www.itsessential.ca

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

The Canadian Language Benchmarks and the Adult Literacy and Lifeskills Survey: A comparative examination of reading components

By: Gail Stewart
Philip Nagy
Stan Jones

Centre des niveaux de compétence linquistique canadiens

200, rue Elgin, pièce 803
Ottawa, Ontario
Canada K2P 1L5
Tél : (613) 230 - 7729 Téléc: (613) 230-9305
Courriel: info@language.ca
Sie web: www.language.ca

Centre for Canadian Language Benchmarks

200 Elgin Street, Suite 803
Ottawa, Ontario
Canada K2P 1L5
Tel : (613) 230 - 7729 Fax: (613) 230-9305
E-mail: info@language.ca
Web site: www.language.ca

The Canadian Language Benchmarks and the Adult Literacy and Lifeskills Survey: A comparative examination of reading components

1. Introduction

The purpose of this discussion paper is to provide a conceptual comparison of the Canadian Language Benchmarks (CLB) and the Adult Literacy and Lifeskills survey (ALL) frameworks as they relate to adult reading. This comparison is intended to supplement and support an empirical comparison using data from selected instruments based on these respective theoretical perspectives.

We begin with an overview of CLB and ALL, examining the purpose and organization of each framework. We then compare the two under a number of headings, including purpose, construct definition, and scaling. Finally, we conclude with a summary and recommendations for empirical study.

2. Overview of the Canadian Language Benchmarks

2.1. Purpose and Development

The Canadian Language Benchmarks were created in response to a national call for standardization across English as a Second Language (ESL) programs throughout the country. The document was created to describe levels of ability in ESL in a comprehensive and systematic manner that could be understood and interpreted consistently by language training providers nationwide. The descriptors within the document are intended to inform classroom placement, curriculum development and outcomes criteria.

The Canadian Language Benchmarks were developed and refined over a five-year period. The initiative began in 1995 with the creation of a draft document (Citizenship and Immigration Canada, 1995). Revisions to this draft document were informed by results of a national field-testing initiative, by the collaborative efforts of a national working group, and by research undertaken by the developers of the first CLB-based assessment, the Canadian Language Benchmarks Assessment (CLBA) (Peirce & Stewart, 1997).

In 1996, the Canadian Language Benchmarks Working Document (Citizenship and Immigration Canada, 1996) was produced. This version of the benchmarks included more detailed descriptors of task requirements and performance criteria. For four years, this working document was used in the field to inform classroom placement and instruction. Subsequently another revision was undertaken to produce the current version, the CLB 2000 (Centre for Canadian Language Benchmarks, 2000a). The comparative discussion in this report is based on

an examination of the CLB 2000, and all commentary on the CLB from this point forward is made in reference to that version of the document.

2.2. Format and Content

The CLB 2000 addresses four language skills – reading, writing, listening, and speaking. Each skill is organized into three stages – basic, intermediate, and advanced – and each stage comprises four levels of ability, or benchmarks, for a total of 12 benchmarks in each skill.

For each skill at each benchmark, a range of information is provided. *Global Performance Descriptors* afford an overview of the general characteristics of performance at a particular level of ability. *Performance Conditions* describe requirements and limitations associated with a given benchmark. *What the Person Can Do* provides a description of the language functions compatible with the benchmark. This information is further supported by *Examples of Tasks and Items*, a more detailed description of the characteristic features of tasks that are considered to be suitable examples of criterion performance by learners at that level of ability. Finally, *Performance Indicators* describe the responses that define successful performance on a given task.

2.3 Definition of the Reading Construct

In order to tease out a definition of the reading construct implicit in the CLB 2000, it is necessary to first understand the evolution of thought that underpins current theories of language competence and proficiency.

In the mid 1960s, a unidimensional view of the construct of language gave way to a theoretical model that acknowledges four skill areas and takes into account the functional and contextual aspects of communication. Canale and Swain (1980) first posited a four-dimensional model comprising linguistic, discourse, strategic, and socio-linguistic competencies, while Bachman and Palmer (Bachman, 1988) later presented a three-pronged approach which included language competence, strategic competence, and psycho-physiological mechanisms. These multidimensional models are considered by the field to be superior to earlier models of general language proficiency because they reflect the best features of the communicative approach to teaching. They describe the ability to use language to accomplish communicative tasks, rather than simply a grammar-based knowledge of language (Swain, 1984).

The CLB defines the language construct in a manner that most closely aligns with the Bachman and Palmer model, acknowledging that proficiency in a language involves aspects of both competence and performance, both of which are influenced by skill and method factors relating to modality, situation, and context. The underlying principle is a belief that language is intended for communication. In the CLB 2000, the target construct is defined as *communicative proficiency* or "a person's ability to accomplish communication tasks" (CCLB 2000a, VIII). The approach is said to be learner-centred, task-based, and competency-based, a competency being defined as "demonstrable application of knowledge and skills" (CCLB 2000a, VIII).

Based on this approach, the reading construct as one component of communicative proficiency can be defined in the following manner:

> For CLB 2000 purposes, communicative proficiency in reading is the ability to demonstrate reading-related knowledge and skill by accomplishing communication tasks as described in the document.

2.4 Setting levels

The CLB reading scales describe a range of ability from complete beginner to extremely advanced. At the low end of the scale, a reader at benchmark 1 is presumed literate in first language, but has a very limited comprehension of reading texts in the English medium. The person can recognize only a very limited number of familiar words and phrases in contexts that are personally relevant. At the opposite end of the continuum, a reader at benchmark 12 is performing at a highly sophisticated level on a range of complex reading tasks that require critical, conceptual, and inferential strategies.

Setting of levels on the CLB benchmark scale is accomplished by means of an assessment instrument designed for this purpose. A variety of such instruments exist across Canada, and most of these have been developed with classroom placement as the main objective. For this reason, they tend to be relatively short tests that have been validated only for use in low-stakes assessment contexts. Many of these tools are confined to an assessment of benchmarks 1 to 8, as these are the levels most commonly associated with placement in language programs.

Of the available instruments, the Canadian Language Benchmarks Placement Test (CLBPT) is the most likely choice for this empirical study. One advantage of using the CLBPT is the fact that it is readily available, efficiently administered, and widely used for the assessment of large groups of learners. Two potential drawbacks to using the CLBPT are as follows: the test is very short and therefore may not be as reliable as some of the longer CLB-based instruments; and the test does not provide benchmarks for learners whose reading ability exceeds benchmark 8.

The CLBPT was originally designed for low-stakes classroom placement. The objective in developing the test was to keep the administration time as short as possible while still allowing trained assessors to place learners into the appropriate ESL classes. The design, format, and approach of the CLBPT represent the best possible compromise between the mandate for a very short test and the requirement to fully represent the domain of behaviour. The CLBPT reading tool comprises four tasks which increase in difficulty. The first is a word-level task that involves matching graphics to single words and phrases; the second is a short, simple story; the third a general-interest article; and the fourth a longer and more technical text. All of the items associated with these tasks are four-option multiple-choice.

The CLBPT tasks and items were written to reflect CLB benchmark descriptors, but no formal validation was carried out to establish the relationship between the individual test items and the benchmark levels. Instead, to investigate the degree to which results on the CLBPT are compatible with the CLB domain, a study was conducted in 2001 to examine the extent of agreement on reading results between the CLBPT and the Canadian Language Benchmarks

Assessment (CLBA), a longer CLB-based test. This study, involving about 500 cases, was followed by an extensive compilation of reviewer responses to the CLBPT, as well as an analysis of internal consistency.

The CLBPT renders reading benchmarks by the application of the following conversions to a total raw score on 29 multiple-choice items.

Raw Score	Benchmark
0-4	Pre-benchmark
5-7	Benchmark 1
8-10	Benchmark 2
11-13	Benchmark 3
14-16	Benchmark 4
17-19	Benchmark 5
23-25	Benchmark 7
26-29	Benchmark 8

3. Overview of the International Adult Literacy Survey

3.1. Purpose and Development

The ALL reading framework is the outcome of several decades of work on measuring the skills of adults. Beginning in the 1970s, a number of attempts were made to find a way to document the reading skills of adults in the United States. Early attempts sought to identify a minimum set of competencies in reading, in line with the then current trend towards mastery learning and minimum competency educational movements and to determine the number of adult "illiterates" in the country (L. Harris Associates, 1970, 1971; Northcutt, Selz, Shelton, Nyer, Hickok, & Humble, 1975).

As part of the *Reading is Fundamental* effort, Murphy (1973, 1975) introduced a more sophisticated approach in his 1970s survey. He asked his respondents to respond to questions based on real documents (as had the functional competency movement), but was not interested in how many adults fell into the literate and illiterate categories. Rather he was interested in how test questions on different kinds of everyday documents affected reading scores and on how well different groups of respondents (young/middle age/older, for example) did with different tasks and texts. Murphy did not classify individuals. This was not his intent, but his methodology also did not allow him to do so. Because his primary interest was in how well adults could read a wide range of texts, and as this range was greater than any individual would be able to deal with in a single test setting, different groups of individuals answered different sets of tasks on different texts.

Since the time of Murphy's work, psychometrics has developed tools for assigning scores on a common scale to individuals who take over-lapping, but not identical, sets of text items. The ALL framework is a direct development of Murphy's approach using these new tools to provide

a profile of skills in the population. It shares with Murphy both a lack of interest in classifying individuals as literate or illiterate and a focus on identifying the skills of individuals in terms of the kinds of reading tasks they can successfully accomplish

The ALL framework, as it currently exists, was developed through a number of studies of adult literacy in North America. Kirsch and Mosenthal (1991) first developed a version of the theory that underlies the framework as a way of interpreting the results of the Young Adult Literacy Survey. Refinement of the theory continued through work on a survey for the Department of Labor (Kirsch & Jungeblut, 1992) and in the National Adult Literacy Survey (Kirsch, Jungeblut, Jenkins, & Kolstad, Andrew, 1993). At the same time, in the survey of Literacy Skills in Daily Activities, Canadian researchers were showing that the framework was appropriate not just for adult reading in English, but that it applied equally well to understanding adult literacy in French (Satin, Kelly, Montigny, & Jones, 1991). The International Adult Literacy (IALS), conducted with the OECD in three waves through the mid- and late-1990s, demonstrated its application across a wide variety of languages. The ALL framework is essentially identical to that used in IALS with one exception, noted below.

3.2. Format and Content

The ALL adult literacy framework is part of a larger adult skills study, which includes frameworks for numeracy, analytical reasoning, teamwork and information and communication technology. Experts in each field developed the frameworks more or less independently. An overarching framework has been prepared, but this has been derived from the individual frameworks, and not vice-versa.

The literacy framework identifies five levels of skill. These levels were empirically derived in that the cut points between levels were established after examining the data and finding points that best seemed to group individuals with similar skills. More recent research (as yet unpublished) seems to support the basic correctness of the cut points in terms of underlying component skill differences. Nonetheless, it is important to stress that the levels of reading skill in the ALL framework were not theoretically pre-determined. It is also important to keep in mind that none of the levels can be interpreted as “illiterate”. Most individuals at the lowest level (1) are usually able to carry out reading tasks of some kind. Instead the levels should be interpreted as an increasing ability to carry out reading tasks of greater variety and complexity. . In the ALL implementation ability is judged probabilitistically, i.e., individuals are placed at a level by getting 80% or none of items of a given level of difficulty correct, but only after theoretical item difficulty has been confirmed empirically.

3.3 Definition of the Reading Construct

The ALL framework defines *literacy* as:

> …using printed and written material to function in society, to achieve one’s goal, and to develop one’s knowledge and potential.

Each of the elements in this definition plays a role in defining the domain of the test and how it is realised in the test setting. The framework identifies three essential characteristics of tasks:

Adult context/content

The materials on which the test is based must be those used by adults as part of their daily life and must be as representative of common texts as possible. While there has been no definitive analysis of the contexts and content of adult reading, enough is known to suggest that sampling from the following categories covers the major areas:

1. Home and family
2. Health and safety
3. Community and citizenship
4. Work
5. Leisure and recreation

The framework is clear that these categories play no direct role in assessing differences in ability. Rather their role is to ensure that the results of the test can be legitimately generalized to the main domains of adult reading. . Although the combination of the BIB matrix design and the statistical techniques used to estimate individual proficiency combined with the sampling of context and content to yield proficiency estimates that are relatively unaffected by familiarity or lack of familiarity.

Materials and texts

The framework identifies two major categories of texts, prose and document.

1. Prose texts are often called continuous texts. They are typically composed of full sentences organised into paragraphs, which may or may not be organised into labelled sections, chapters, etc.
2. Document texts are often called non-continuous texts. The TV program grid is a common document in North America. The basic structure is a list, which may be organised into matrix patterns or a graphic pattern.

While these, too, do not contribute directly to consistent differences in ability (prose texts are not inherently more difficult than document texts), population groups do differ in the overall success with different types. For example, in IALS German adults were able to deal more easily with document tasks than were Canadian adults, but these two groups did not differ in their abilities with prose-based tasks.

Document texts, especially graphic texts, frequently contain numerical information. While tasks requiring arithmetic operations (adding, subtracting, etc.) were included in a third text type, quantitative literacy, in IALS and previous work, these tasks now form part of a new numeracy framework. Still, tasks using numerical information, but not requiring arithmetic operations, are common in everyday life and remain part of document literacy in ALL as they were in IALS.

Process/strategies

The characteristics of reading tasks that do affect their relative difficulty concern the structure of the information (Type of Match), the semantic complexity of the information (Type of Information), and the complexity of relevant information in the text (Plausibility of Distractors). The theory of process and strategy that underlies the ALL framework is a rich and complex one

and cannot be fully summarised here (see the full ALL document framework), but a brief description of each follows:

> *Type of Information*: The semantic properties of the information required to complete a cognitive task, such as a reading task, have long been known to affect difficulty. Thus, cognitive tasks requiring concrete information are easier than those requiring abstract information. The ALL framework has an algorithm for evaluating the relative abstractness of the information of a task.
>
> *Type of match*: The task of obtaining relevant information from a text may be as simple as finding a single match to a key phrase or as complex as having to combine information from one or more texts with prior knowledge to generate a new idea. The prominence of the information in the text also plays a role in the difficulty in locating it. The ALL framework uses another algorithm, derived from analysis of a variety of reading tests, to evaluate the complexity of the match.
>
> *Plausibility of distractors:* Obtaining the necessary information is easier if there is no conflicting information in the text(s). (Distractors here, refers to information in the text, not in the test question.). There is an ALL algorithm for assessing this, as well.
>
> *Readability of text:* The standard estimates of text readability (such as the popular Flesch measure) turn out to have little effect on task difficulty, once the other three process / strategy characteristics are taken into account. This is as one would expect, because reading tests have long been constructed with items of widely varied difficulty based on one text. A very simple text (in the Flesch sense of simple) might not support a task with a complex *type of match* property, but generally any text would support both relatively easy and relatively difficult tasks.

ALL items are written with these process/strategy characteristics in mind and test forms are constructed to have tasks with a full range of process difficulties. It is important, however, to understand that tasks are not put in an ALL level based on their process difficulties; the level of an item is determined solely by its IRT values. The correctness of the process/strategy framework depends on how well the process characteristics predict the empirical IRT difficulties.

3.4 Setting levels

ALL, like IALS, uses a two-parameter item response model to score the tests. IRT models are dual models, providing an estimate of individual ability and task difficulty on the same scale. This allows the ability of an individual to be interpreted directly in terms of item properties. IRT estimates of ability and difficulty typically are centred on 0, yielding scores that mostly lie between –3 (low scores) and +3 (high scores). To avoid negative numbers, scores and difficulties on ALL are transformed by the formula: IRT*50 + 250. The result is a range of scores from 0 to 500, with most scores between 100 and 400.

Because item difficulty is not a single number but a curve of probabilities of success, it is necessary to select a single probability point for all items. Because ALL is concerned with the ability to consistently use printed and written material, item difficulties are considered to be the point at which individuals have a .8 probability of success.

Because an individual's score is the likelihood of successfully completing tasks at a given level of difficulty, it is possible to relate that score to characteristics of the items. For these item characteristics to provide a description of individual ability requires that items with similar degrees of difficulty share many characteristics; that indeed, appears to be the case. In studies over a number of surveys, consistently high correlations between the process/strategy variables and item difficulty measures strongly argue that these are the characteristics related to difficulty and, hence, to ability.

Because it is often difficult to work with the continuous range of numbers that the IRT model generates as scores, a series of cut-points were established to provide categories of ability. The correlations seemed to work best with five categories:

Level 1: 0 – 225
Level 2: 226 – 275
Level 3: 276 – 325
Level 4: 326 – 375
Level 5: 376 – 500

Each task is then assigned a level based on the ability required to have a success likelihood of .8. Individuals are then assigned to a level such that they have a greater than .8 likelihood of answering items in the level below (except level 1) and a less than .8 probability of answering items in the level above (except level 5). For example, an individual with a score of 285 would be assigned level 3 because such an individual would have a greater than .8 probability of success with a level 2 task (which requires no greater score than 275 for the most difficult task) and a less than .8 probability of success with a level 4 (which requires an ability of 326 for .8 success with the easiest task). In reports from ALL, individual ability will be described in terms of characteristics of the items that match that ability.

4. Comparative Examination

Points of comparison can be drawn between the CLB and IALS frameworks with respect to their fundamental purpose and design, the extent to which their basic constructs overlap, and their likelihood of producing test score interpretations that allow for mapping onto a scale.

4.1 Fundamental Purpose and Design

The CLB is designed to provide a descriptive summary of proficiency in ESL. Its descriptors are aimed at capturing the performance of an immigrant population within the Canadian context. The original design and content of the document is based on the intuitive predictions of a group

of experts in the ESL field, and the scales presented in the document were not empirically validated as a condition of its development. The only validation that has been carried out to date has been on certain ESL assessment instruments aligned to the document's specifications.

In contrast, the IALS is a measure of official language literacy. It has been developed to describe the performance of individuals using types of written and printed materials common within a country. The basic validation of the framework is that the results are consistent with the expectations of the underlying theory of adult reading theory. In addition, the scores have been demonstrated to be highly related to a range of social-economic educational and health outcomes and to be the product in large measure of formal instruction.

Because the ALL framework provides specifications for the tasks and items included in the survey, it can be used to create a test blueprint, whereas the CLB is not a test blueprint. While the CLB may serve as an appropriate underpinning for test development, it does not set forth explicit specifications, nor does it present a hierarchy of difficulty that has been empirically validated. Its purpose is to provide a framework for describing what learners tend to do, so that educators can conceptualize classroom levels in a standardized manner across the country and can plan programs and curricula accordingly. As such, the CLB is used as an interpretive tool by a variety of professionals and organizations, ranging from individual instructors to program coordinators, funding bodies, and government agencies. In contrast, the ALL is intended for testing purposes. While it has been used for instructional applications in some contexts, its purpose is to facilitate a fair and valid comparison of literacy levels among the populations of different countries, including both those with substantial immigrant populations, and those without such populations.

4.2 Definition of the Reading Construct

The CLB and ALL definitions of reading both focus on understanding for different purposes and in different contexts. Both are task-based and competency-based. In a broad sense, then, they are compatible, and contain substantial overlap.

ALL tasks are generally short answer, while many of the standardized tests of reading based on the CLB rely on multiple-choice items. The rationale for the multiple-choice item type tends to be based on administrative convenience, efficiency, and the importance of obtaining a reading measure that is not contaminated by influence from other skills, particularly writing. Those aspects of the CLB reading domain that are not testable in an objective format, such as "explain how something works (in nature or man-made) based on a text" (CLB, p. 91), or "use two bus route maps/schedules: locate a time of departure; coordinate with transfer to reach destination" (CLB, p. 89) are more likely to be found in classroom assessments or curriculum materials.

Most of the assessments that have been developed to align with the CLB 2000 avoid the explicit use of numeracy items. For example, a reading item that requires knowledge of mathematics or a text involving a bar chart or pie graph would be excluded as a potential source of construct-irrelevant variance. Thus, although the CLB document includes task types that require the application of numeracy skills, many of the existing CLB-based tests do not. As set out above, because matrix documents with numerical information and graphs are an important part of

everyday literacy, they form an essential part of ALL document literacy. Previous versions of the ALL framework (such as that for IALS) did include a separate measure of quantitative literacy, but a richer numeracy framework that includes, but goes beyond, the IALS quantitative literacy measure has been developed for ALL. ALL does include texts with numerical information in the document literacy framework as these are common read world tasks, but none of them require any calculations.

Because both frameworks are based on a communicative model of language, the representative tasks, as described, are authentic, meaningful, and embedded within relevant social, academic, or business contexts. This feature lends a high degree of face validity to both documents, as users tend to readily accept the relevance and practical application of descriptors that appear to describe real-life contexts and situations. In the ESL community, the authentic and communicative approach to the CLB is considered very beneficial in terms of positive curricular influence. What this means is that the CLB framework ensures that learners are taught to really comprehend and express meaning, rather than to simply internalize grammar rules and memorize lists of vocabulary.

However, for testing purposes within an ESL population, the complexities of a communicative approach introduce certain tensions that can sometimes be problematic. The requirement of authenticity, for example, must be carefully balanced against potential interference from cultural bias. A truly authentic task, by definition, cannot be separated from the cultural context in which is carried out, and yet, a test of ESL must be equally fair to learners from a variety of linguistic and cultural backgrounds (Norton & Stewart, 1999). For this reason, the majority of standardized CLB-based tests, and in particular those intended for high-stakes usage, tend to steer away from tasks that are considered to be culturally biased, and as a result, these instruments may not reflect the full impact of cultural variables on learner performance. Several aspects of the ALL design mitigate against problems of cultural bias. First, items are subjected to a sensitivity review that identifies and eliminates problematic content. Second, a matrix design is available to provide much greater content coverage, reducing the potential impact of any individual items. Third, procedures can be used to identify aberrant response patterns in empirical data, thus providing yet another guard against bias.

4.3 Scaling the Reading Construct

It is possible to predict empirical task difficulty from item characteristics in both the ALL and CLB frameworks. In the ALL framework, specific algorithms are available to evaluate tasks for the type of information requested, the plausibility of distracting information, and the nature of the cognitive task required of the reader/ test taker. These rules associate lower scale values with, for example, concrete tasks and searches for simple information, and higher values with more abstract tasks and searches for more complex information. Different schemes are presented for prose and document literacy. The result is a “difficulty score” for each item. These scores have been compared with empirical difficulty data, with two results. First, judgmental rating of items on four variables can be used to predict a substantial proportion of variance in item difficulty indices (87% for prose tasks and 76% for document tasks). Second, detailed comparison of the item ratings and empirical difficulty indices reveals consistent shifts in the nature of the ratings as difficulty increases. Thus, success on items of specific rated difficulty

can be associated with specific ranges of scaled test scores. That is, scores can be associated with skill levels, and the skill levels can be defined in terms of item characteristics.

This is theoretically possible in the CLB framework as well. However, the prediction of item difficulty is much more exact in the ALL framework compared to CLB. There are three reasons why this is so.

First, holistic judgments are made in CLB item rating compared to ALL, resulting in placement of an item on the benchmark scale. Thus, there is only one variable available for the prediction equation. In one test currently under development, these item ratings account for some 50% of variance[1]. This is poorer than the correlation of single best predictor in ALL prose tasks (80%) and document tasks (72%).

Second, the CLB is primarily an instructional rather than assessment document. As such, it focuses on the relevant tasks and strategies that are assumed to characterize each benchmark, rather than on quantifiable differences that distinguish between ability on contiguous benchmarks. The use of the CLB for assessment purposes is an adaptation and an extension of its original purpose.

Third, we would expect better prediction of empirical difficulties using the ALL scheme because items were written with these quite specific criteria in mind. In contrast, manipulation of the difficulty levels of items designed to fit the CLB document would much less exact, being based on a less specific blueprint.

4.4. Summary

The reading constructs used by CLB and ALL overlap considerably and are compatible with each other. There are two areas in which each is unique. The CLB construct includes areas of reading that can be tapped via short-answer, multiple-choice, oral response and written summary. While these varying response modalities are indicated in the document, the full range is seldom explored in CLB-based assessment. The ALL construct focuses on areas of the reading construct associated with open-ended written response. Most CLB-based assessment tools tend to exclude texts using numerical information while ALL includes them.

In both the ALL and CLB frameworks, item judgments can be used to predict empirical difficulties and define scales. This has not yet been done for a CLB-based test. In the development and validation of the only likely candidate, the CLBPT, sample sizes were too small for IRT analysis. One would expect less success in predicting empirical difficulties from item ratings in a CLB framework. CLB item ratings are holistic, yielding only one predictor variable; multiple-choice distractors cannot be rated through the CLB framework; and the CLB is an instructional framework adapted for assessment, rather than an assessment framework *per se*.

[1] Cited by permission of the Centre for Language Training and Assessment.

5. Recommendations for Empirical Study

Any empirical comparison of scores from CLB-based and IALS-based tests would require supportive conceptual analysis. As noted, the CLBPT is the most readily-available instrument for such a study. The study would need to include:

- Individuals taking both the ALL reading test and the CLBPT.
- Determination of appropriate scores and levels for each test
- Development of a crosswalk to relate scores and levels on ALL to scores and benchmarks on CLB
- It is not necessary to put CLBPT items on the ALL scale or ALL items on CLB scale as the study should not attempt to use ALL items to determine CLB benchmark or vice-versa.

Given these ancillary studies, the main data would come from recently-arrived immigrants with a full range of language skills as defined by the two scales. Item level data from both tests would be required, administered in counterbalanced order, sufficiently close in time that little learning has taken place. The most likely source of lower-skilled candidates would be English language classes. Candidates of higher ability tend to be harder to locate.

References

Adult Literacy and Lifeskills Survey. (Undated). *An Overarching Framework for Understanding and Assessing Lifeskills.* Working Draft.

Bachman, L.F. (1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language Acquisition*, 10, 2, 149-164.

Canale, M. (1988). The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67-84.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1, 1-47.

Centre for Canadian Language Benchmarks (2000a). *Canadian Language Benchmarks 2000: English as a Second Language for adults.* CIC, Ottawa.

Centre for Canadian Language Benchmarks (2000b). *Canadian Language Benchmarks 2000: ESL for Literacy Learners.* CIC, Ottawa.

Citizenship and Immigration Canada (1996). *Canadian Language Benchmarks: English as a Second Language for adults.* Ottawa.

Citizenship and Immigration Canada (1995). *Language Benchmarks: English as a Second Language.* Ottawa.

Harris L. & Associates. (1970). *Survival Literacy: Conducted for the National Reading Council*. Louis Harris and Associates, New York.

Harris L. & Associates. (1971). *The 1971 national reading difficulty index: A study of reading ability for the National Reading Council*. Louis Harris and Associates: New York.

Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding What Was Measured.* Research Report RR-01-25. Princeton NJ: Educational Testing Service.

Kirsch, I. S., & Jungeblut, A. (1992). *Profiling the literacy proficiencies of JTPA and ES/UI populations*. Educational Testing Service, Princeton, NJ.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. National Center for Education Statistics, Washington, DC.

Mosenthal, P. B., & Kirsch, I. S. (1991). Toward an explanatory model of document literacy. *Discourse Processes* 14: 147-80.

Murphy, R. T. (1973). *Adult functional reading study*. (PR 73-48). Educational Testing Service, Princeton, NJ.

Murphy, R.T. (1975). *Adult functional reading study*. (PR 75-2). Educational Testing Service, Princeton, NJ.

Northcutt, N., Selz, N. Shelton, E., Nyer, L., Hickok, D., and Humble, M.. (1975). *Adult functional competency: A summary*. Division of Extension, University of Texas, Austin, TX.

Norton, B. & Stewart, G. (1999). Accountability in Language Assessment of Adult Immigrants in Canada. *The Canadian Modern Language Review*, 56, 2, 223-244.

Peirce, B. N. & Stewart, G. (1997). The development of the Canadian Language Benchmarks Assessment. *TESL Canada Journal*, 114, 2, 17-31.

Satin, A., Kelly, K., Montigny, G., & Jones, S. (1991). Canada's Survey of Literacy Skills Used in Daily Activities: Survey preparation and measurement issues. L. Benton, & T. Noyelle (eds.), *Adult Literacy and Economic Performance in Industrialized Countries*. The Eisenhower Center for the Conservation of Human Resources, Columbia University, New York.

Swain, M. (1984). Teaching and testing communicatively. *TESL Talk*, 15, 7-17.