www.itsessential.ca

Canadian Language Benchmarks/ Essential Skills | **Niveaux de compétence linguistique canadiens/ Compétences essentielles**

Canadian Language Benchmarks and Essential Skills Comparative Framework: Report on Responses From the Field

July 2005

www.language.ca

CLB-ES Comparative Framework

Report on Responses from the Field

July 21, 2005

Philip Nagy
Gail Stewart
Karen Geraci

Executive Summary

This report describes a small study that was conducted to elicit responses from English as a Second Language (ESL) professionals on the accuracy and utility of a Comparative Framework (CF) designed to align the Canadian Language Benchmarks (CLB) with Essential Skills (ES). The study was based on a three-pronged approach that included user feedback on the value of the framework as a tool to inform classroom instruction, a panel session in which respondents were asked to place selected tasks using the framework, and respondent opinion on an introduction that had been written to describe the framework.

Analysis of data for the first prong of the study revealed that, in general, response from the field was positive and welcoming in regard to the overall concept of devising a framework to depict the relationship between the CLB and ES. Respondents affirmed the inherent value of exploring and establishing a clear linkage between these two important national standards, and they seemed to appreciate the convenience of having a concrete tool to assist them in understanding this complex relationship. In the course of their participation in the study, our respondent group provided a great deal of anecdotal feedback to suggest ways in which the framework might be clarified, fleshed out, and improved. Had this single prong comprised the full extent of our study, we might well have reached the simple conclusion that the Comparative Framework should undergo minor to moderate revisions and then be made fully operational. However, when we moved on to the analysis of data for the second prong, some significant issues arose.

In the second prong of the study, panels of experts were assembled in order to demonstrate the degree to which users of the framework would be able to apply its principles with consistency and confidence. In summary, when panels attempted to use the Comparative Framework to place selected Reading and Document Use tasks, the degree of disagreement at the higher levels was significant enough to cast some doubt on the basic utility of the framework and to raise concerns regarding the potential for high-stakes misuse. As a result of these findings, our ultimate recommendation is that the present version of the Comparative Framework be re-conceptualized.

1

Background and Methodology

In 2002, a project was initiated to investigate the relationship between the Canadian Language Benchmarks (CLB) and the Essential Skills (ES). Out of this investigation, the BC Construction Industry Skills Improvement Council (SkillPlan), in consultation with the Centre for Canadian Language Benchmarks (CCLB), produced a Comparative Framework (CF) that represents a draft attempt to align the skills and competencies reflected in these two national standards. The alignments presented in the framework were based on the holistic judgements of a research team who brought a background in both CLB and ES to the task. The framework was subsequently reviewed by a National Advisory Committee and was then introduced to a representative sample of potential users to elicit feedback on its content, format, and utility. This report describes the methodology and results of a study that was conducted with these potential users.

The sample for the study included a range of participants with background in one or more of the following: English as a Second Language; Canadian Language Benchmarks; Essential Skills; and workforce preparation.

The study was designed as follows.

1. The CF was sent out to a group of ESL professionals working in CLB-based language programs across the country. By means of a survey, responses were collected from this group to determine their satisfaction with the content and utility of the framework. Suggestions for changes and improvements to the framework were solicited from the group, and this feedback was analyzed, compiled, and reported in the summary and recommendations at the end of this report.

2. A panel of users was assembled in four focus groups (N = 36) across the country. These groups were asked to provide feedback on the framework and to judge the CLB levels for selected Reading and Document Use tasks which had pre-determined ES levels. These judgment data, collected both individually and after group discussion, were used to estimate the degree to which the CLB levels assigned by respondents aligned with the ES levels as described in the current CF.

3. The Introduction to the Comparative Framework was sent to participants, who were asked to comment on its comprehensiveness, clarity, usefulness, and accuracy. They were also asked to estimate their level of comfort and confidence in applying the information in the introduction to their use of the Comparative Framework. Finally, they were encouraged to suggest revisions that would improve the introduction.

This report is structured around the results of each of these three prongs of the investigation. Note that, as part of Step 2, participants filled out a demographic and background survey. While most of these responses are reported in Step 2 below, a couple

of questions concerning plans and expectations for the trial period are included in the results of Step 1, following immediately.

Feedback on the Framework

Plans and Expectations from Demographic Survey

Respondents were asked, in the initial demographic survey, which programs they would be representing during the pilot time span. A list, organized under several categories, is included in Appendix A. In this same survey, respondents were also asked how they expected their participation in the study to benefit them and their teaching. A large percentage of respondents perceived their participation in the pilot and focus groups as an opportunity to learn more about Essential Skills, assessment, and the theory and application of the two scales (ES and CLB). Many respondents indicated that their participation in the pilot process would enhance their ability to provide more targeted teaching material for learners with employment goals. Slightly fewer respondents saw their participation as a way to learn more about the Canadian Language Benchmarks, materials selection and research generally. Responses are summarized in Table 1.

Table 1: How Participation in This Project will Benefit Respondents and Their Teaching

Benefit	N Expected
Benefit Respondent	
Learn more about Essential Skills	25
Learn more about the CLB	19
Learn how to select materials	19
Learn more about assessments	24
Knowledge of theory and application of the two standards	24
Develop more interest in research	18
Other	7
Benefit Teaching	
Provide more targeted material	23
Facilitate more employment-related content	20
Provide more relevant materials for learners with employment goals	22
Other	6

As can be seen, respondents had high hopes for the project, with a heavy emphasis on learning more about Essential Skills and other approaches to enhance programming for employment-oriented ESL learners. The "Other" expected benefits are listed in Appendix A.

The Usage Sample

Over a period of seven to eleven weeks, 45 participants worked with the framework, and at the conclusion of their trial period, they responded to a questionnaire on its usefulness, clarity, and comprehensiveness. The majority of the respondents came from Ontario (N = 20) and Alberta (N = 11), with five from each of BC and Manitoba. There were two from Newfoundland, and one each from Nova Scotia and New Brunswick. They were employed by school boards (N = 9), community colleges (N = 11), non-profit organizations (N = 9), immigrant-serving agencies (N = 7) and private schools (N = 3). Six described themselves as working in other related roles. The majority described themselves as either LINC (N = 12) or ESL (N = 14) teachers, or as ESL administrators (N = 6). Eight were self-described as CLB experts, two as ES experts, and three as experts in both.

The Framework and Essential Skills

The core of the survey consisted of six statements with associated prompts for additional suggestions (see Appendix B). Results are summarized in Table 2.

Table 2: Summary of Responses to Usage Survey

Statement	1	2	3	4
1. The framework gives me a better understanding of how Essential Skills relate to the benchmarks	30	12	1	2
2. I feel confident in using the Framework to choose and gauge materials to be used in my ESL classroom	19	20	4	2
3. I feel confident using the Framework to place Canadian Language Benchmarks tasks along the Essential Skills scale	19	20	4	2
4. I found the Framework accessible and easy to use	17	19	7	2
5. I found the information in the Framework to be complete and comprehensive	17	21	5	2
6. I found the information in the Framework to be accurate and representative of the underlying scales (the CLB and the Essential Skills)	23	20	1	1

(1) = Agree Completely, (2) = Agree Somewhat, (3) = Disagree Somewhat, (4) = Disagree Completely

After each question, respondents were asked for specific suggestions, as detailed in Appendix B.

In response to Statement 1, it is clear that most users felt that the CF gave them a better grasp of the relationship between the CLB and ES. Where suggestions for improvement were offered, these centred on the need for more examples, practice, discussion, and detail. In addition, there were a few editorial and organizational suggestions. Detailed comments,

slightly edited, are listed in Appendix B. Comments that were simply supportive without offering any further suggestions have been omitted.

In response to Statement 2, the most common comment was a request for further examples. Also relevant to this statement are concerns about how low-level ESL students (beginners, LINC 1 and 2) might be introduced to ES materials, given that these levels are not addressed in the Comparative Framework. The fact that these levels are absent from the CF seems to imply that it is perhaps unsuitable to introduce ES tasks at the lower benchmarks. However, instructors recognize that this is not necessarily the case; certain workplace tasks can be accessed, in whole or in part, by learners at the lowest benchmarks.

For Statement 3, the most common comment was that further practice would be required. One respondent felt the framework was confusing and inefficient, while another suggested that efforts to apply the framework were too time consuming. When asked for comments on Statement 4, there were about ten specific and pointed suggestions relating to layout. Several respondents also requested more examples.

With respect to Statement 5, many respondents commented on the number of examples. There were a couple of comments about layout, and some questions concerning the completeness of the document. Also relevant to this statement are comments that acknowledge other factors which contribute to an individual's ability to perform on ES tasks. These include previous work experience, literacy skills, formal first-language education, cognitive ability, and background knowledge. These considerations do not appear in the framework, yet respondents recognize their significance, which seems to underscore an inevitable limitation of the CF and other documents of its kind – the fact that it may not be possible to capture on paper all of the reasons why task performance might vary from one individual to the next.

In responding to Statement 6, some of the participants felt that they did not have sufficient background in Essential Skills to understand how the complexity of an ES task is determined. Seven respondents said that this insufficiency left them unqualified to comment on the accuracy of the framework. In relation to this statement, there were also several suggestions for expanding the framework.

Respondents were given the opportunity to add additional comments, and many did so. Most of these comments emphasized the importance of the project. There were, as above, several suggestions for an improved format.

It is perhaps worth emphasizing that throughout the survey, on many of the questions, respondents called for more examples. These were viewed as important adjuncts that could serve a number of key purposes for users.

It is also important to note that several respondents indicated a need for further training or other supports in order to use the framework effectively. This is supported by the

comments received after the focus group, which showed a high level of satisfaction among participants who, in fact, viewed the focus group experience as a workshop or training session.

Commentary

The CF could be improved and enhanced by the inclusion of example materials that typify tasks, texts, and documents at all levels. Exemplars could be used to illustrate the types of documents and texts described by both the CLB and ES, thereby giving instructors a better understanding of how specific tasks might be of benefit for specific instructional purposes. In addition, exemplars could assist those users who find it difficult to fully comprehend the alignments shown in the framework and the key distinctions between levels. A series of sample tasks could clearly illustrate, for example, how complexity levels vary across ES tasks, even when those tasks are associated with a single document. The addition of exemplars could thereby assist in clarifying which criteria support task placement at different levels across the two scales. In these ways, examples would enhance and support a richer interpretation of the Comparative Framework, by exposing what is addressed by the framework and what, because of the different natures of the two scales, is beyond the scope of the CF.

In focus group conversations and in written feedback, respondents articulated a clear desire to participate in professional development on the relationship between ES and CLB. A majority of respondents were new to Essential Skills, and expressed a lack of confidence in how to introduce, within their classrooms, the types of tasks and methodologies which the ES framework seems to support. Focus group participants indicated that they felt more confident about ES and its potential connections to language learning after the focus groups themselves, which they perceived as training sessions. This increase in confidence can likely be attributed to opportunities within the focus groups to learn more about ES, to participate in hands-on placement activities, and to discuss the framework's potential applications and limitations with colleagues. Professional development which provides participants with an overview of both scales (ES and CLB), and affords them opportunities to see how the two scales complement each other, would support the Comparative Framework's consistent and considered application in the classroom.

<u>Results of Text and Task Placement Activities</u>

This section of the report describes the results of text and task placement exercises in Reading and Document Use. For the focus group agenda and copies of the assigned tasks and exercises, please refer to Appendix C.

There are ratings data from 36 respondents, 28 of whom provided responses to the demographic survey. The main data set consists of CLB benchmark ratings for tasks associated with 10 Reading texts and 12 Document Use texts. The tasks had pre-assigned

ES levels which respondents did not know. Respondents were asked to assign a benchmark range to each text and a single benchmark to each task. They did this individually and after group discussion. Thus, the data shell consists of 82 judgments (22 assigned benchmark ranges for the texts, and 60 assigned benchmarks for the tasks). Due to time constraints in the panel sessions, many respondents did not complete assignments on all texts and tasks. This accounts for the many blanks in this data shell.

Respondents used two rating sheets to record their benchmark assignments, one before discussion and another after. They were instructed to record a benchmark on the second (after discussion) sheet only in cases where there had been a change in decision from their initial (before discussion) rating. In the analysis, therefore, the assumption was made that if there was no entry on the second sheet, the benchmark assigned on the first sheet still stood.

There are many blanks on both the before and after discussion sheets. It is clear from the data sheets, and from discussions with those who ran the data collection sessions, that some participants did not have enough time to give due consideration to all texts and tasks.

In addition to benchmark ratings, respondents were also asked five questions about each of the 22 texts. Before discussion, they were asked:
1. How certain are you that you have placed the material associated with this text at the correct benchmarks?
2. How easy was it for you to place this material using the comparative framework?
3. How likely would you be to select this material for use in a language program?

After the discussion, they were asked
4. Following discussion with your group, how certain are you that you have now placed the material associated with this text at the correct benchmarks?
5. How easy was it for your group to reach consensus on the placement of this material?

Each of the five questions was accompanied by an appropriate four-point scale. The many blanks in the responses to these questions also reflected a lack of time. Despite these difficulties, in all cases there are enough respondents to each question that reliable information can be obtained from the data.

This report on this data analysis is organized as follows:
- Characteristics of respondents
- Extent of agreement on benchmark placement among respondents. For ease of comparison between the *individual* and *discussion* data, both these conditions are reported together.
- Reported confidence and ease of assigning ratings alone. As a side bar, the relationship between confidence and accuracy of ratings is explored.
- Reported confidence in and ease with assigning ratings after discussion.

Characteristics of Focus Group Respondents

The four focus groups were held in Edmonton (N = 9), Hamilton (N = 12), Toronto (N = 12) and Vancouver (N = 3). The bulk of participants came from Ontario (N = 14) and Alberta (N = 8), with two from each of B.C and Manitoba, and one from each of New Brunswick and Newfoundland.

A large variety of ESL-related occupations were reported by the group, as reported in Table D1. The majority of the groups worked for either a school board (N = 6), a community college (N = 8), or a non-profit organization (N = 8). In addition, there were two who worked for an immigrant serving agency, and one at each of a private school and an adult learning centre. One was self-described as a consultant, and another as self-employed.

Respondents were asked about the orientation of the program in which they work. Seven (of 28) reported Academic Preparation, 16 reported Workplace Preparation, and five said the question was not applicable. They were also asked what types of academic skills were covered in their program. For each of the options offered in the questionnaire, academic writing, academic reading, advanced speaking, study skills and test preparation, only three to six respondents checked *yes*. In addition, there were three write-in responses of listening, pronunciation, and high school preparation.

Participants were selected for their complementary backgrounds in ESL, CLB and workforce preparation. In particular, the instructors with some experience in workforce preparation were targeted in the belief that they were more likely to be familiar with the Essential Skills framework, and perhaps even using aspects of it in their classrooms. Researchers felt that a range of ESL programming experience amongst participants would form a representative snapshot of English as a Second Language programming across Canada.

Respondents were asked about the workplace skills their home programs focused on. Half of the 28 reported *resume preparation, job interview skills,* and *occupational terminology*, while slightly fewer (between 8 and 12 of the 28) reported *workplace reading, workplace writing, career direction* and *role-playing*. There were also a handful of other reports, as outlined in Table D2.

Respondents were asked about their use of the CLB. A substantial majority reported the three choices offered: to identify learning outcomes (N = 22), to develop materials (N = 20) and to check learner progress (N = 17). Experience with the CLB ranged from two years or less (N = 5) through 3-5 years (N = 10) to more than five years (N = 12).

The CLB levels with which the respondents had experience are summarized in Table 3.

Table 3a: Levels of Present Experience

Level	N with Experience
CLB 1 – CLB3	4
CLB 4 – CLB 5	2
CLB 5 – CLB 6	3
CLB 7 – CLB 8	10
CLB 9 – 10	2
N/A	6

Table 3b: Levels of Past CLB Experience of Respondents

Level	N with Past Experience
ESL Literacy	13
CLB 1-3	23
CLB 4-5	25
CLB 5-7	23
CLB 7-10	13
CLB 11-12	2

Table 3a, depicting current experience, seems to indicate that the participant sample is not a perfect reflection of the teaching population. However, when past experience (Table 3b) is taken into account, the respondents appear to be a typical representation of the distribution of experience across the population. Given the immigrant population found in most ESL programs, one would expect that the teaching population would distribute in a manner that is more consistent with Table 3b, that is with the greater number of classes and instructors falling at the CLB mid-range, and fewer at the high and low extremes.

Table 4: Use of CLB Tools

Tool	N With Experience
CLBA	10
CLBPT	5
LINC	16
On Target	16
Additional Sample Tasks	13
Theoretical Framework	16
Guide to Implementation	14
CLB 2000: ESL for Literacy	11
Learners	14
Other resources	11

Those who checked "other resources" generally did not supply details, except for three reports of CLB2000 – ESL for Adults. Respondents generally have extensive experience with a large variety of CLB tools.

The level of Essential Skills experience reported by the respondents is summarized in Table 5. A distinction was made between general familiarity and actual experience. In general, respondents displayed reasonable prior familiarity but little experience actually using Essential Skills.

Table 5a: Aspects of Essential Skills where Respondents Have Some Familiarity

Aspect	N Familiar
Familiar with ES Reading	14
Familiar with ES Writing	15
Familiar with ES Oral Communication	15
Familiar with ES Document Use	11
Familiar with ES Thinking Skills	11
Familiar with ES Computer Use	8
Familiar with ES Team Work	11
Familiar with Measure Up	7

Table 5b: Aspects of Essential Skills where Respondents Have Experience

Aspect	N Experienced
Heard of ES in general	16
Was familiar with Measure Up	7
Have used TOWES	5
Have read something about Essential Skills	18
Use Essential Skills methodology in my teaching	7
Have visited the HRSDC Essential Skills web site	15
Have no familiarity with any of the above	8

Respondents were asked how often they use Essential Skills in their teaching of CLB-based ESL and which ES tools they have used in the past. The level of experience was low. Six of 28 responded that they used ES every class, and another one that s/he used ES once a month. No one else reported any actual ES experience.

Use of specific ES tools, presumably outside the classroom, is reported in Table 6. Less than half of respondents indicated they had used some tool associated with Essential Skills. Twelve of 28 indicated they had used sample materials; 11 had used Essential Skills Profiles, and 10 had used a document entitled 'Understanding Essential Skills'. A small minority had used other tools listed: four respondents had used TOWES; three said that they had used a document which they called the 'Essential Skills Resource Guide'.

Table 6: Essential Skills Tools Used in the Past

Tool	N Used
TOWES	4
Essential Skills Profiles	11
'Understanding Essential Skills'	10
Sample Materials	12
'Resource Guide'	3

Overall, respondents indicated low levels of experience and familiarity with the Essential Skills framework and tools. Less than half of respondents had any familiarity with Essential Skills, and even fewer had accessed any of the tools produced from the Essential Skills framework or research. Fewer still indicated that they had used the Essential Skills as part of their programming. Given the relatively recent appearance of workforce preparation within Canadian Language Benchmarks programming, this is very likely to be true for the English as a Second Language field generally.

Focus Group Methodology

At the beginning of each focus group, participants were asked to consider the CLB range that could be associated with each of the Reading or Document Use texts (passages) without looking at the tasks (questions) associated with these texts or documents. In completing this initial exercise, they did not use the Comparative Framework. Instead, they were asked to approach each text just as they would approach a piece of realia that they might be considering for use in an ESL classroom. Typically, when approaching a new text or passage, an ESL instructor reads through the piece and considers the range of ESL levels that might benefit from classroom exposure to the format and content. He or she then thinks in terms of the questions or tasks that would be appropriate for the target level of instruction.

Our introductory activity was designed to mirror the classroom preparation process and to provide a familiar context for ESL practitioners so that they would begin the panel session with the big picture, a sense of the overall suitability of each text or passage for ESL instructional purposes. In this manner, a foundation was laid for more targeted focus on the complexity level of each particular task. From a research perspective, the results of this initial text-rating activity provide some insights into the collective mind of the panel group. They help us to see to what degree the participants seem to be on the same page with regard to general application of CLB principals.

Following the text rating activity, participants were asked to assign a CLB level to each of the associated Reading and Document Use tasks. The two tables below show the relationship that is set out in the Comparative Framework. In each table, the left column indicates the ES level assigned to a workplace task, and the right column shows the CLB benchmarks which are associated in the CF with a task at this level of complexity.

ES Reading Text Level	Corresponding CLB Reading Level(s)
1	3, 4, 5
2	6
3	7, 8,9
4	10
5	11, 12

ES Document Use Level	Corresponding CLB Reading Level(s)
1	3, 4, 5
2	5,6
3	7, 8
4	9, 10
5	11, 12

The above relationships are drawn directly from the Comparative Framework and are based on the collective expertise of a research group who collaborated in mapping CLB levels onto the ES framework. In our study, we asked respondents to assign CLB levels to tasks with pre-determined ES ratings so that we could see whether these CLB ratings would line up with the ES correspondences as shown in the above tables. In this manner, we hoped to examine the consistency of the CF. Our assumption was that the accuracy and usefulness of the above correspondences would be reflected in the degree to which our panellists assigned a benchmark that aligned with the known ES level for each task.

In this study, we knew the ES level for each task because this had been established by a qualified ES rater. We did not reveal these ES levels to our panellists. Instead, we asked the panellists to assign a CLB benchmark to each task. By doing so, we hoped to determine whether or not our experts would assign benchmarks that fell within the ranges established by the research group who originally created the CF tables.

Levels of Agreement

Texts

It is reasonable to assume that experts should find texts and tasks at lower benchmarks easier to rate, and one could perhaps predict a relationship between benchmark level and both extent and ease of agreement. For this reason, passages were rearranged in approximate order of benchmark ratings within each group, text or document. Figure 1 gives the range of ratings for each passage *after discussion*. The extreme ratings are marked by a vertical line and the median by an asterisk. More details, including the ratings before discussion, are included in Tables D3 and D4.

In general, the discussions after initial ratings did not change the ratings very much. Almost all median ratings and ranges of ratings remained the same. The numbers of ratings provided ranged from 35 for those passages earlier in the presentation order to 19 for those at the end. (There were several participants in these panels who did not respond to the demographic and experience survey just discussed.)

```
Text       3     4     5     6     7     8     9    10    11    12
B
J                |----*----|
E                |----*--------------|
D                      |---------*--------------|
F                      |--------------*---------|
A                      |--------------*---------|
I                           |---------*--------------|
C                      |--------------*--------------------|
H                |----------------------------*---------|
G                           |-----------------*--------------|
                      |---------------------*--------------|

Document
K          |---------*--------------------|
L          |---------*---------|
U                |----*---------|
O          |---------*---------------------------|
R          |--------------*------------------------------|
V                |---------*-------------------|
S                     |----*--------------|
T                |--------------*--------------|
N                           |---------*--------------------|
M                           |---------*--------------------|
Q                      |--------------------*--------------|
P                           |--------------*--------------|

           3     4     5     6     7     8     9    10    11    12
```
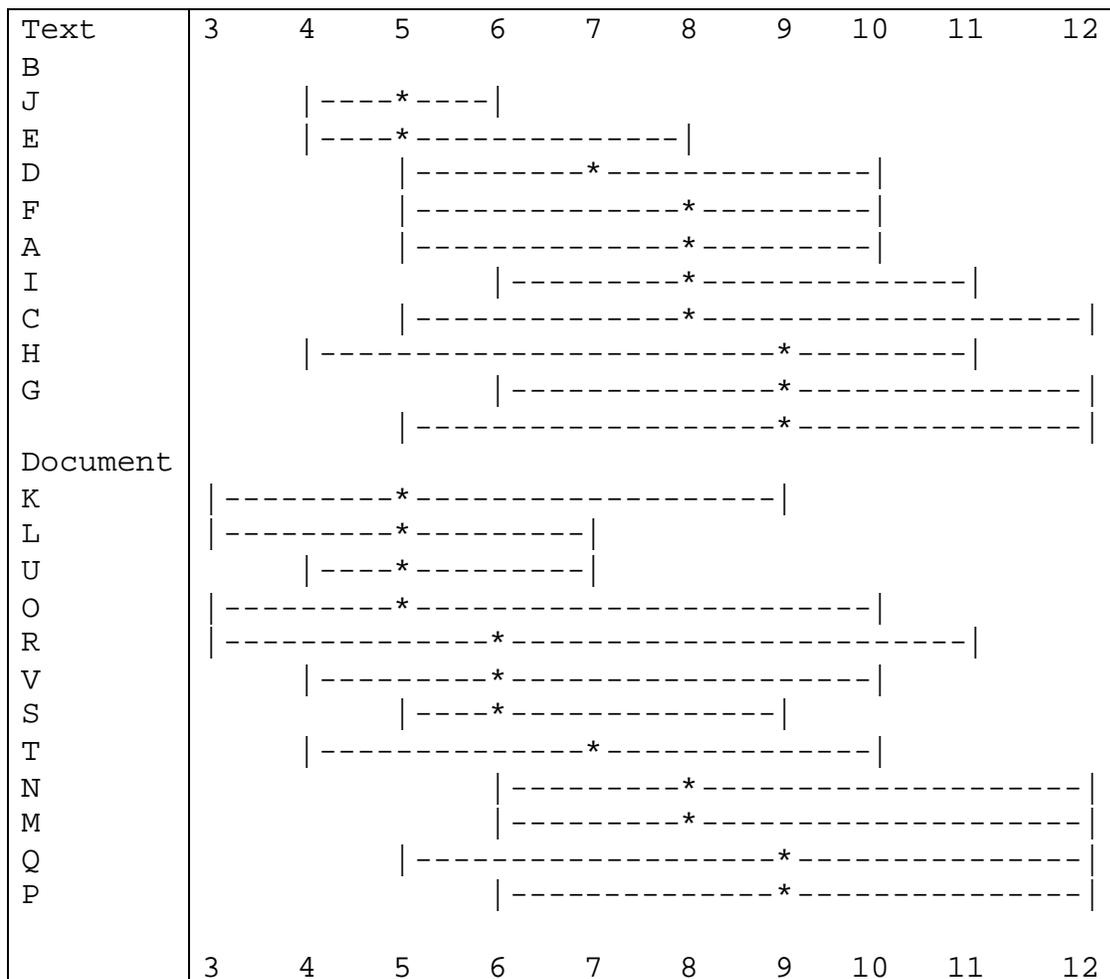
Figure 1: Ranges and Medians of Passage CLB Ratings

Most passages received ratings across a range of at least five benchmarks, which is not unexpected. A single text can apply across a wide range of CLB levels, with the associated tasks indicating suitability for one particular benchmark over another.

Reading Tasks

A total of 26 tasks were associated with the 10 Reading texts. The information in Tables 7a and 7b shows the number of panellists who chose each benchmark and indicates which

benchmarks selections fell into the ranges stipulated by the Comparative Framework. Tables D5 and D6 in Appendix D show the range of benchmarks assigned to the individual tasks in relation to the Comparative Framework. The second column in each of Tables D5 and D6 indicates the "correct" CLB benchmark range, in other words, the benchmarks that would be assigned by a practitioner who knew the true ES level for each task and who then applied the existing Comparative Framework to assign a benchmark level.

Table 7a: Summary of CLB Ratings for Reading Tasks at each ES Level

Benchmark Assigned by Panel	Accepted ES Rating			
	1	2	3	4
	N = 2 items	N = 7 items	N = 10 items	N = 3 items
1				
2				
3	5 Accepted	4	3	
4	14 Accepted	26	13	
5	16 Accepted	36	40	2
6	6	29 Accepted	53	3
7		18	35 Accepted	6
8	3	6	32 Accepted	19
9		2	18 Accepted	5
10			7	10
11				Accepted
12		2		Accepted
Total Accepted	35 (80%)	29 (24%)	85 (42%)	0 (0%)
Total High	9 (20%)	26 (21%)	7 (3%)	0 (0%)
Total Low	0 (0%)	66 (55%)	109 (55%)	45 (100%)
Overall Total	44	123	201	45
Mean	5.8	5.6	6.5	8.3

Each column in Table 7a gives the number of times a particular benchmark rating was assigned by one of the judges to a task with a given ES rating. For example, in the ES "rating 1" column, there are data from two of the 26 tasks that were rated at ES 1. Of the 44 ratings assigned by judges to these items, five were benchmark 3, 14 benchmark 4, etc. The average benchmark rating across these two items was 5.8.

For comparison, the accepted benchmark values for each ES level from the Comparative Framework are also indicated. The extent of agreement is acceptable for ES 1, but not for other ES levels. The group tended to peg the tasks at lower rather than higher benchmarks than expected. The problem intensifies as the ES levels increase.

Document Use Tasks

A total of 34 tasks were associated with the 12 Document Use passages. These tasks all had accepted ES ratings. Table D6 in Appendix D provides task-by-task details. Table 7b, below, is a summary grouped by accepted ES ratings.

Table 7b: Summary of CLB Ratings for Document Use Tasks at each ES Level

Benchmark Assigned by Panel	Accepted ES Rating				
	1	2	3	4	5
	N = 9 items	N = 15 items	N = 10 items	N = 2 items	N = 1 item
1					
2			1		
3	16 Accepted	25	3		
4	54 Accepted	49	14		
5	39 Accepted	70 Accepted	36	3	2
6	13	26 Accepted	25	9	4
7	1	18	23 Accepted	6	2
8		7	7 Accepted	11	4
9		2	9	2 Accepted	1
10			5	7 Accepted	5
11				3	1 Accepted
12				1	Accepted
Total Accepted	109 (89%)	96 (49%)	30 (24%)	9 (21%)	1 (5%)
Total High	14 (11%)	27 (14%)	14 (11%)	4 (10%)	0 (0%)
Total Low	0 (0%)	74 (37%)	79 (65%)	29 (69%)	18 (95%)
Overall Total	123	197	123	42	19
Mean	4.4	5.0	6.1	7.9	7.9

Each column in Table 7b gives the number of times a particular benchmark rating was assigned by one of the judges to a task with a given ES rating. For example, in the ES "rating 1" column, there are data from the 9 of 26 tasks that were rated at ES 1. Of the 123 ratings assigned by judges to these items, 16 were benchmark 3, 54 benchmark 4, etc. The average benchmark rating across these 9 items was 4.4. Again, accepted benchmark ratings according to the Comparative Framework are indicated. The pattern of agreement is similar to that shown in Table 7a, with an intensifying effect at the higher levels.

Commentary

These results may be indicative of certain fundamental difficulties that practitioners seem to have in using the CLB document to rate tasks, and it corroborates the observations made by the writers in an earlier report on the basic distinctions between CLB and ES – that in essence, the ES is intended as a device for rating tasks, while the CLB is intended to rate learners. In addition, given the fact that the worst results are apparent at the highest levels of ability, these results indicate a pressing need to re-examine the format and content of the existing CF, as the potential for high-stakes misapplication will be at its most rampant and consequential in contexts where these higher levels are targeted.

Levels of Confidence and Ease

Respondents were asked about their level of confidence placing questions at a benchmark level. Responses were on a scale from very certain (1) to very uncertain (4). An average of 2.5, therefore, marked the boundary between the *certain* and the *uncertain* zones of the scale. Mean ratings per passage varied from 1.8 to 2.4, that is, slightly more certain than uncertain. (See Table D7 in the appendix for detailed results by passage). There is a modest correlation of 0.36 between the extent of disagreement and the level of uncertainty.

At the individual level, there was virtually no relationship between accuracy and certainty. Taking the median rating as the "correct answer," 77% of those with the "correct answer" reported themselves as *certain* or *fairly certain*, compared to 74% of the entire sample.

A similar pattern occurs concerning the question of how easy respondents found the placement task. (See Table D8 in Appendix D for detailed results by passage). Again, there is a modest correlation of 0.45 between the extent of disagreement and the level of reported ease.

Table D9 gives responses to Question 3, whether the respondent would use the given passage in the classroom. There was a large variation in acceptability of the passages, with Passage I being particularly unpopular. Generally, the text passages are more popular than the document passages. The correlation between possible usage of the passages and level of agreement on placement is 0.45, indicating that respondents are more likely to use passages that they have less difficulty classifying.

Table D10 gives the responses to question 4, concerning level of certainty in assigning a benchmark after discussion. For six of the 22 passages, respondents were less certain after discussion. It should be noted that extremely small numbers of respondents had the time to respond to this question for many of these items.

Table D11 contains detailed responses to the final question concerning ease of reaching agreement on benchmark placement. The results are in contrast to the high level of

disagreement found. Respondents generally found it easy to reach agreement within their discussion groups. However, the discussion groups were small, generally three rather than a "committee of the whole." Thus, groups found it easy to agree, but they were apparently agreeing on a different benchmark than their neighbours. Thus, without some indication of which respondents worked together, these data cannot be interpreted.

Feedback on the Introduction

Respondent comments on the introduction fell into three main categories - formatting, written content, and charts.

Formatting comments can be summarized as follows:

- The introduction should be simple, clear, concise, and user-friendly
- The format is too busy, overwhelming, and cluttered.
- It looks dense and technical.

Comments on the content of the introduction can be summarized as follows:

- The purpose section clearly explains the intent of the document.
- More information is needed on application of the framework
- It is important to stress that this is not an equivalency table.
- Clearer descriptions are needed of CLB and ES and their relationship.

Feedback on the tables used to illustrate comparable items took the following direction:

- There are too many charts, and they are confusing.

For a specific breakdown of comments on the introduction, see Appendix B.

In response to feedback on the introduction, revisions have been made with a focus on the following:

- Improve format to make the document more approachable.
- Simplify wording to make concepts more accessible.
- Provide more background information on the underlying scales.
- Create and augment sections describing applications and limitations.
- Remove charts because they are confusing and because they will not accurately support a re-conceptualized Comparative Framework.
- Clarify and strengthen the point that the CF is not an equivalency table.

Summary and Recommendations

An analysis of qualitative data captures the inherent appeal of bringing together the worlds of CLB and ES. Responses from potential users indicate that the original concept of a Comparative Framework is, in general, well received by the field. Intuitively, practitioners see the value of attempting to understand and describe the relationship between these two very important national standards. In particular, ESL instructors are inspired by the very idea of having a framework that supports them in their efforts to select appropriate workplace tasks for classroom use. It is therefore clear that the development of a high-quality comparative tool will benefit the field by bolstering practitioner confidence and opening channels to facilitate the enhanced use of ES materials in ESL contexts. As a working model, the current version of the CF provides a starting point for this development.

The small study described in this report constitutes the first step in a course of research aimed at uncovering the strengths and weaknesses of the current CF working model. Simply put, this study has revealed that the existing Comparative Framework does not yet fully depict the complex relationship between the CLB and ES. The high degree of disagreement observed on task placement activities at the advanced levels for the skills of Reading and Document Use indicates that further research will be required if this working model is to ultimately fulfil its basic purpose. One of the strengths of the existing CF is its acknowledgement of the many-to-one relationship between the ES and CLB, allowing, for example, that a Reading or Document Use task at a given ES level may be associated with a range of CLB benchmarks. However, the current working model does not yet effectively capture the true intricacy of this relationship, nor does it support users in making clear distinctions between characteristics and features that are unique to one scale and those that are common to both.

This apparent weakness in the CF working model suggests the need for a re-conceptualization. We would therefore recommend that the next stage in the CF development process take the form of a serious study involving experts from across the country. The quantitative aspect should focus on determining, for each of the ES and CLB, which criteria form the core of task and learner placement and which criteria can vary independent of ratings. A pilot study involving CLB-based and ES-based instruments could perhaps be undertaken as part of the empirical investigation. For this purpose, a bank of reliable ES tasks, such as those in the TOWES, might be administered in combination with a credible benchmarks-related instrument, such as the CLBPT. Along with this quantitative work, a qualitative component of the study should more fully explore points of intersection and divergence between the CLB and ES. By delving more deeply into the essence of the CLB-ES relationship, both quantitative and qualitative, such a study would pave the way to a re-conceptualized framework that would perhaps reflect a more accurate and nuanced approach. At the conclusion of the CF development project, when the re-conceptualized framework has undergone a suitable period of trial use, a course of

training and professional development should be put in place to orient potential users to the appropriate and accurate application of its principles.

A possible starting point for exploring the true multi-dimensionality of the CLB-ES relationship might be a visual display that moves the focus away from a linear mapping and toward an exploration of alternative modes of conceptualization. For example, perhaps a Venn diagram could be used as a way to visually capture the key points of convergence and divergence that exist within an intersecting and overlapping flow of criteria. Similarly, a simple shaded chart might be of assistance in examining the grey areas of varying compatibility between the two standards. For samples of basic templates that might be considered as potential starting points for creating these visual displays, please refer to Appendix G.

To sum up, it is our opinion that the existing CF provides a starting point for the necessary research and development that will be required to produce a high-quality comparative tool. However, in its current format, with tables that seem to imply a precise linear relationship, this version of the Comparative Framework may be misleading. In this very small study, we have observed a fairly benign example of what can happen when the framework is put to practical use – ESL practitioners experience a false sense of ease and confidence, while their efforts to apply the framework in selecting level-appropriate workplace tasks for classroom use are inconsistent and inaccurate, particularly at the higher levels. This finding points to a very real potential for misapplication of the framework. Despite strong warnings about the limitations of the CF, its current format could engender the same false sense of ease and confidence in environments where its use is not warranted. In such contexts, where the stakes and confidence levels are high but accuracy and consistency low, the existing framework could have serious consequences. We therefore believe that continued use of the current Comparative Framework should be undertaken only in low-stakes environments. Potential users should be made aware of its inherent limitations and should understand that, as a work in progress, the existing CF is simply a loose set of guidelines that may prove somewhat useful for instructional purposes. Thus informed, practitioners who wish to use the framework can be encouraged to do so in the expectation that their insights, commentary, and feedback will further facilitate the ongoing process of revision and development.