Geographical distribution of Adult Literacy Skills in Canada based on local area estimates

J. Douglas Willms
Canadian Research Institute for Social Policy and
KSI Research International
with the assistance of
Richard Chan and **Teresa Tang**

Human Resources and Social Development Canada
Ressources humaines et Développement social Canada

Canada

Acknowledgements

We are grateful for advice on this project from Satya Brink, Scott Murray and Urvashi Dhawan-Biswal. The methods for mapping social outcomes used in this project developed over a period of several years with support from the Social Sciences and Humanities Research Council for Willms's position as the Canada Research Chair in Human Development at the University of New Brunswick, and for the research program, Raising and Levelling the Bar in Children's Cognitive, Behavioural and Health Outcomes. We are also thankful for support from the Atlantic Networks for Prevention Research and the Atlantic Rural Centre which have provided an infrastructure for conducting this kind of research in Atlantic Canada.

Table of Contents

Geographical distribution of Adult Literacy Skills in Canada based on local area estimates

1. The CRISP Method for Mapping Social Outcomes

The aim of this report is to provide maps of adult literacy skills for each of the provinces and territories in Canada and for its three largest cities. The data for the project come from the 2003 International Adult Literacy and Skills Survey (***IALSS***) conducted by Statistics Canada and the Organisation for Economic Cooperation and Development[1], and the 2001 Canadian Census. The project uses a mapping technique developed by the Canadian Research Institute for Social Policy over the past two years (Willms & Chan, 2005).

The aim of the CRISP mapping technique is to estimate a score on an outcome variable for all Canadian citizens, based on the best available information for each individual, and then display the resulting scores on provincial or local area maps. The approach uses the 2001 Canadian census data to create a file for each province that includes a "pseudo-record" for every individual in the province, based on the distribution of people by gender and age in each Dissemination-Area (DA). An estimate of a person's outcome (in this case a literacy score) for each person in the pseudo-record file is then estimated using multilevel multiple regression techniques, based on the following data: (a) information at the individual level from a Statistics Canada survey (in this case the ***IALSS***) about how well other people of the same age and gender scored in the person's DA, and in other DAs in their local area, and (b) information at the DA-level on the average outcome scores and the demographic characteristics of all DAs in the province. The work is carried out in several steps.

[1] Statistics Canada and Human Resources and Skills Development Canada (2005). Building on Our Competencies: Canadian Results of the International Adult Literacy and Skills Survey. Ottawa: No. Catalogue 89-617-XIE,

Step 1. Creation of provincial files of "pseudo-records".

The Census data include variables describing how many males and females in particular age ranges are within each DA. For example, variable V5 indicates the number of males aged 0 to 4, V6 the number of males aged 5 to 9, and so on through to V22 which is males 85 and older. This is followed by the comparable set of variables for females. These data are used to create an *individual-level* file for each province that includes the full population of people. For each DA the records of individuals have the age and sex distribution as described in the DA-level Census file. These files are called the provincial pseudo-files. For Newfoundland, for example, there are 1220 DAs, and the total population is 511,272. The Census file indicates that the first DA, number 10010001, has a total population of 210 people, comprised of 0 males aged 0 to 4, 5 males aged 5 to 9, and so on through to 5 females aged 75 to 79, 0 females aged 80 to 85, and 0 females aged 85 and older. Therefore, the individual level file has 210 records for DA 1001001, with three variables: DA, sex and age. Sex is coded 0 for males, 1 for females. The age variable is given a score in the middle of the category age range (i.e., 2 for age 0 to 4, 7 for age 5 to 9, etc.). The pseudo-file for Newfoundland has one record for each person, with a total of 511,272 records.

In total there are 13 pseudo-files, one for each province and territory. Each file has the same number of records as there are people in the province, as indicated by the Census.

Step 2. Construction of 3-level contiguity file.

Each DA has one or more other DAs that are contiguous with it (except for DAs that are small islands). These "neighbours" are called the level-1 contiguity neighbours. The first DA in Newfoundland, 10010001, has only one contiguous neighbour, 10010002. It is then determined which neighbours are contiguous to the level-1 contiguous neighbours. In this case, there are three: 10010003, 10010004, and 10010005. These are called level-2 contiguous neighbours. Going one more level out, we identify the level-3 contiguous neighbours, and in this case there are four of them: 10010006, 10010010, 10010470 and 10010471. These 9 DAs (the target DA and its neighbours) comprise what we call a DA-group. Our software constructs the DA-group or cluster for 10010001, and then proceeds to the next one, 10010002, and identifies its level 1, level 2 and level 3 contiguous neighbours. In this case there are 4 at level 1, 4 at level 2, and 11 at level 3. CRISP software completes this process for all 52,924 DAs in Canada to create our

“master contiguity file” (the creation of this file takes several weeks of computing time on a fairly fast computer). Thus, the master contiguity file has X records for each DA, where X is: (a) the number of level-0 neighbours (i.e., one, the DA itself), plus (b) the number of level-1 neighbours, plus (c) the number of level-2 neighbours, plus (d) the number of level 3 neighbours. Each record lists the DA, a neighbour, and the level of contiguity (0, 1, 2, or 3) of that neighbour. Overall, then, there is one DA-group (the DA and its neighbours) for each DA.

Step 3. Construction of census variables for estimating DA-level equation.

This step uses the Census data to construct DA-level variables that are used in the final analysis to estimate literacy scores. Several variables were constructed, but our model for adult literacy scores included 7 DA-level variables:

Years Education: The average number of years of education in the DA.

Transience: The percent of people that had moved in the previous 5 years.

Income: Average level of family income.

Unemployment Rate: Percent in the DA that were unemployed.

Percent Social Class 1 and 2: Percent in the DA who were in professional or semi-professional occupations.[2]

Percent Social Classes 4, 5, 6: Percent in unskilled labour occupations, and unclassified occupations.

Percent Recent Immigrants: Percent of people that had immigrated in the previous five years.

Imputation of missing data with local information. For most variables there is some missing data in the census file at the DA level. For example, many DAs do not have an estimate of average family income. For all DAs we estimated a "smoothed value". The smoothed value for a DA is the *weighted average of the scores for all DAs in its corresponding DA-group*. A DA's smoothed mean weights DA scores in its DA-group by the total population of each DA, and level of contiguity, with level 0 weighted 1.0, level 1 weighted 0.75, level 2 weighted 0.50, and level 3 weighted 0.25.

The income level for a DA is then set to its reported census value, if available. However, if the data are missing, it is set to its smoothed value.

Step 4. Construction of *IALSS* variables for estimating individual-level equation.

Thus far only data from the DA-level Census file have been used. ***IALSS*** data is now added to create an individual-level file that includes data on the following variables:

Province

DA

[2] Included occupations described in v989, v994, v1001, v1004, v1009, v1010, v990, v995, v1005, v1011, v1013, v1016, v1017 and v1027.

Survey Design Weight

Normalized Survey Design Weight. This is the population survey design weight, normalized to have an average of 1.0 at the provincial level.

Jackknife Replicate Weights. These are not used in this descriptive analysis, but they can be used later if we want to estimate the standard errors of each of our local area estimates.

Literacy Scores. "Plausible Value" are included for scores for prose, document, numeracy and problem-solving. In this report, only the five plausible values for prose literacy are used.

Age. Age of respondent.

Sex. Sex of respondent (coded 1.0 for females, 0 for males).

A separate version of this file is saved for each province.

For each province a DA-level file that indicates how many ***IALSS*** respondents were in each DA is also created.

Step 5. Construction of *IALSS* DA-group file.

The goal here is to construct a file that has the available ***IALSS*** records for each DA-group; that is for each DA and for each of its neighbouring DAs. It is easier to describe this file if just one DA-group is considered.

For example, consider a "target" DA that happened to have 4 ***IALSS*** respondents in it. The first four records in the DA-group file for that target DA would have the data from our ***IALSS*** individual file described in step 4. If the target DA had 3 level-1 contiguity neighbours, and these neighbouring DAs had, respectively, 3, 0, and 6 ***IALSS*** respondents in them, then our DA-group file for the target DA would have 9 more records in it, each with the data describing literacy scores and background variables from the individual ***IALSS*** file. This is done again for level-2 and for level 3, such that the file now has records for all respondents who were sampled from DAs in the DA group.

This process is repeated for all DA-groups in a province. Note that although some DAs did not have any ***IALSS*** respondents in them, their level 1 neighbouring DAs might, and in nearly all cases, there are some ***IALSS*** respondents when considering the DAs out to the third level of contiguity. Thus, nearly all DA-groups have ***IALSS*** respondents in them. To ensure

confidentiality, estimates that do not include at least five respondents in a DA group are not reported.

Step 6. Create master file for HLM analyses.

Next the file from step 5,is merged onto the census-level data from the DA-level file. Thus the micro-data file would now include for each DA-group:

The scores and covariates of all ***IALSS*** respondents in that DA;

The scores and covariates of all ***IALSS*** respondents in the neighbouring DAs; and

The census-level variables corresponding to each DA, attached to each record.

Step 7. Save Person-level and DA-group-level files for hierarchical model analysis.

The file created in (6) is taken and a person-level file is saved that includes information on the literacy scores, age and sex for each DA-group. Within a DA-group, people are weighted according to their level of contiguity. In this case, the variable for sex was centered on 0.5, and age was centered on age 45. This means that the estimated literacy scores for each DA will be the average of the expected scores for a male and female, both aged 45.

Recall that in step 6 the census variables were merged to the data for each individual. The DA-group level file has the average scores on these census variables for that group, weighted such that people in the target-DA are weighted 1.0 when computing the average for the DA-group, level-1 neighbours, 0.75, etc. Our DA-group-file has one record for each DA.

Step 8. Conduct HLM analysis

At this step an hierarchical linear model is fit to the data, with people nested within DA-groups. The within-DA-group model regresses prose literacy scores on sex, age, and age-squared for each DA-group. The DA-group level model regresses the intercepts from the within-DA-group level model on the 7 census variables (Years Education, Transience, etc.). This analysis yields a separate set of coefficients that comprise a prediction equation for each DA-group. In the case of Newfoundland, for example, where there are 1220 DAs, there are 1220 separate prediction equations. These analyses are carried out province by province.

Step 9. Predict scores for all people in each DA.

The pseudo-file created in step 1 is next used, and for each individual their predicted literacy score is estimated, based on the prediction models generated at step 8. Note that the prediction model was estimated locally, using survey information for all people in the DA group that actually did the survey, as well as more global information that pertains to how DA level characteristics are related to literacy scores at the DA level. The prediction equation also includes a random normal covariate, based on the results of the prediction model. This is added to each person's score.

This step is also very computationally intensive, as it is estimating the scores for about 30 million people, using results from nearly 50,000 separate regression equations. The analyses for most provinces can be done in a few hours, but the results for Ontario and Quebec require two to three days. In the end there is a person-level file with a predicted literacy score for each person.

Step 10. Mapping the data.

The final step uses the information to create a map for each province, displaying the data for each DA. The mapping software used, ARCVIEW (which is also used by Statistics Canada), requires the identification of fixed categories (e.g., about 8 to 10) for displaying in differing colours on the maps. To determine the cut-points the distribution of scores for all of Canada was examined using the final population pseudo-file, and attempted to make reasonable judgments on a variable by variable basis.

Although this approach is very computationally intensive, one of its strengths is that the data can be rolled up to higher levels, such as CSDs, cities, etc., based on our population pseudo-file results.

2. Maps for Prose Literacy

This report includes 180 maps plus two "posters" for all of Canada. The maps are organized in sets of 10, with a set for each of the 10 provinces, the 3 territories, and the 3 largest cities (Vancouver, Montreal, and Toronto). In addition, separate sets maps were produced for South-eastern Ontario and southern Quebec because of the high population densities in those areas.

The literacy scores for adults in the International Adult Literacy and Skills Survey are classified into five categories, based on people's levels of skills. The tasks required for each level are summarized below (see Statistics Canada and OECD, 2005, p. 17):

Level 1 "Most of the tasks in this level require the respondent to read relatively short text to locate a single piece of information which is identical to or synonymous with the information given in the question or directive.

Level 2 Some tasks in this level require respondents to locate a single piece of information in the text; however, several distractors or plausible but incorrect pieces of information may be present, or low-level inferences may be required. Other tasks require the respondent to integrate two or more pieces of information or to compare and contrast easily identifiable information based on a criterion provided in the question or directive.

Level 3 Tasks in this level tend to require respondents to make literal or synonymous matches between the text and information given in the task, or to make matches that require low-level inferences. Other tasks ask respondents to integrate information from dense or lengthy text that contains no organizational aids such as headings. Respondents may also be asked to generate a response based on information that can be easily identified in the text. Distracting information is present, but is not located near the correct information.

Level 4 These tasks require respondents to perform multiple-feature matches and to integrate or synthesize information from complex or lengthy passages. More complex inferences are needed to perform successfully. Conditional information is frequently present in tasks at this level and must be taken into consideration by the respondent.

Level 5 Some tasks in this level require the respondent to search for information in dense text which contains a number of plausible distractors. Others ask respondents to make high-level inferences or use specialized background knowledge. Some tasks ask respondents to contrast complex information." (Statistics Canada and OECD, 2005, p. 17).

Each set of maps includes the following maps:

1 Percent Level 1 Prose Literacy
2 Percent Level 2 Prose Literacy
3 Percent Levels 1 and 2 Prose Literacy
4 Percent Levels 3, 4 and 5 Prose Literacy
5 Percent Levels 4 and 5 Prose Literacy
6 Average Scores for Adults at Levels 3, 4 and 5
7 Number of Adults at Level 1 Prose Literacy
8 Number of Adults at Level 2 Prose Literacy
9 Number of Adults at Levels 1 and 2 Prose Literacy
10 Number of Adults at Levels 4 and 5 Prose Literacy

3. Interpreting the Maps

There are three basic types of maps in this report: maps showing the percentage of adults at particular levels of prose literacy, maps showing the number of adults at each level of prose literacy, and maps showing average scores on the prose literacy test. An understanding of the interpretation for these three maps enables one to interpret all of the maps.

Maps Showing Percentage of Adults at a Particular Levels of Prose Literacy

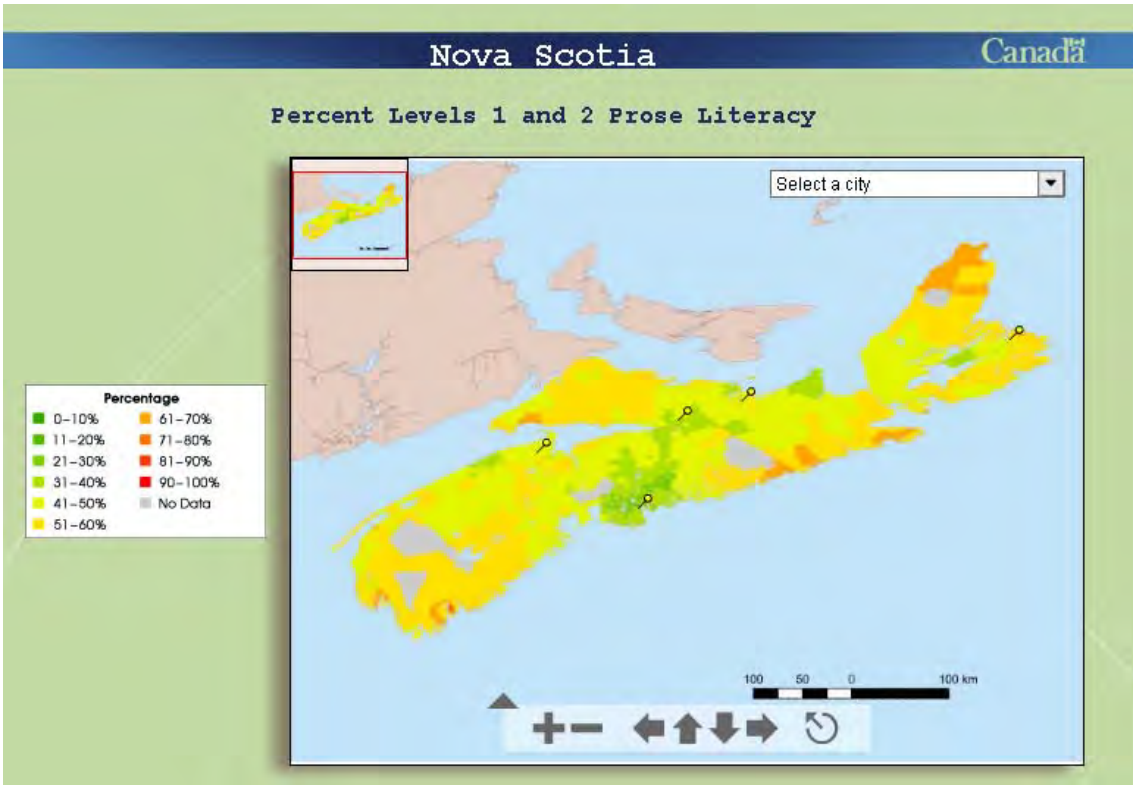

Figure 1. Percentage of Adults at Level 1 and 2 Prose Literacy in Nova Scotia

Figure 1 provides an example of a map depicting percentages of adults at levels 1 and 2 prose literacy. It shows this for Nova Scotia. The percentage-ranges for all maps showing percentages at levels 1 and 2 were set in 10 percentage-point categories (i.e., 0 – 10%, 10 – 20%, etc.). In these maps the colours range from dark green (generally considered a positive outcome – in this case, a low percentage of adults at the two lowest levels of literacy) through to dark red. Areas where there were no data available (or insufficient data to display results) are indicated in light

grey. For example, if one looks at the north-east tip of the province, the area is predominantly light orange, indicating that about 61 to 70% of the adults aged 16 to 65 in that area have literacy skills at Levels 1 and 2.

A cautionary remark is order. One cannot zone in on a particular Dissemination Area and say with certainty that there are, say, 60 to 70% of adults in that area with Level 1 and 2 prose literacy skills. The estimation technique estimates the result for a dissemination area based on the ***IALSS*** results of people in that area that did the survey, combined with results from people in the neighbouring dissemination areas. It also uses information on several characteristics of the area, measured at the DA level, to estimate the result. For example, imagine that only conducted a global regression analysis was conducted that modeled prose literacy scores on the adults' individual-level characteristics (i.e., age and sex) and the characteristics of their dissemination area (i. e., Years Education, Transience, Income ,Unemployment Rate, Percent Social Class 1 and 2, and Percent Social Classes 4, 5, 6) for the full weighted ***IALSS*** data set. One could then use those regression results (i.e., the unstandardised coefficients) to predict the literacy score for any adult in Canada, or for any dissemination area in Canada, based on the person's age and sex and the area were they lived. The results for an individual would not be very accurate, but the results for a DA would probably be quite reasonable. This approach essentially does this, but instead of a global all-Canada regression, it uses data for each local area to estimate the regression equation. Thus, it is much more accurate, but one should bear in mind that it is still a predicted value with a degree of uncertainty attributable to sampling and measurement error.

One way to think about this is to imagine placing a small drop of red paint on a blue sky while in the course of creating a water-color. The red droplet would "bleed" into the blue, creating a circle with colours that blended gradually from red through purple, and out to the blue sky. The data for the 23,038 respondents to the ***IALSS*** are like the droplets of red paint, but these range in colour according to people's actual performance on the literacy test. Also, the droplets are not deposited on a blue background; rather, they are placed on a background with a colour that reflects each person's local context, based on the economic and social characteristics of their neighbourhood as determined by information from the census. The colours bleed to form a picture that depicts the range of literacy skills throughout the province.

Maps Showing the Number of Adults at a Particular Level or Levels

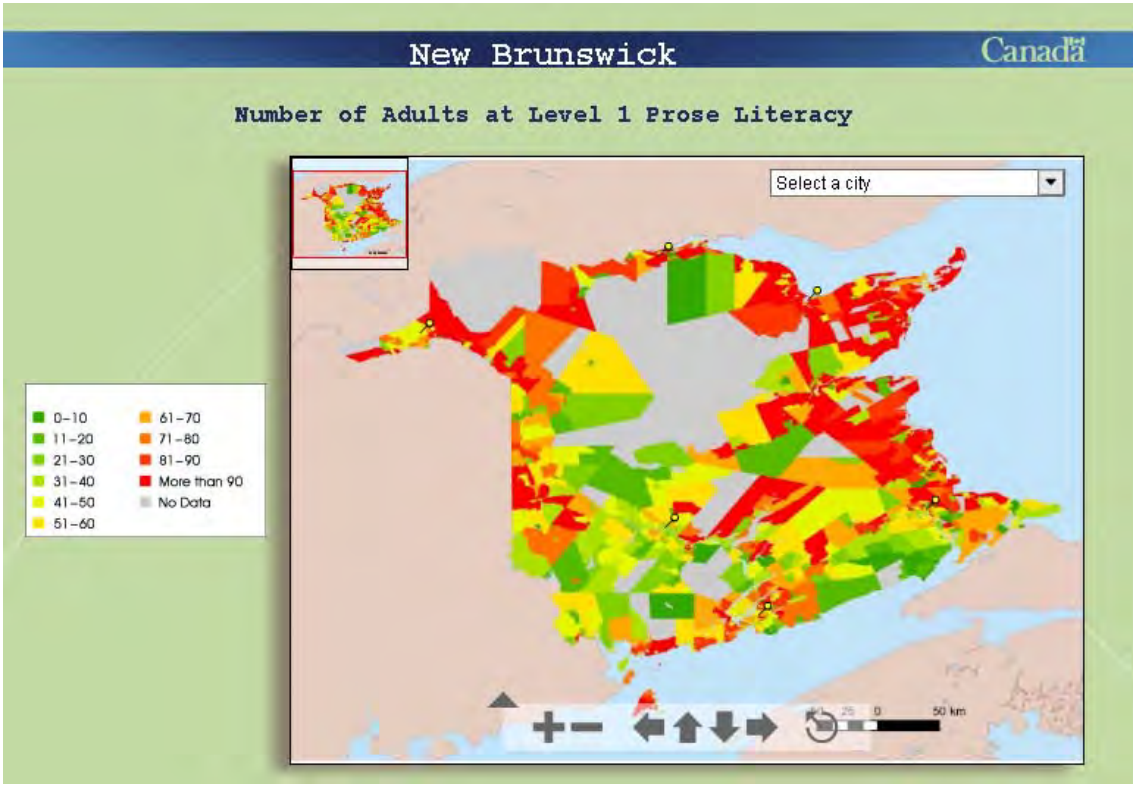

Figure 2. Number of Adults at Level 1 Prose Literacy in New Brunswick

In the course of preparing the maps based on percentages, it was discovered that their geographical presentation can be somewhat misleading, as one may mentally equate the area with numbers of people. However, the percentages pertain to the percentage of people living in a particular geographical area. For example, imagine a region where people with high literacy scores are concentrated in a large city, while those with low literacy skills are over-represented in rural areas. The map would be predominantly red, as the green areas would be mainly in the large city, which would not include much of the area of the map. It would give the impression that most people in the region had relatively low literacy skills even though the majority of people, who mostly live in the large city, have relatively high literacy skills.

Consequently, a set of maps were also produced that display the number of adults in the province or city with skills at a particular level. Figure 2 provides an example. It shows the number of people in each Dissemination Area in New Brunswick with literacy skills at Level 1 and 2. Like the map that shows the percentages of adults with low literacy skills, this map shows

that there are many adults in the rural areas with low skill levels. But if one zooms in on the three large cities, Fredericton, St. John, and Moncton, it also shows that there are areas within those cities with low literacy skills. Also, some of the rural areas have a very low population density, and these have relatively low numbers of people with low skills, even though the percentage may be relatively high. In many respects, these maps are the most useful for discerning where best to target particular programs.

Maps Showing the Average Literacy Scores

One can use exactly the same techniques described above to generate the average literacy score for each area, or the median score, or any descriptive statistic that is useful for policy purposes. A particular interest for Human Resources and Skills Development Canada was the average score for adults who had scored at Levels 3, 4, and 5. The map for Toronto, shown in Figure 3, illustrates this.

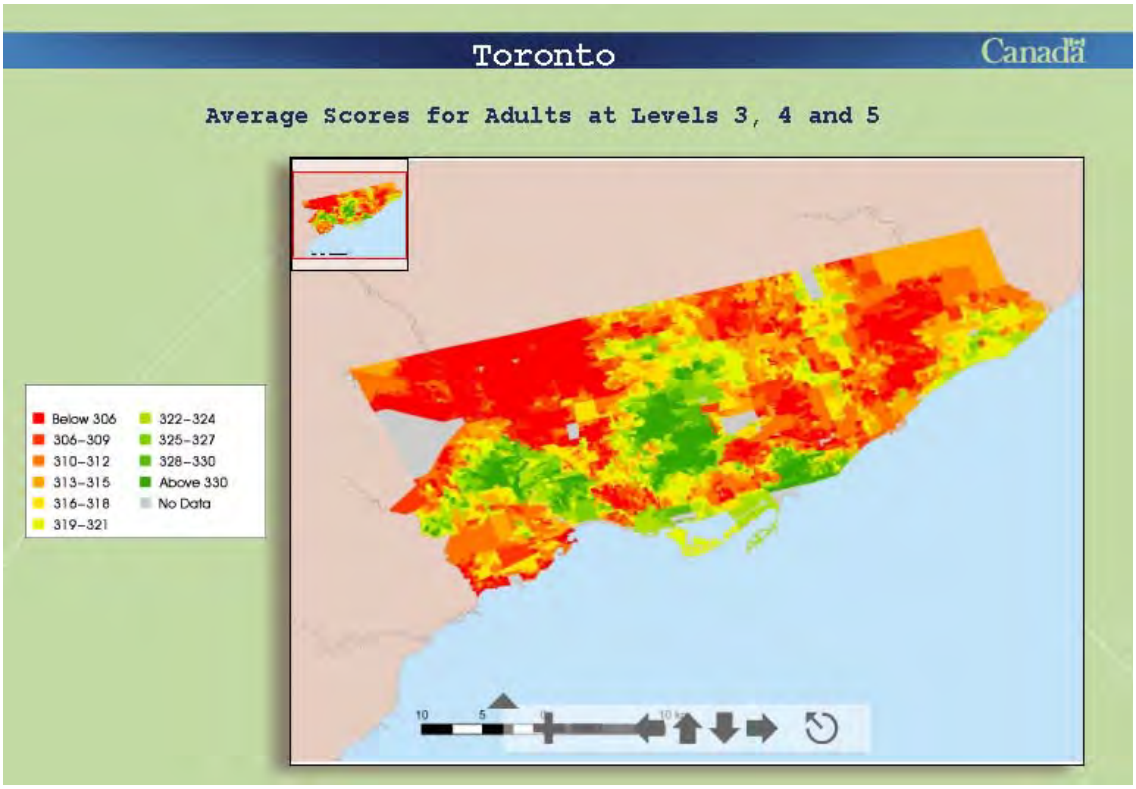

Figure 3. Average Prose Literacy Scores for Adults Scoring at Levels 3, 4 and 5 in Toronto.

In this case, red areas are places where the average score among adults scoring at the top three levels is quite low, below 306. The dark green areas are places where there is a high average level of scores among adults that scored at the top three levels. As this statistic is largely driven by the distribution of scores at the upper end of the distribution, one might think of the dark green areas as places where there is a preponderance of people with very high literacy scores.

Canadian Statistics

A performance indicator typically derives its meaning in one of three ways: by comparison to some fixed standard or set of standards, by trends over time, or through comparisons among jurisdictions. The ***IALSS*** allows for all three types of interpretation. First, the testing framework provides a means of categorizing people's scores into proficiency levels (i.e., Levels 1 to 5) that have meaning in that they are anchored to particular kinds of skills. The interpretations discussed above are based on the standards inherent in this classification, and require the person interpreting a map to be familiar with what is meant by scores at each proficiency level. The second kind of interpretation requires longitudinal data on individuals, or data describing the samples from the same population at two or more periods. The ***IALSS*** used the same tests and questionnaires (with minor modifications) such that one can compare results from one period to another.[3] Although this kind of analysis is not employed here, it would be feasible to do this using data that included Dissemination Areas for the 1994 and 2003 adult literacy data. The third approach requires an understanding of the amount of variation among jurisdictions. For the maps, this occurs naturally as one can see the extent of variation within a province based on the colour scheme of the map. However, it is also useful to see the results in the context of all of Canada. Table 1 shows the variation among provinces for each of the ten indicators shown in a set of maps. The analysis is conducted at the DA level, as it is the DAs that are displayed on the maps. However, the results are weighted by the number of adults in each DA, as it is these comparisons that are most pertinent to social policy.

[3] For example, see:
Willms, J. D. (2005). Skills, parental education and literacy practice in daily life. In Statistics Canada and OECD (Eds.), *Learning a Living: First Results of the Adult Literacy and Life Skills Survey (pp. 225-239)*. Ottawa and Paris: Statistics Canada and OECD.

Table 2. Descriptive Statistics for Adult Prose Literacy for Provinces and Territories

Province	Number DAs	Population Size (1000's)	% Level 1		% Level 2		% Levels 1 & 2		% Levels 3, 4, & 5		% Levels 4, & 5		Mean Score Levels 3, 4 & 5		Number Adults Level 1		Number Adults Level 2		Number Adults Levels 1 & 2		Number Adults Levels 4 & 5	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Newfoundland	920	511	18.5	9.6	33.0	5.1	51.5	13.1	48.5	13.1	16.3	7.8	315	6.3	102	62	193	96	294	149	100	71
Prince Edward Island	207	135	14.1	4.5	30.7	3.7	44.8	7.4	55.2	7.4	20.8	5.6	319	4.5	80	42	175	73	255	111	118	62
Nova Scotia	1276	906	11.4	5.7	29.2	6.6	40.6	11.7	59.4	11.7	23.3	9.8	320	7.0	67	40	175	79	242	114	141	97
New Brunswick	1223	728	16.6	8.4	34.1	5.1	50.7	12.4	49.3	12.4	15.7	7.2	314	5.9	87	59	179	79	265	130	84	59
Quebec	11558	7228	15.4	6.4	34.6	4.7	50.0	10.1	50.0	10.1	15.2	6.1	313	4.9	80	47	181	87	261	129	79	48
Ontario	17745	11378	15.8	11.0	30.8	6.8	46.6	15.7	53.4	15.7	19.6	10.6	317	8.3	93	111	185	194	278	293	116	134
Manitoba	2024	1117	12.9	7.7	30.6	6.1	43.5	12.8	56.5	12.8	21.0	9.1	318	6.9	59	49	140	79	199	124	95	63
Saskatchewan	2387	974	9.3	7.5	29.3	8.4	38.6	14.5	61.4	14.5	22.5	11.1	317	8.2	34	31	118	78	152	104	98	99
Alberta	4789	2970	11.5	5.4	28.5	5.2	40.0	9.9	60.0	9.9	24.2	7.8	321	5.8	68	49	179	126	248	172	160	135
British Columbia	6647	3897	12.2	7.5	27.3	6.4	39.6	13.2	60.4	13.2	26.1	10.9	323	8.1	61	49	139	83	200	127	131	85
Yukon	83	28	8.9	3.3	25.8	4.1	34.7	6.5	65.3	6.5	28.1	4.9	324	3.8	41	26	125	79	167	104	148	103
Northwest Territories	62	37	19.0	8.1	29.6	3.6	48.6	10.4	51.4	10.4	20.7	6.9	322	5.5	109	50	186	93	294	135	145	105
Nunavut	33	27	45.5	11.0	28.9	2.7	74.5	9.8	25.5	9.8	8.3	5.5	314	5.3	297	153	181	71	479	215	48	27
CANADA	48954	29936	14.4	8.9	31.1	6.6	45.5	13.9	54.5	13.9	19.9	10.0	317	7.9	79	80	173	141	252	212	112	108

Figure 4. Number of adults at Levels 1 and 2 prose literacy in Canada

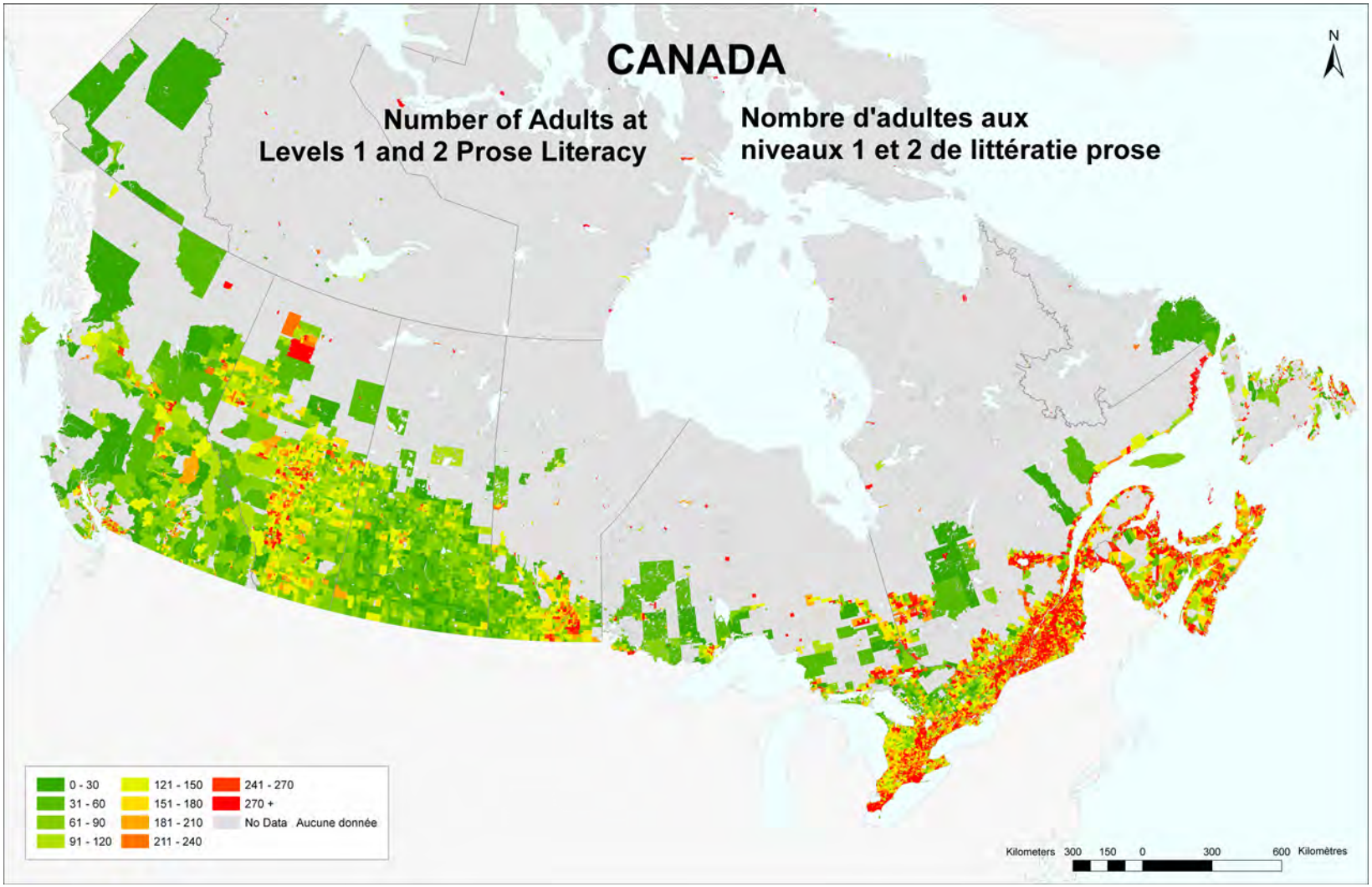

Figure 4 displays the number of adults at levels 1 and 3 prose literacy for all of Canada. This map was developed as a high resolution “poster” to be used as an introduction to the CRISP mapping technique, and as a policy tool for Human Resources and Skills Development Canada. The projection used in this map is the "Lambert conformal conic", which allows one to exaggerate the size of areas at more southern latitudes. This is often used in maps of Canada as most of the population is relatively close to its southern borders.

4. Concluding Remarks

To the best of our knowledge this work represents the first attempt to map social outcomes on a large scale using Statistics Canada's survey data. It respects the confidentiality of individuals, as the estimates for any small areas are based on reasonably large local area samples. This work opens up several possibilities for the mapping of social outcomes, such as health outcomes based on the Canadian Community Health Survey or early childhood outcomes based on the National Longitudinal Survey of Children and Youth. Moreover, the technique can be used with data from any survey; the critical requirement is data describing people's age and sex and the Dissemination Area in which they live. The accuracy of the estimates depends on the density of the data in the region being examined.

The techniques presented in this paper can be improved upon in several ways:

(1) If the pseudo-file of all Canada included information on people's level of education, as well as their sex and age, the individual-level model predicting literacy outcomes would be stronger. This would require Statistics Canada to produce a cross-tabulation of "level of education by sex by age" for each Dissemination Area. For other surveys, breakdowns of other DA characteristics would likely be called for, such "income by age by sex". With these data the prediction equations at the micro level would be stronger.

(2) This analyses used the plausible values inherent in the ***IALSS*** test scores, but did not use the multiple jack-knife design weights. Their use would enable the generation of standard errors for each of the small area statistics. This is relatively straightforward computationally; it more or less requires repeating our computations thirty times and aggregating the results. Practically, though, the CRISP program uses SPSS for most of its calculation, and as noted above, some of the runs require 2 to 3 days. If software that would generate standard errors were to be designed, it would likely use a program like S-Plus, or perhaps develop specialized software that did not rely as heavily on pre-packaged statistical routines. It might also be challenging to map both the indictor and its standard error on the same map.

(3) It might be worthwhile to figure out a way to present data that essentially standardizes the size of the geographical unit. We noted the difference between trying to map the percentage of people with a particular outcome versus the number of people. The problem is that the DAs vary not only in their size in terms of the number of people in each of them, but also in their geographic

area. For example, the average DA size is about 600 people, but they vary considerably, ranging from 40 to 11,657, and a standard deviation of 356. Also, the distribution is skewed (skewness = 3.97) with 90% of the DAs having a total population between 261 and 1035. What is needed is a technique that would enable the generation of estimates of statistics for spatial units that had each covered the same geographic area (e.g., in hectares). This is a much greater challenge than the two previous modifications.

Considering the ***IALSS*** specifically, this study opens up several possibilities. Much of the work at CRISP has focused on socioeconomic gradients; that is, the relationship between particular social outcomes (e.g., literacy scores, childhood obesity, early childhood outcomes, secondary school achievement) and socioeconomic status. The work presented here could be modified to discern whether socioeconomic gradients varied locally. The hypothesis would be that in areas where there are strong literacy outcomes there would be smaller gaps between people of low and high socioeconomic status[4]. The same ideas and methods could be applied to examining the literacy gaps between males and females, or between immigrants and non-immigrants, and then mapping these for each of Canada's major cities. In that research vein, there is also interest in whether the gaps between people of differing status varies as a function of the extent to which they are segregated in certain areas of Canada's major cities. For example, we might expect that the gap in literacy scores between immigrants and non-immigrants would be less in cities where immigrants are not as highly segregated in particular city neighbourhoods.

[4] For example see: Willms, J.D. (2003). Literacy proficiency of youth: Evidence of converging socioeconomic gradients. *International Journal of Educational Research 39*(3), 247-252.