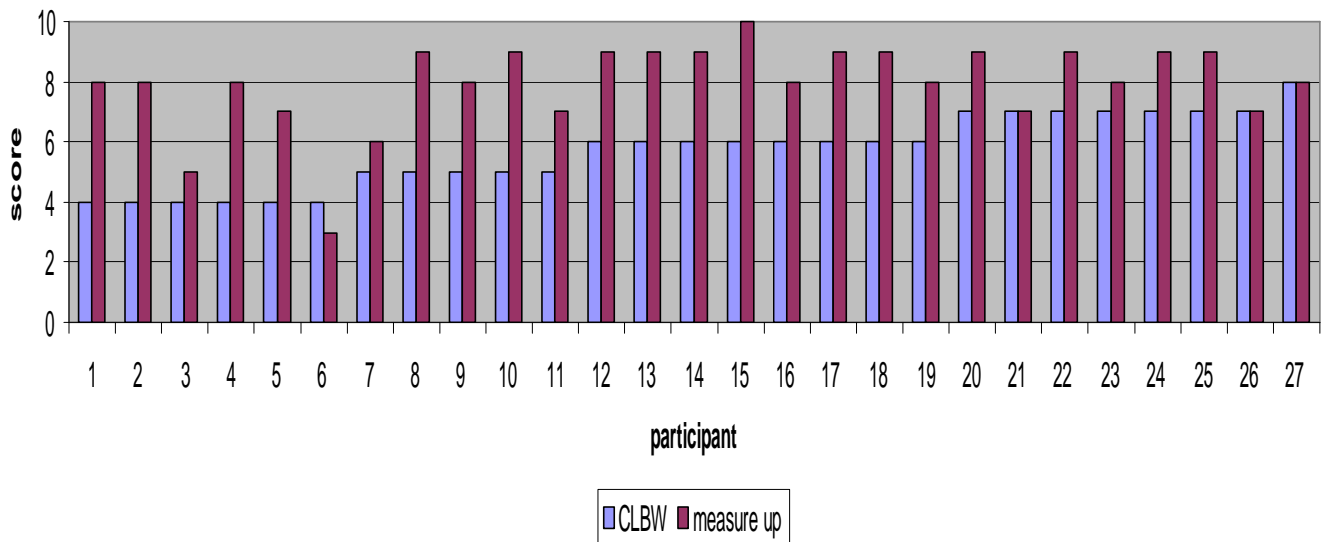
The following graphs show the results of each part of the CLBA compared to the Measure Up tasks.



Comparison CLBA Stage 2 Speaking and Measure Up Tasks Levels 1 & 2



Comparison CLBA Stage 2 Writing and Measure Up Tasks Levels 1 & 2



5. Participant Comments

During the listening and speaking interview, the participants were invited to make comments about the *Measure Up* tasks. 14 of them said that it was easy. 5 said it was not difficult. 3 said it was okay or good. 1 said it was interesting. 1 said it was confusing. 3 participants did not make comments.

Elaborating on the ease of the test, students remarked that the tasks were “straight forward” and “not abstract”, used “easy vocabulary” and that “by looking at the context I could guess”. Two students labeled the tasks as scanning, which they had studied in class: “all the information was in there. I don’t need to think.” One student remarked that the tasks were “appropriate for benchmark 5 or 6.”

About the content of the readings, participants tended to appreciate the familiar topics. They said the readings were “realistic” and “related to our daily life- a good topic”.

Two participants pointed out the benefits of short answer as opposed to multiple choice tests. One said, “I liked it because I don’t have to change my mind”. Another said, it was “more useful to measure how people understand”. Meanwhile another participant was troubled by the short answer format because as he said, “I have no idea how to finish it because I can’t paraphrase the answer.”

Some participants were quite concerned about the amount of time they spent. They made remarks such as: “I think it is used to measure the time.” “I tried to do it faster because I was timing.” “I should have taken less time.” And “It took me 4 minutes to get oriented.”

Only two students had complaints about the test or thought they did poorly. One participant admitted that she found the vocabulary a confusing and “how to apply the words to the context was a little difficult.” She got 5/10 correct. The other participant complained that the test was not well organized because the questions were not numbered.

6. Recommendations

In order to further explore the relationship between the CLBA and the *Measure Up* tasks, the following steps could be taken:

- Select participants for a broader range of initial benchmarks
- Use a broader range of *Measure Up* tasks, including more difficult tasks.
- Focus on CLBA Reading only as opposed to the full CLBA, which is very time intensive.