Citizenship and Immigration Canada
Citoyenneté et Immigration Canada

An examination of the Canadian Language Benchmark data from the Citizenship Language Survey

Tracey M. Derwing
University of Alberta
Murray J. Munro
Simon Fraser University
Marlene Mulder & Marilyn Abbott
University of Alberta

March 2010

Canada

Research and Evaluation

Acknowledgements

The authors would like to acknowledge Martha Justus and Jessie-Lynn MacDonald of Citizenship and Immigration Canada for their support of this study. We also thank Mehti Krasniqi for his work recoding some of the data. We thank the Prairie Metropolis Centre for support. Fay Sylvester and Lan Chan-Marples of the Faculty of Education at the University of Alberta were most helpful. Finally, the first two authors both taught ESL and citizenship classes to adult immigrants many years ago, and still value those experiences. It is our hope that this study, and other related research, will help lead to the elimination of barriers for newcomers that keep some from participating as full citizens in Canadian society.

Ci4-42/2010E-PDF
978-1-100-16164-8

Table of contents

Executive summary iii

Introduction 1

Definition of 'Eligible case' 1

Characteristics of the cases 1

Countries *1*

Languages *1*

Language training *2*

Formal education in Canada *2*

Occupation *3*

Language at work *3*

Citizenship test information *3*

Canadian Language Benchmark analyses 4

Country of origin 4

First language 4

Language training in Canada 7

Formal education in Canada 9

Occupation in Canada 9

Language at work 11

Citizenship test scores and CLBA scores 11

Differences across cities 12

Additional correlational and regression analyses 13

Simple correlational analyses 13

Multiple regression analysis 14

Problems with the data set 16

Data cleaning 16

Non-response or missing data *16*

Inconsistent coding *16*

Lack of probing *17*

Selection of participants *17*

Shortcomings of the survey instrument 17

Insufficiently focused or ambiguous questions *17*

Inadequate participant background information *18*

Changes to survey midstream *18*

Summary statement and recommendations 19

Summary statement 19

Recommendations 19

References 23

Appendix A: Number of cases by country 24

Appendix B: Coding of countries in which English is an official language 25

Appendix C: Coding of countries by geographic region 26

Appendix D: Coding of languages as originally reported 27

Appendix E: Classification of occupations into three levels of job skills 28

List of tables

Table 1: Native speakers of English, French and cases by language group. 2
Table 2: Formal education in Canada accessed by participants. 3
Table 3: Mean CLBA scores by geographic region. 4
Table 4: Mean CLBA scores according to sex and mother tongue. 7
Table 5: Mean CLBA scores according to level of formal education in Canada. 9
Table 6: CLBA mean scores and N's for occupations with over 25 participants reporting. 10
Table 7: Mean CLBA scores by city. 12
Table 8: Intercorrelations among selected variables of interest. 13
Table 9: Final regression variables with significant effects. 15

List of figures

Figure 1: Source of language training. 2
Figure 2: Mean CLBA score by language group. 5
Figure 3: Mean CLBA scores by source of language training. 8
Figure 4: Mean CLBA scores for five Canadian cities by language training source. 8
Figure 5: Job skill level by immigration class. 11

Executive summary

At the request of Citizenship and Immigration Canada, we undertook an examination of an existing data set based on a pilot survey of immigrants in six cities. The data covered a range of variables, including demographic information, questions about language training, citizenship test scores, and Canadian Language Benchmark Assessment (CLBA) scores for listening and speaking combined. We undertook a descriptive analysis of 3,827 cases in the dataset, focusing on age, age at immigration, gender, length of residence in Canada, immigration class, country of origin, mother tongue, language training received in Canada, formal education received in Canada, most recent and longest held occupation in Canada, language used most frequently at work, and citizenship test scores.

In accordance with the terms of the contract, we evaluated the relationships between many of the factors listed above and Canadian Language Benchmark Assessment listening and speaking scores. One of the noteworthy findings was a large range of average CLBA scores across mother tongues, such that speakers of several East Asian and Southeast Asian languages tended to score significantly lower than the rest of the language groups. Scores also varied according to type of language training in Canada. Those who reported studying in LINC programs scored significantly lower than those who studied in fee-based programs, who in turn scored lower than those in high school/college/university programs. Other influences included formal education in Canada, immigration class, and occupation.

Responses to the question about language used most frequently at work indicated that 85% of immigrants use either English or French, whereas the most common non-official languages reported were Cantonese and Mandarin, but these made up only a small percentage of the total.

We conducted a multiple regression analysis aimed at identifying the best combination of predictors of CLBA scores. A final model covering variables relating to language training, mother tongue, level of education in Canada, age at immigration, immigration class and city of residence accounted for over 41% of the variance in CLBA scores. Conclusions about causal factors in language proficiency cannot be drawn from this analysis, but the outcome is suggestive of directions for further research.

Included in this report is a list of problems associated with the data, and a set of recommendations for future studies of the language development of adult immigrants in Canada.

Introduction

In the summer of 2006, Citizenship and Immigration Canada requested a comprehensive analysis of an existing data set. Data for the pilot test were collected in six cities from immigrants who were waiting to take their citizenship test. Assessors administered the combined listening and speaking component of the Canadian Language Benchmark Assessment tool (CLBA). In addition, the participants provided demographic information on a wide array of variables. The chief purpose of this research was to examine the relationships between these variables and the language proficiency of the immigrants, as determined by the CLBA scores. The variables of interest were gender, mother tongue, country of origin, length of residence in Canada, nature of language training (LINC or other), full time versus part time language training, length of language training in months, level of formal education, age (chronological), age on arrival, citizenship test score, immigration status and language used at work.

In this report, we will present an overview of the characteristics of the eligible participants in the study, followed by an analysis of the relationships between CLBA scores and participant variables. Additional correlational and regression analyses will then be presented. Problems with the data set will be presented, and finally, recommendations for future research will be suggested.

Definition of 'Eligible case'

For our study we defined an 'eligible case' as an individual whose "mother tongue" was not an official language, who completed the CLBA and who provided consent to participate in the study. There were a total of 3828 cases, later reduced to 3827 that met these criteria (one case was lost in the analysis).

Characteristics of the cases

Slightly more of the participants were females (53.9%) than males (46.1%). The ages of the participants ranged from 18 to 61, with a mean of 35.9 years. Age at immigration ranged from 3 to 56 years, with a mean of 30 years. Length of residence in Canada ranged from 2 to 26 years, with a mean of 5.9 years. According to participants' reported immigration class, 20% were refugees, 34% were family class, and 46% were independent class.

Countries

Only 3 countries contributed more than 5% of the respondents – China (19.3%), India (10.1%), and the Philippines (8.4%) – for a combined total of 37.8%. An additional 20 countries contributed at least 1% of total cases each. See Appendix A for a breakdown by country.

Languages

More than 108 mother tongues were represented. A precise number is not possible because of vague classifications or possible misclassifications at the interviews. For example, some individuals were recorded as 'Chinese' (vague), 'Swiss' (no such language), or Gad (impossible to know whether this was an abbreviation for a language name such as Gaddang). The most frequent 5 mother tongues, each representing more than 5% of the total eligible cases, were Mandarin (15%), Cantonese (7.5%), Tagalog (7.4%), Punjabi (6.5%), and Arabic (5.8%). An additional 18 languages contributed at least 1% of the total cases each. Table 1 presents a breakdown by language grouping (see discussion below) of eligible cases along with counts of English and French speakers.

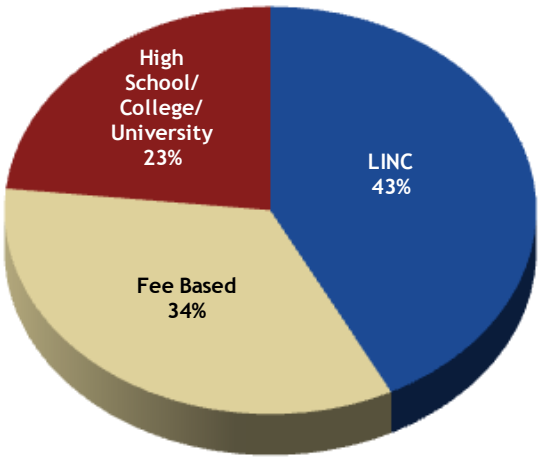
Table 1: Native speakers of English, French and cases by language group.

Language	N	Language	N
Mandarin	575	Tamil-Dravidian	133
English	513	Korean-Japanese	117
Hindi-Punjabi	350	Serbo-Croatian	106
Semitic	267	Other Slavic	111
Filipino-Indonesian	301	Romance	117
Cantonese	289	Other languages	98
Iranian	260	Vietnamese-Cambodian	59
Somali-Oromo	247	Niger-Congo	49
Other East Indian	223	Turkic	28
Other Chinese	191	French	31
Other European	166		
Russian-Ukrainian	140	**Total**	**4,371**

Language training

The cases were almost evenly split between those with language training in Canada (50.7%) and those without (49.3%). Length of training ranged for as long as 114 months, with a median of 6.0 months. Nearly two thirds (61.3%) of those who accessed language training attended full-time, with 38.7% taking part-time classes. With respect to language training progress, only 6.2% indicated that they had completed their language studies, while 67.7% indicated that they were still in the process of studying, and 26.1% reported that they had not completed their language training and were no longer studying an official language.

Figure 1: Source of language training.



An examination of Figure 1 reveals that 42.5% of those who received training accessed LINC, another 34.2% obtained fee-based official language training, and 23.3% studied an official language at a high school, college, or university. A Chi square analysis indicated no difference across immigration class in attendance across the three language training groupings (Chi square [df = 4], 4.22, p = .377).

Formal education in Canada

The lack of information from the participants about their formal education in their country of origin greatly reduces the usefulness of the information about education in Canada. Half the participants (49.7%) had received formal education in Canada. However, a level was recorded for only 62% of these cases. The data on level of training are therefore also of limited value. It is worth noting that nearly half of those participants who indicated the type of education they had accessed identified studying for a university or college diploma (see Table 2). The next most common response was continuing education, but unfortunately, the diverse array of offerings within this category (including language training) makes it impossible to draw any meaningful conclusions.

Table 2: Formal education in Canada accessed by participants.

Level of Education	Percent
University or College Diploma	46.5
Continuing Education	23.5
High School	12.3
Computer Training	11.5
Apprenticeship	6.2

Two thirds of those individuals who had attended formal education programs in Canada were full-time students. Sixty-one percent reported having completed their programs, 26% were still in progress and the remaining 14% identified their status as incomplete.

Occupation

In order to interpret and analyze the questions on recent and longest held occupations, we recoded the data employing the National Occupation Coding system (NOC). This resulted in 26 categories; the numbers of participants employed in recent occupations are shown in Table 6. Only seven categories accounted for more than 5% of the total participants each. When we compared recent occupations with the longest held occupations, we discovered practically no difference; again, the same seven categories accounted for more than 5% of the total participants. The only difference worth noting was the change in position of Clerical Occupations and Professional Occupations in Natural and Applied Science. In the recent occupation data, 8.3% of the eligible participants reported holding a Clerical Occupation, and 8.1% held a Professional Occupation in Natural and Applied Science. These figures were slightly different in the Longest Held Occupation category, where 8.0% of the respondents were in a Professional Occupation in Natural and Applied Science, and 7.8% held a Clerical position. Because there were so few differences in these two analyses, we have reported only the recent occupations.

Language at work

More than 85% of participants reported using primarily an official language at work in their most recent occupation: 81.4% of cases reported English, while 4% identified French. The most common nonofficial language category was Chinese: Chinese (3.4%), Cantonese (2.9%), and Mandarin (2.4%). No other language contributed more than 1% of cases to the total. The findings for language used most often at work in the longest held occupation are remarkably similar to the recent occupation percentages.

Citizenship test information

The overall pass rate for the citizenship test was 96.2%. Ninety eight percent of the native English speakers and 100% of the native French speakers passed the test, while 95.9% of the participants with another mother tongue passed. There was no statistically significant difference across the three groups (Chi square [df = 2] = 4.02, $p = .134$). Nor was there a difference in mean scores on the test, with English speakers scoring an average of 18.8/20, French speakers averaging 19.2 and speakers of other languages averaging 18.5, F (2, 2562) = 2.58, $p = .076$.

Canadian Language Benchmark analyses

As outlined in the original letter of agreement, the chief purpose of this analysis is to examine the relationship between scores obtained on the Canadian Language Benchmark Assessment tool for speaking and listening (the combined version) and an array of factors that might influence second language learning progress. In the sections that follow we will consider each of these factors in turn: country of origin, first language, language training in Canada, formal education in Canada, occupation in Canada, language at work, citizenship test scores, and city where the test was administered.

Country of origin

To facilitate a general analysis, we have created 8 broad geographic categories, as indicated in Table 3 (below). An examination of CLBA scores revealed the highest scoring groups were the South Pacific group (Fiji), followed by Europeans and Central/South America/Caribbean. The area showing the lowest CLBA score was East Asia; this score was significantly lower than the scores from any other region. The largest representation within this category was from China. The totals include only speakers of languages other than English or French.

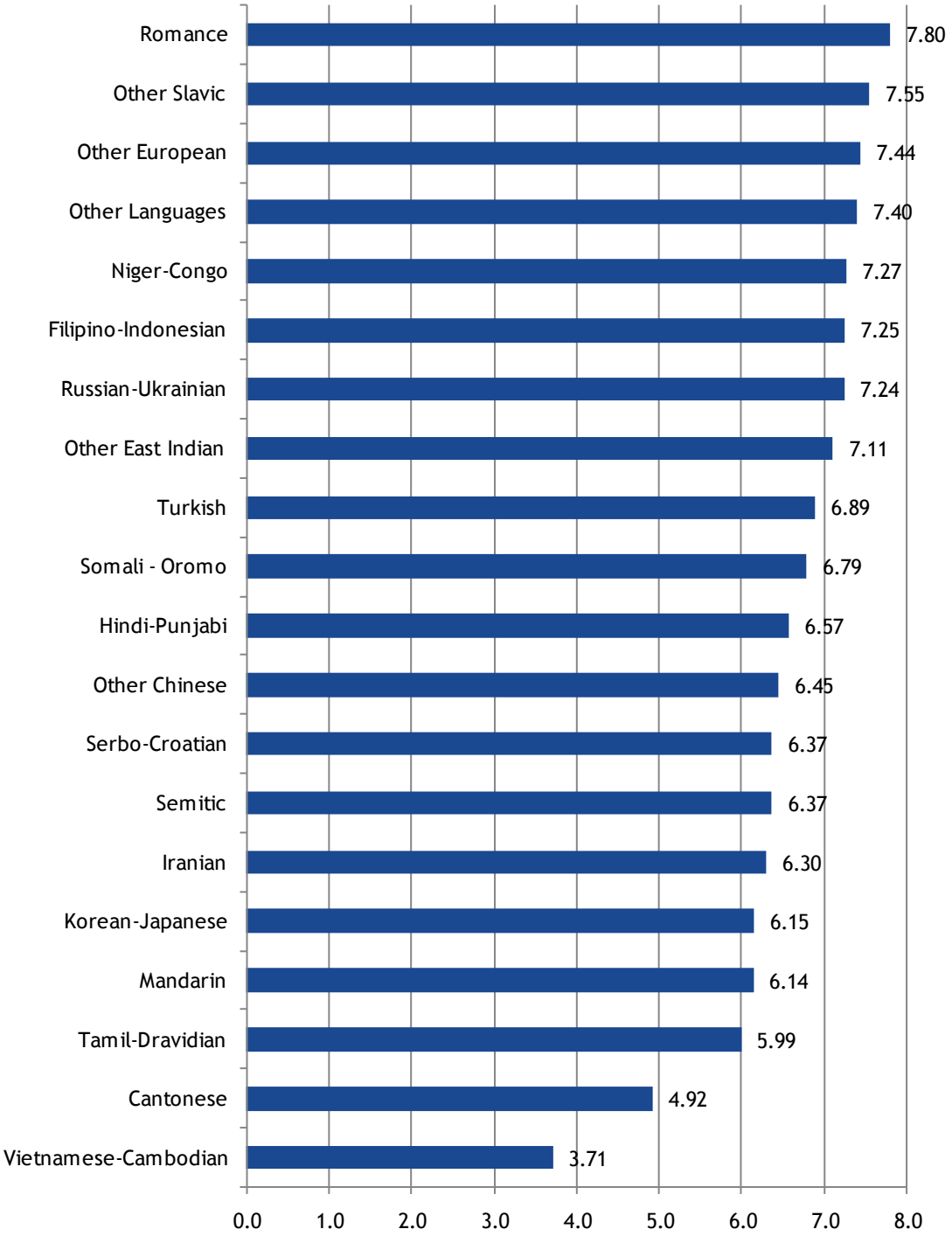
Table 3: Mean CLBA scores by geographic region.

Geographic Region	Example Countries *(not the whole set)*	N	CLBA Score
South Pacific	Fiji	48	7.4
Europe	Russia, Romania, Poland, Germany	547	7.3
Central/South America & Caribbean	Mexico, El Salvador, Columbia, Cuba	202	7.1
Africa	Somali Republic, Ethiopia, Ghana, Nigeria	216	6.9
Countries with English as an official language	United Kingdom, United States, Jamaica,	71	6.7
South Asia	India, Pakistan, Sri Lanka, Bangladesh	727	6.5
Middle East	Afghanistan, Iraq, Lebanon, Israel	452	6.4
East Asia	China, Philippines, Hong Kong, Taiwan, Vietnam	1,559	6.1

First language

A factor closely related to country of origin that is more germane to this analysis is mother tongue, or first language. As noted earlier, a very large number of first languages were represented in the sample. However, small Ns in many cases do not permit us to analyze each language separately. For this reason, we reclassified the languages into 20 major categories based on the languages most frequently represented in the sample, genetic relationships among languages, and geographical region. This resulted in 20 categories, summarized in Figure 2.

Figure 2: Mean CLBA score by language group.



When we examine these CLBA scores, we see that the extremes are for speakers of Romance languages with a mean of 7.8 versus Vietnamese/Cambodian with a mean of 3.7. This indicates a wide range of proficiency across these groups. There may well be a variety of influences here. Typological similarity, for instance, may explain why speakers of Romance and other European languages scored high, whereas speakers of many non-Indo-European languages scored much lower. However, this factor cannot fully explain the outcome. In particular, Filipino/Indonesian speakers

were among the highest scoring, a finding which may be related to the significant presence of English in such areas as the Philippines and the Pacific Islands. Other factors that may have affected language proficiency include degree of cultural distance, size of compatriot communities in Canadian cities, language teaching methods in the countries of origin and differences in overall levels of proficiency on arrival.

To gain insight into differences among first languages, we examined more closely the CLBA scores for the most frequent five mother tongue categories represented in the sample. Here we find the highest mean score in the Tagalog group (N = 281, M = 7.2), followed by Arabic (N = 222, M = 6.5), Mandarin (N =575, M = 6.1), Punjabi (N = 248, M = 6.0), and Cantonese (N = 289, M = 4.9). An analysis of variance revealed a significant effect of first language on CLBA scores, $F(4, 1610) = 36.1$, $p < .001$. Post hoc Bonferroni t-tests ($p < .01$) indicated that the Cantonese speakers scored significantly lower than any other group, whereas the Tagalog speakers scored significantly higher than all other groups. The other three groups did not differ significantly from each other. The relatively low language scores of the Mandarin and Cantonese speakers are especially striking given that these groups have more representation in the independent immigration class than do the other top 5 mother tongues (79.2% for Mandarin and 53.4% for Cantonese), which suggests a high level of education based on immigrant selection criteria. In fact, when we examine the whole data set, we find that independent class immigrants scored significantly higher (M = 6.8) on the CLBA measures than did either family (M = 6.3) or refugee (M = 6.0) classes, which also differed significantly from each other (Bonferroni, $p < .05$). Thus the scores of the Mandarin and Cantonese speakers, who comprise a large component of the sample, appear to be inconsistent with other members of the independent class.

It is not possible to draw any firm conclusions about why the Mandarin and Cantonese speakers seem to be at such a disadvantage. Apart from the fact that these languages are typologically distinct from English, educational practices and cultural factors in China and Hong Kong may well play a role. Future research should examine the types of English language training received by these groups before their arrival in Canada to determine how to best meet their oral language needs after they immigrate. It is unfortunate that the reading and writing CLBA scores were not done for these groups because we expect that the Chinese speakers would show greater proficiency in these skill areas than in terms of oral skills. Still another area that merits careful attention is the degree of exposure to English on a daily basis after arrival in Canada. Previous research (Derwing, Munro & Thomson, 2007) has indicated that Mandarin-speaking immigrants to Canada interact less frequently with English speakers than do immigrants from Slavic backgrounds, a tendency that may partially explain their slower oral language development.

Both Hong Kong and Taiwan have experienced more western influence than China. For this reason we undertook a further analysis comparing CLBA scores across participants from China, Hong Kong, and Taiwan. There were no statistically significant differences, in spite of a tendency for the speakers from China to score lower than those in the Hong Kong and Taiwanese groups.

Overall, a significant effect of sex was found, such that the mean CLBA score for males (6.7) was higher than that for females (6.4), t (3825) = 3.49, $p < .001$. It is obvious that a wide range of other variables interact with sex in determining language proficiency. For instance, it is well known that high academic achievement predicts higher attainment in a second language (Gardner, Polyzoi & Rampaul, 1996) and that education level can often differ between the sexes. However, we do not have information about the participants' previous educational experiences before arrival in Canada. We also compared male and female CLBA scores for each of the 20 language categories as shown in Table 4.

Table 4: Mean CLBA scores according to sex and mother tongue.

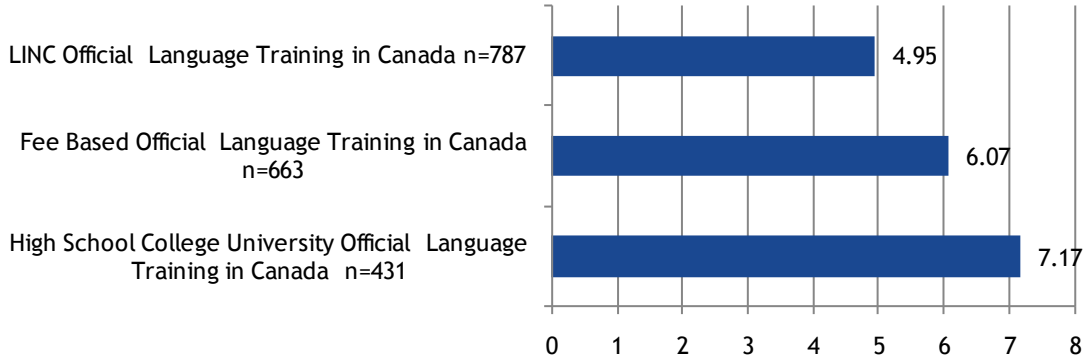
Language	*M* for Females	*M* for Males	Significance (*t*-test)
Cantonese	4.8	5.0	ns
Filipino/Indonesian	7.6	6.6	$p < .001$
Hindi/Punjabi	6.2	7.0	$p = .002$
Iranian	6.2	6.4	ns
Korean/Japanese	6.3	6.0	ns
Mandarin	6.1	6.2	ns
Niger/Congo	7.1	7.4	ns
Romance	7.9	7.7	ns
Russian/Ukrainian	7.0	7.6	$p = .026$
Semitic	5.8	6.9	$p < .001$
Serbo-Croatian	6.5	6.3	ns
Somali/Oromo	6.7	6.9	ns
Tamil/Dravidian	5.7	6.3	ns
Turkic	7.2	6.5	ns
Vietnamese/Cambodian	3.3	4.6	$p = .064$
Other Chinese	6.4	6.6	ns
Other East Indian	6.7	7.5	$p = .005$
Other European	7.6	7.3	ns
Other Slavic	7.6	7.5	ns
Other Languages	7.6	7.2	ns

For several countries, there were no significant differences in language proficiency in English between males and females, but we note that in five instances, females scored significantly lower than their male counterparts. In one instance, the Filipino/Indonesian group, females scored significantly higher than males. We can only surmise that these differences are related to educational practices in the home countries.

Language training in Canada

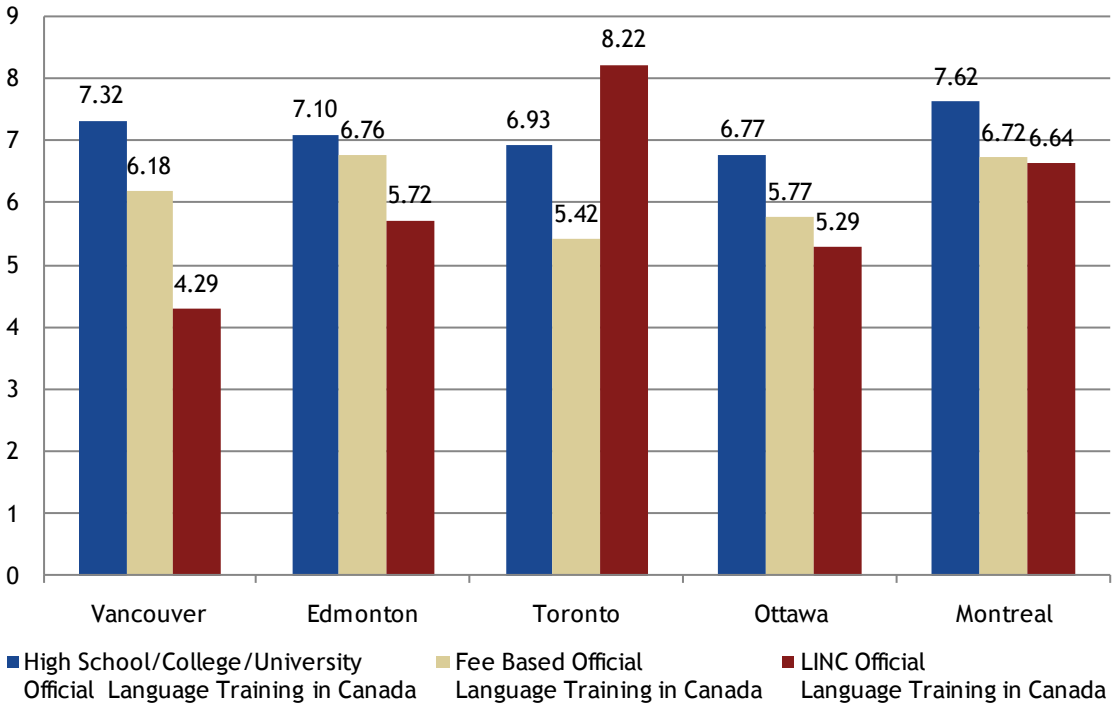
In the survey, three different sources of language training were specified: Language Instruction for Newcomers to Canada (LINC), fee-based official language training, and high school/college/university official language training. CLBA scores across the three sources differed significantly, $F(2,1848) = 171.2$, $p < .001$. Bonferroni *t*-tests showed that the scores in each training category were significantly different from the others: the high school/college/university category registered the highest mean score (7.2), followed by the fee-based training category (6.1) and LINC (5.0) (see Figure 3). This is not altogether surprising, since LINC is intended for low proficiency learners, and many other language providers design their programs to dovetail with LINC to avoid duplication of services. Thus the other language source categories tend to cater to higher proficiency learners.

Figure 3: Mean CLBA scores by source of language training.



Important differences emerged in the relationship between type of language training and CLBA scores when we looked at data for individual cities across Canada. In Vancouver, the pattern followed the overall Canadian data described above. In Edmonton, LINC CLBA scores were significantly lower than both fee-based and high school/college/university scores; however, the latter two categories did not differ from each other. In Ottawa, CLBA fee-based scores did not differ from LINC scores. Finally, in Montreal, the high school/college/university scores were significantly higher than those of the fee-based programs; the small number of LINC cases (N = 11) precluded any statistical analysis. The disparities across cities may reflect differences in the upper limits of LINC at the time of the study, which were lowest in British Columbia at level 3, followed by Alberta at level 4, and by Ontario at level 5. Thus, there was potentially more overlap in curriculum coverage in LINC and fee-based programs in Ontario. Figure 4 shows CLBA scores for five Canadian cities according to language training source.

Figure 4: Mean CLBA scores for five Canadian cities by language training source.



An additional examination of CLBA scores for participants enrolled in full-time versus part-time language training revealed slightly higher scores for the former group (M = 6.34 vs 6.06). However, this difference was not statistically significant (p = .07). There was a slight tendency for refugees to attend full-time more (68.2%) than members of the family (57.7%) or independent (61.0%) classes. However, this trend was not statistically significant.

Formal education in Canada

Table 5 provides mean CLBA scores according to participants' formal education in Canada. While a significant difference in scores was observed across the categories [$F(5, 1299) = 6.90$, $p < .001$], post hoc Bonferroni tests revealed that this was due to significantly higher CLBA scores for those reporting a university or college diploma as compared with all other categories except apprenticeship training.

Table 5: Mean CLBA scores according to level of formal education in Canada.

Level of Formal Education	N	Mean CLBA Score
University or College Diploma	543	7.6
Apprenticeship	72	7.6
Computer Training	134	7.1
Continuing Education	274	6.9
High School	144	7.1
Other	138	7.0
Total	1,305	7.3

Occupation in Canada

We examined the NOC codes and determined that it would be most informative to run an analysis of CLBA scores against the degree of expertise required for various occupations. In an examination of the most recent occupation data, we reclassified all of the occupations into three categories of job skills: Highly skilled, professional and senior management (N = 697); Skilled, technical and middle management (N = 1453); and Low/unskilled, assisting occupations (N = 817). The mean CLBA scores for these groups differed significantly across the categories, $F(2, 2964) = 42.3$, $p < .001$. The scores for highly skilled professional and senior management were significantly higher than the scores of the other two groups, Bonferroni, $p < .05$. There was no significant difference in the latter groups' scores. We also considered the most frequently occurring recent occupations. In Table 6, the occupations in which at least 25 immigrants worked are listed, along with CLBA scores in rank order. From this information we see a wide range of average CLBA scores, from 8.48 for registered nurses, to 3.0 for sewing machine operators. As expected, there is a clear relationship between the communication demands at work and the participants' CLBA scores. It is interesting to note, for example, that bakers (5.0) and cooks (5.0) had lower language skills than food service counter attendants/preparers (6.1). At first glance one might assume that the skill level of these occupations would be similar, but more oral interaction is required at the service counter, which requires a higher level of language proficiency.

Because the survey question on longest held occupation exhibited a very similar pattern of responses, there is little value in providing a separate analysis. It is regrettable that there was no question in the survey about occupation in country of origin. Thus we are unable to determine what percentage of these individuals have reentered their professions, and to what degree language skills have played a role in whether or not they have done so.

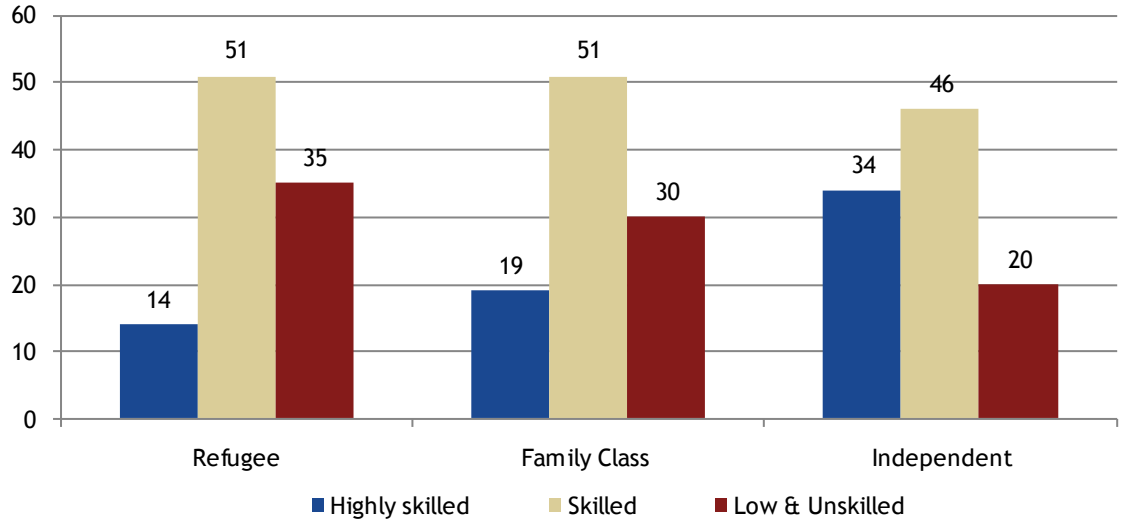
Table 6: CLBA mean scores and N's for occupations with over 25 participants reporting.

Most Recent Occupation	N	CLBA Score
Registered Nurses	29	8.48
Computer Systems Analysts	30	8.23
Financial Auditors and Accountants	52	7.94
Web Designers and Developers	55	7.82
Visiting Homemakers, Housekeepers & Related Occupations	25	7.80
Computer Programmers	52	7.70
Post-Secondary Teaching & Research Assistants	30	7.67
Customer Service, Information & Related Clerks	44	7.57
Retail Trade Managers	28	7.43
Self-Employed (no specification)	28	7.20
Retail Sales Person, Sales Clerk	168	7.14
Community and Social Service Workers	27	7.11
General Office Clerks	36	7.08
Shippers and Receivers	35	6.91
Nurse Aides and Orderlies	44	6.82
Security Guards and Related Occupations	34	6.82
Cashiers	98	6.57
Babysitters, Nannies and Parents' Helpers	44	6.52
Business/Business Owners (no specification)	105	6.50
Restaurant and Food Service Managers	49	6.35
Truck Drivers	46	6.30
Material Handlers	26	6.23
Food Service Counter Attendants & Food Preparers	26	6.08
Food and Beverage Servers	70	5.99
Kitchen and Food Service Helpers	33	5.70
Mechanical Assemblers and Inspectors	48	5.40
Janitors, Caretakers and Building Superintendents	33	5.36
Light Duty Cleaners	74	5.32
Construction Trades Helpers and Labourers	44	5.07
Bakers	25	5.04
Cooks	74	5.03
Other Labourers in Processing, Manufacturing & Utilities	67	4.96
Sewing Machine Operators	28	3.00
Total	**1,607**	**6.49**

We performed one further analysis, in which we examined job skill level according to immigration class. As noted earlier, independent class immigrants exhibit significantly higher CLBA scores than the family and refugee classes. However, this does not appear to afford as much of an advantage as

might be expected with respect to the jobs they have obtained. Figure 5 shows that within the independent group there is a larger representation of individuals working in highly skilled positions than in the other groups. Nevertheless, nearly two thirds (66%) had jobs in either the middle skilled or low skilled categories, even though most were selected to come to Canada, in part, on the basis of their superior job skills and formal education.

Figure 5: Job skill level by immigration class.



Although participants were asked whether they were currently employed, it is not possible to tell whether those who answered in the negative were actually in the labour market. Nonetheless, we have carried out an additional analysis of CLBA scores comparing employed versus not employed participants, excluding from the latter category those individuals who reported having dependents at home. This permits us to gain a rough idea of how employment status relates to language proficiency. We found a significantly higher mean CLBA score in the employed group ($M = 6.78$) compared to the not employed group ($M = 6.19$), $t\,(1728) = 6.77$, $p < .001$.

Language at work

As noted above, 85% of respondents identified an official language as the one they used most often at work. However, this information is limited in that it does not reveal how much people are required to communicate at work or for which audiences. Nor does it indicate the extent to which the respondents might use more than one language in their workplaces. For example, a sales person might use one language with customers and another with coworkers. What we can surmise is that lower CLBA scores are probably associated with less use of English or French at work. For example, sewing machine operators have little opportunity to interact with others in their jobs, whereas good communication skills are a bona fide qualification for customer service employees. Given the many complexities associated with language use at work, attempting to extract further information from these data is not feasible.

Citizenship test scores and CLBA scores

CLBA scores and citizenship test scores exhibited a small but statistically significant Pearson correlation ($r = .391$, $p < .05$). It is unsurprising that language skills should be somewhat predictive of performance on a written test. The effect of language skills on actual pass rates, however, appears to be minimal, given that there were no statistically significant differences among native and non-native speakers of official languages, as noted earlier. Nonetheless, immigration category did exert a significant effect on citizenship test scores, $F\,(2, 1930) = 81.74$, $p < .001$, such that refugees scored

significantly lower (M = 17.34/20) than family class immigrants (M = 18.11), who, in turn, scored lower than members of the independent class (M = 19.16).

Differences across cities

Table 7 presents mean CLBA scores across cities. Significant differences were observed between the cities, F (5, 3821) = 37.1, p < .001. Post hoc Bonferroni tests revealed that the scores for Montreal were significantly higher than for all other cities. This finding might be partially explained by the level of language skills of participants on arrival. In Montreal, only 7% of cases who reported English language training took LINC classes (designed for beginner level or lower proficiency speakers of English). In contrast, the percentage in Edmonton was 46. This indicates that, in Montreal, a greater proportion of individuals had some proficiency in English before arriving in Canada. The average CLBA scores from Edmonton were significantly higher than those of the other English-speaking cities. In part, these differences may be explained by the relative proportions of refugee, family and independent class immigrants represented in each city's data. Whereas Montreal has a majority of independent class respondents (61%), thus increasing the overall CLBA mean, Ottawa has a disproportionately large representation of refugees (27%) and Edmonton's refugee percentage (14%) is the smallest of the five cities involved.

Table 7: Mean CLBA scores by city.

City	N	Mean
Greater Vancouver	1,369	6.3
Greater Toronto	801	6.2
Montreal	597	7.4
Ottawa	531	6.3
Edmonton	520	6.9
Halifax	N	N

Additional correlational and regression analyses

Simple correlational analyses

Additional correlational analyses were carried out to investigate relationships not discussed above. Table 8 provides intercorrelations (Pearson *r*) among a number of continuous and nominal variables of interest. Among the most noteworthy findings, a significant negative correlation was observed between CLBA scores and age of immigration. This outcome is consistent with findings well-documented in the second language acquisition literature, which show that ultimate proficiency is closely related to age of second language learning (Long, 1990). In fact, participants' chronological ages (not shown in Table 8) were also significantly correlated with CLBA scores ($r = -.258$, $p < .01$), such that older individuals overall tended to have lower proficiency.

It is striking that there is no significant correlation between the number of years in Canada and CLBA scores. In fact, second language research literature indicates that length of residence effects on ultimate language attainment are typically much smaller than age of learning effects (Oyama, 1976). Furthermore, no significant relationship was observed between language training (in months) and CLBA scores. However, it is important to recognize that without a measure of language proficiency on arrival in Canada, it is not possible to assess the overall effect of language training. In addition, dramatic variations in the quality of instruction in language training programs adds considerable complication to the interpretation of these results.

Table 8: Intercorrelations among selected variables of interest.

	CLBA Score	Sex	Years	Lang. Source	Lang. P/F	Lang. Months	Educ. Level	ImAge	Citizen	ImCat
CLBA Score	1									
Sex	.056**	1								
Years in Canada	.025	-.036*	1							
Source Lang. Training	-.395**	-.068**	-.090**	1						
Lang Part/ Fulltime	.073	.042	.024	-.010	1					
Lang Months	.015	-.005	.057*	-.202**	.172**	1				
Education Level	.150**	-.016	-.096**	-.002	-.238**	-.185**	1			
Immigration Age	-.274**	.051**	-.200**	.341**	-.116**	-.149**	.128**	1		
Citizenship Test Score	.391*	.020	-.084**	-.107**	.059	-.047	.053	-.002	1	
Immigration Class	.149**	.005	-.041*	-.016	-.051	-.106**	.160**	.163**	.278**	1

** significant at the p<.01 level, *significant at the p<.05 level

Multiple regression analysis

As an exploratory investigation of the contributions of participant variables to CLBA scores we carried out a multiple regression analysis (pairwise), in which CLBA scores served as the dependent variable and a wide range of predictors were included. The purpose of this analysis was to identify the best combination of variables that would predict CLBA scores. This evaluation must be regarded as preliminary because of some difficulties with the data identified above. However, it may prove useful in providing directions for further work.

Total variance accounted for by the final model (Adjusted R Square) was 41.6%. A list of variables that contributed significantly to the regression is provided in Table 9, along with slope of the regression line, beta coefficients, *t*- scores and significance levels. LINC training (as compared to other language training), official language training in general, and fee-based training were all tied to lower CLBA scores. It is important not to interpret this outcome as an indication that LINC or other training has a negative impact on language proficiency. This finding is probably a result of the fact that participants who sought such training were likely to have lower or no language proficiency in English on arrival. Education in Canada was positively associated with CLBA scores. Level of education was also a contributor, because, as noted earlier, participants with university or college training scored considerably higher on the CLBA than did most other participants. Age of immigration continues to be negatively correlated with CLBA scores, even when other factors are taken into account. Immigration class was associated with CLBA scores because, as noted above, the independent class achieved higher average CLBA scores than either the family or refugee classes. Toronto residence tended to be associated with lower CLBA scores. This finding is worth investigating in future studies. There are several possible explanations, none of which can be confirmed with this data set. For example, the under-representation of eligible cases for the city of Toronto may have affected the results. Furthermore, because Toronto receives more immigrants than any other location in Canada, the existence of large compatriot communities may have had an impact on language learning. In some cases, it may not have been necessary for individuals to learn an official language in order to obtain employment or services. Finally, the results may be due to a complex interplay of demographic variables that affects Toronto differently than the other cities.

Two significant negative predictors of CLBA scores were 'East Asian Mother Tongue' and 'Southeast Asian Mother Tongue'. Overall, the former group scored significantly lower (M = 5.89) than other major language groups in the analysis. This group comprised speakers of Mandarin, Cantonese, Filipino/Indonesian, Other Chinese, Korean/Japanese, and Vietnamese/Cambodian. However, it should be noted that 69% of this group consisted of speakers of Mandarin, Cantonese, and Other Chinese. In fact, as noted earlier, the Tagalog speakers (from the Filipino/Indonesian grouping) scored significantly higher overall than the Chinese speakers. Participants from Southeast Asian languages (M = 6.41) also were at a disadvantage compared to the other language groups. Of the several Southeast Asian languages represented in the sample, the Tamil-Dravidian language speakers exhibited the lowest CLBA scores (M = 5.99).

Table 9: Final regression variables with significant effects.

Variable	Values	Slope	Beta	t score	Significance
LINC Training	Other training, LINC	-1.576	-0.335	-7.421	.000
Level of Education in Canada	High school, computer training, apprenticeship, university or college	0.395	0.263	6.738	.000
Official Language Training	no, yes	-1.218	-0.262	-7.781	.000
Education in Canada	no, yes	1.013	0.218	6.360	.000
East Asian Mother Tongue	no, yes	-1.098	-0.218	-3.639	.000
Age at Immigration	Years	-0.043	-0.167	-4.654	.000
Immigration Class	Refugee, family, economic	0.393	0.130	3.526	.000
Fee Based Language Training	Other training, fee based	-0.729	-0.149	-3.466	.001
Toronto Resident	no, yes	-0.722	-0.126	-2.792	.005
Southeast Asian Mother Tongue	no, yes	-0.755	-0.108	-2.310	.021

Other variables used in the analysis which did not contribute significantly: sex, number of years in Canada, Edmonton residence, Ottawa residence, Vancouver residence, African mother tongue, and Indo-European mother tongue.

Problems with the data set

The main problems that we have identified with the data set include the need for extensive data cleaning, and recording procedures, and finally, shortcomings of the survey questions themselves. We will deal with each of these problems separately.

Data cleaning

When we received the data in mid-January of 2007, we noted that the Toronto and Vancouver files had fewer questions than the files from the other cities, and that they were in SPSS format while the other files were in Excel. In order to make direct comparisons, all the data had to be merged into a single file, which required a great deal of data management. Considerable additional cleaning was also needed because of inconsistencies in data entry procedures. Furthermore, we recoded the occupation variables using the National Occupational Coding (NOC) system. These procedures were carried out between March and June of 2007. Statistical analyses were then performed on the resulting data set.

At the same time that we received the data, we also received documents entitled "CIC Language Surveys: Sample Development and Data Management" and "Description of Data Files." Although these documents provided us with some background details, they were of limited value in helping us understand some of the data collection procedures.

We have resolved, to the best of our ability, some of the problems arising from inadequate data cleaning. The data from this project, now combined into a single file with consistent coding, are available to CIC for further perusal. Two of the larger problems encountered in the cleaning phase were as follows:

- The data included incompatible Excel and SPSS files, requiring extensive manipulations prior to merging
- Extensive recoding had to be carried out. For example, there were nine different codes for sex! This was a relatively simple irregularity that had to be corrected. However other modifications were far more complex. Much of the trouble with the data set was due to faulty data collection procedures that will be discussed in the next section.

Non-response or missing data

There are high numbers of non-responses or missing data across many categories. Perhaps many of the participants did not understand particular questions and thus could not respond, but the numbers are so high that we suspect that there is assessor error involved here as well.

Inconsistent coding

Another assessor-related problem was the inconsistent coding of data throughout. Although some data cleaning is inevitable in this type of work, assessor training in this area might have eliminated the high degree of variability in the coding of responses, and would have cut down on the number of hours required to clean the data. For instance, the many spelling irregularities made it impossible to automatically convert string to numeric variables, and inconsistent recording of dates had to be rectified manually. Furthermore, there were instances of mixing of string and numeric coding across cities, which meant that city files could not be merged until recoding was carried out.

A related problem we encountered was the assignment of the same ID numbers to individuals in different cities. This apparently happened because participants were assigned IDs at the testing sites rather than through a more centralized process.

Lack of probing

Some responses to questions indicated inadequate probing on the part of the assessors. For example, responses to the question about current occupation included such things as 'works at the Bay', 'technician', and 'owner'. These vague descriptors cannot be interpreted or classified into the NOC system. For example, an employee at the Bay could work in maintenance, food service, clerical, sales and service, or in management. An example of another question that would have benefited from more careful probing was language – responses included non-existent languages, e.g., Swiss (Switzerland has four official languages, none of which is properly called 'Swiss'), and multiple responses to questions that required a single answer, e.g., language used most often at work. Another type of inadequate probing was a failure to collect the full complement of information on questions with multiple parts. For example, many participants who reported having accessed language training did not provide information as to type of training.

Selection of participants

It is unclear why, in a study of second language acquisition, native speakers of English and French were surveyed so extensively. It appears that considerable resources were devoted to collecting data that would be of little value.

Shortcomings of the survey instrument

Insufficiently focused or ambiguous questions

In order to efficiently collect data that will be useful in planning better language programming, a focus should be placed on questions that are pertinent to the issues being investigated. In this study, it appeared as though a number of disparate areas were covered. Not only were demographic and language questions asked, but citizenship questions that appeared to have little relevance to matters of language learning were also included. It is unclear why data were collected on citizenship matters (e.g., name of judge) in a study that was ostensibly conducted to gauge language development, particularly when Citizenship and Immigration Canada has other records on citizenship pass rates, etc. It would be a better use of resources to restrict the survey to questions that are directly relevant to the purpose of the study. Adding more to the study increases time, demands on participants and assessors, and overall cost.

Another example of an insufficiently focused question concerns language at work. Participants were asked which language they use most frequently, but there is no indication as to the nature of language use, for instance, the types of tasks required at work such as using formulaic language (e.g., a waitress uses the same phrases over and over) answering the phone, reading on the job, interacting with co-workers, training others, making formal oral presentations, writing correspondence and reports, etc.

Because of the way the question about current employment status was worded, it is not possible to determine which participants were unemployed versus not working by choice. This question could have been worded differently to elicit more useful information.

Some questions were ambiguous, such that participants may have had difficulty knowing how to interpret them. For example, when queried about language training, participants were asked whether they took LINC, fee-based, or high school/college/university courses. The latter category and fee-based instruction are not mutually exclusive. Moreover, LINC is sometimes offered at institutions that identify themselves as colleges. Concepts such as full-time versus part-time training and 'continuing education' are also highly problematic because of the wide range of interpretations that can apply to these terms. The only useful measure of amount of language training is number of hours of contact. Part-time attendance could entail a very small number of hours, or a very large number of hours per week.

Inadequate participant background information

On the other hand, there were insufficient questions about participants' experiences before coming to Canada, such as previous education and prior occupation. Furthermore, the participants who reported having received language training should have been asked about their CLBA score when they were originally tested. Although some individuals might not have accurately recalled their score, most would be able to provide helpful information. The lack of this information makes it impossible to assess actual language progress among the participant group after their arrival in Canada. Had these types of questions been included, the usefulness of the data set would have been greatly enhanced, and stronger conclusions could have been reached regarding the effectiveness of the language training the participants received.

Changes to survey midstream

The data collection in Toronto and Vancouver was completed prior to the data collection in the other four cities. Subsequently, changes were made to the questionnaire, resulting in incompatible data files. In order to make direct comparisons across all cities, it was necessary for us to exclude some of the information that was added midstream in the data collection process.

Summary statement and recommendations

Summary statement

An examination of data of the type employed in this study has the potential to yield useful, although incomplete, information about the language learning outcomes of adult immigrants. Our analysis has identified a number of ways in which the research design and data collection procedures can be improved. The data presented here reveal a variety of factors that predict participant CLBA scores. In fact, a regression model which included ten factors predicted over 41% of the variance in the scores. Given the complex array of influences on language learning, this is an unusually high multiple correlation. Among the statistically significant predictors were language training, education in Canada, age at immigration, immigration class, and city of residence. Perhaps the most noteworthy finding of the study was the effect of mother tongue. Members of both East Asian (with the exception of Tagalog speakers) and Southeast Asian language categories appear to have been disadvantaged relative to other language groups. This finding points to the need for language training that is targeted to listening and speaking skills for members of these language groups, especially since they comprise the largest cohorts of newcomers to Canada.

Although the analysis revealed a negative relationship between language training in Canada and CLBA scores, it must be recognized that there is no basis for believing that this is a causal relationship. Rather, those individuals who accessed language training most likely did so because they entered Canada with limited or no official language skills. This expectation is confirmed by the fact that LINC training, in particular, designed for beginners and low proficiency learners, showed the largest negative contribution to CLBA scores. The finding of a negative relationship between age of immigration and CLBA scores supports previous empirical evidence that ultimate attainment in a second language is affected by developmental factors, such that younger learners have an advantage. Another finding that is not altogether unexpected is the differences in CLBA scores across the three immigrant classes and the resulting significant contribution of immigration class to the multiple regression model. The higher scores for members of the independent class may be explained by selection criteria. Independent immigrants are chosen, in part, on the basis of their knowledge of an official language, as well as their formal education. Family class immigrants are not expected to meet the same stringent requirements, and because refugees, who received the lowest CLBA scores overall, are no longer selected on the basis of adaptability, prior language knowledge and formal education are not a consideration.

Recommendations

The following recommendations are based on our analysis of the limitations of the pilot study. We recommend that they be implemented in any further collection of data that CIC may undertake. These suggestions will lead, we believe, to more useful information for assessing the effectiveness of language training than is contained in the current data set.

1. Define a clear, sufficiently focused set of goals at the outset of the study.

The current study, as noted above, attempted to address several distinct issues at the expense of an adequate focus. For example, there is no reason to believe a priori that citizenship test scores and other citizenship related questions would contribute to the overall goal of the study to determine predictors of second language ability. A sharper focus on factors known to affect second language acquisition will lead to more meaningful outcomes in future studies. Although it may be tempting in work of this type to collect as much data as possible from participants, this approach can be counterproductive in that lengthy interviews increase the likelihood of assessor and participant fatigue as well as erroneous and missing data.

To ensure a maximally useful final data set, participant selection procedures must be consistent across the entire study. Evidence from "CIC Language Surveys: Sample Development and Data Management" (Government Consulting Services) suggests that this was not the case in the current data set. It is also clear that the questions in the survey were not finalized at the start of the study; after data collection was completed in Toronto and Vancouver, significant changes were made to the survey.

2. Develop and implement a clear training protocol for assessors and coordinators.

Any future data collection should entail face-to-face training workshops for assessors and others associated with the project. In addition, a training handbook should be prepared in which potential problems are anticipated and dealt with. It would be practical to employ a full-time trainer/coordinator to oversee the project. This individual could conduct the training sessions in all participating cities, observe assessors in their initial interviews, and debrief them regularly. The investment in training would add value, not only by eliminating the need for extensive and expensive data cleaning, but by resolving problems that might otherwise result in missing or uninterpretable data.

3. Streamline and improve consistency of data entry procedures.

An essential step in a large quantitative study of this type is the development of a codebook for data entry. To reduce the margin of error, and to facilitate analysis, all variables should be coded numerically. If future data are collected in the same manner, data entry should be centralized to ensure consistent recording. However, an alternative approach, which we believe to be superior, is the implementation of web-based data input requiring users to select appropriate responses from menus of choices. This type of data recording will minimize the potential for error and missing data. The assessors could enter the data directly, eliminating the need for separate data entry personnel.

In the future, occupations should be coded according to the NOC protocol.

Response codes should be included indicating whether the respondent did not understand the question or did not know, or the question was not applicable. This would be more informative than a blank response.

4. Collect more information on factors known to be relevant to second language acquisition.

The single, most important factor not included in this study was the CLBA score of participants when they were first assessed in Canada. Without this information, it is impossible to quantify the effects of language training or any linguistic progress that the participants may have made. There are many factors that influence language development, but without a sense of the participants' starting points, the relative contributions of each cannot be evaluated. If data sources from the original assessments are not available, participants should be asked to report their first CLBA scores.

Two additional variables that are known to be related to language learning are formal education and language training in the home country. Without knowing what resources the individual brings with him/her on arrival, it is very difficult to account for patterns and degree of language development. It is well established that formal education is positively correlated with second language proficiency. With regard to language training in the country of origin, it would be useful to know what type of instruction the participants received both to understand what skills they have already acquired and to identify potential gaps (e.g., oral/listening skills, pragmatics, reading/writing skills, grammar). The fact that many English as a foreign language programs focus almost exclusively on reading and writing development often leads to limited oral skills in individuals who may have a relatively good command of vocabulary and grammar.

Although occupation in home country usually is not directly related to language acquisition, such information would be useful because participants' personal linguistic goals and motivation may be

tied to reentering their previous occupation. The linguistic requirements of those occupations may in some way determine the participants' ultimate attainment in their second language.

More detailed information about language used at work should be collected. Although it is interesting to know which language is used most frequently, without a clarification of the actual types of language use, there is no helpful information regarding opportunities for ongoing linguistic development. Participants should be asked to specify the extent to which they use routinized, formulaic language (e.g., a waitress in a restaurant) versus conversational language (e.g., with clients and coworkers). In addition, participants should report the extent to which they are required to read and write at work, as well as the level of complexity of the tasks they are expected to perform. Some of the language at work questions could be posed in scalar fashion, for instance, percentage of time an official language is spoken (0%, 10%, 20% … 100%). Furthermore, participants' use of an official language could be probed more extensively to elicit information about language use at home and in social interactions outside the workplace. The latter, in particular, could lead to a better understanding of degree of integration.

While the three-way breakdown of immigration class in the current study was informative, more detailed specification of the independent and refugee classes may assist in the development of targeted language programming that would better address the needs of particular cohorts within the three classes. In the refugee class, for example, the language learning needs of the various categories of refugees may differ.

Length of language training should be specified in hours. Any increase in the time required to respond to this question would be offset by the increase in the usefulness of the answer.

Participants should be asked directly about the usefulness of their language training experiences. This could be achieved, for example, by having them respond on a scale to a question such as "How helpful have you found the language training you received?" 1 = not at all helpful, 5 = extremely helpful. They should also be asked to rate the degree of emphasis placed on particular language skills in the program(s) they attended (listening, speaking/pronunciation, reading, writing, grammar, vocabulary development).

Separate rather than combined assessments of listening and speaking should be used. Furthermore, the speaking component of the test should be digitally recorded for additional analysis. An informal consideration of the assessors' comments indicated that a large number of observations related to the participants' pronunciation. These comments could not be used in the current analysis because of a lack of standardization, but recordings could permit a thorough evaluation of pronunciation along with other speech variables such as fluency. Through appropriate analysis, data supplementing the CLBA test could be a source of extremely valuable information for researchers and curriculum developers. It would also allow for large scale cross-validation of this aspect of the CLBA assessment tool.

The CLBA test itself could be supplemented with an additional task, in which participants are asked to repeat a small number of utterances after a model (to be digitally recorded). From a methodological standpoint, this would permit an easy comparison across all participants because the content would be identical. Whether or not the participants could complete the task would also offer confirming evidence of their language proficiency.

5. ***Consider additional, complementary ways of evaluating the effectiveness of language training.***

Although a survey of this type can help pinpoint factors that influence second language learning, as well as ways in which current federally-funded language training can be improved, it provides only a limited perspective on these concerns. In order to gain a better understanding of a more complete range of issues that affect second language attainment, language programs themselves, including curricula, classroom practices, qualifications of instructors, quality and appropriateness of assessment

tools, and relevance of course content to learners' goals should also be studied. Without taking these factors into consideration, it would be difficult to make effective improvements to existing programs. One approach to addressing program efficacy would be to study innovative programs that report a high success rate, as measured by CLBA improvement.

6. Miscellaneous

We suggest limiting the assessor comments category to matters that concern highly unusual information about the participants. Many of the remarks in the current study were repetitions of information that had already been collected, or highly idiosyncratic information that was of little use.

References

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359-380.

Gardner, S. Polyzoi, E. & Rampaul, Y. (1996). Individual variables, literacy history, and ESL progress among Kurdish and Bosnian immigrants. *TESL Canada Journal*, 14 (1), 1-20.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285.

Krahn, H., Derwing, T. M., Mulder, M. & Wilkinson, L. (2000). Educated and Underemployed: Refugee integration into the Canadian Labour Market*, Journal of International Migration and Integration*, 1, 59-84.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 261-283.

Appendix A: Number of cases by country

Country	N	Country	N	Country	N
China	737	Nigeria	16	Uganda	3
India	386	Cuba	16	Rwanda	3
Philippines	320	Slovakia	15	Nicaragua	3
Iran	149	Malaysia	15	Nepal	3
Hong Kong	148	Bolivia	15	Libya	3
Pakistan	136	Singapore	14	Georgia	3
Taiwan	121	Sudan	13	Estonia	3
Afghanistan	100	South Africa	13	Costa Rica	3
Sri Lanka	93	Jordan	13	Austria	3
Romanic	82	Hungary	12	Algeria	3
Vietnam	66	Chile	12	Yemen	2
United Kingdom	65	Kenya	11	United Arab Emirates	2
Iraq	62	Czech Republic	11	Tunisia	2
Russia	60	Switzerland	10	Thailand	2
North Korea	57	Peru	10	Mauritania	2
Lebanon	56	Myanmar (Burma)	9	Haiti	2
Poland	55	Guatemala	9	Dominica	2
Ukraine	53	Venezuela	8	Azores	2
Somalia	53	Portugal	8	Zambia	1
Bangladesh	52	Morocco	8	Swaziland	1
Fiji	48	Kazakhstan	8	Seychelles	1
Mexico	41	Ecuador	8	Senegal	1
South Korea	40	Argentina	8	Panama	1
Bosnia & Herzegovina	34	Palestine	7	Norway	1
Yugoslavia Serbia & Montenegro	32	Belarus	7	Mongolia	1
Ethiopia	32	Tanzania	6	Moldova	1
El Salvador	30	Sierra Leone	6	Malawi	1
Israel	29	Saudi Arabia	6	Madagascar	1
Colombia	29	Greece	6	Luxembourg	1
Bulgaria	28	Honduras	5	Lithuania	1
Kuwait	26	Finland	5	Liberia	1
Croatia	25	Zimbabwe	4	Laos	1
Turkey	24	United States	4	Italy	1
Ghana	22	Sweden	4	France	1
Egypt	21	Malta	4	Djibouti	1
Albania	21	Macau	4	Cyprus	1
Japan	20	Eritrea	4	Brunei	1
Netherlands	19	Congo	4	Belgium	1
Germany	18	Cambodia/Kampuchea	4	Azerbaijan	1
Syria	17	Uzbekistan	3	Australia	1
				Armenia	1

Appendix B: Coding of countries in which English is an official language

Countries where English is an official language include:

Australia

Bahamas

Barbados

Bermuda

Canada

Jamaica

New Zealand

Saint Kitts and Nevis

Saint Lucia

Saint Vincent and the Grenadines

United States of America

United Kingdom: England, Scotland, Wales

Appendix C: Coding of countries by geographic region

South Asia includes:

Bangladesh, Bhutan, India, Maldives, Nepal, North Korea. Pakistan, Sri Lanka

East Asia includes:

Brunei, Cambodia/Kampuchea, China, Hong Kong, Indonesia, Japan, Laos, Malaysia, Mongolia, Myanmar/Burma, Philippines, Singapore, South Korea, Thailand, East Timor, Vietnam, Macau, Taiwan

Middle East includes:

Afghanistan, Armenia, Azerbaijan, Bahrain, Egypt, Georgia, Iran, Iraq, Israel, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Tajikistan, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Yemen, Palestine

Africa includes:

Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Côte d'Ivoire, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somali Republic, South Africa, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zambia, Zimbabwe

Central/South America and the Caribbean includes:

Belize (although English is the official language of Belize, most people native to that country speak Spanish as their mother tongue), Guatemala, Honduras, Mexico, Nicaragua, Costa Rica, Panama, Antigua and Barbuda, Cuba, Dominica, Dominican Republic, El Salvador, Grenada, Haiti, Trinidad and Tobago, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Surinam, Uruguay, Venezuela

South Pacific includes:

Fiji, Kiribati, Marshall Islands, Micronesia, Nauru, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu

Europe includes:

Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, USSR/Russia, San Marino,Yugoslavia, Serbia & Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, Vatican City, Northern Ireland, Azores

Appendix D: Coding of languages as originally reported

East Asian includes:

Cantonese, Chinese, Fukinese, Hakka, Japanese, Korean, Mandarin

Southeast Asian includes:

Bengali, Bijaiya, Gujarati, Hindi, Hindko, Kankani, Kashmiri, Konkani, Malayalam, Marathi, Nepali, Oriya, Punjabi, Sindhi, Sinhalese, Tamil, Telugu, Urdu

Indo-European includes:

Afrikaans, Albanian, Amharic, Arabic, Aramaic, Bosnian, Bulgarian, Cambodian, Cebuano, Chaldean, Croatian, Czech, Dari, Dutch, Estonian, Farsi, Filipino, Finnish, Flemish, French, Frysan, German, Greek, Hebrew, Kiligayon, Hungarian, Ilongo, Indonesian-Bahasa, Italian, Khmer, Kirdish, Lebanese(Arabic), Lithuanian, Macedonian, Malay, Maltese, Other European, Pampango, Pashto, Polish, Persian, Portuguese, Romanian, Russian, Serbian, Serbo-Croatian, Slovak, Slovene, Spanish, Sri Lankan, Swedish, "Swiss", Tagalog, Tigrinya, Ukrainian, Vietnamese, Visayan, Yiddish, Yugoslavian

African includes:

Aemavic, Akan, Alba, Bemba (Chibemba) Bini (Edo), Chichiwa, Efik, Ewe, Fanti-Fante (Akan) Foulla-Fula (Fulani) Ga, Ibo, Kikongo (Kongo), Kinyarwanda, Kirundi, Lingala, Oromo, Somali, Sukuma, Swahili, Swazai, Twi, Uhrobo, Wolof, Yoruba

Other languages includes:

Azeri (Azerbaijani), Burmese, Creole (Krio), Gad (Gaddang), Ilican (Turkic), Kakwa, Frio, Laotian,Taiwanese, Turkish, "other"

Appendix E: Classification of occupations into three levels of job skills

1. ***Senior Management, Professional and Highly Skilled Occupations:***

   Legislators and Senior Management; Professional Occupations in Business and Finance; Professional Occupations in Natural and Applied Science; Professional Occupations in Health; Professional Occupations in Social Science, Education, Government Services and Religion; Professional Occupations in Art and Culture, Trades and Skilled Transport and Equipment Operators; Skilled Occupations in Primary Industry; Processing, Manufacturing and Utilities Supervisors and Skilled Operators

2. ***Middle Management and Skilled Occupations:***

   Skilled Administrative and Business Occupations; Technical Occupations Related to Natural and Applied Sciences; Technical and Skilled Occupations in Health; Paraprofessional Occupations in Law, Social Services, Education and Religion; Technical and Skilled Occupations in Art, Culture, Recreation and Sport; Skilled Sales and Service Occupations; Intermediate Sales and Service Occupations; Intermediate Occupations in Transport, Equipment Operation, Installation and Maintenance; Intermediate Occupations in Primary Industry

3. ***Assisting, Low/Unskilled Occupations:***

   Occupations; Assisting Occupations in Support of Health Services; Elemental Sales and Service Occupations; Trades Helpers, Construction Labourers and Related Occupations; Labourers in Primary Industry; Labourers in Processing, Manufacturing and Utilities