

Répartition géographique des compétences en littératie des adultes au Canada selon les estimations locales

J. Douglas Willms

Institut canadien de recherche en politique sociale et
KSI Research International

avec l'aide de

Richard Chan et Teresa Tang



Remerciements

Nous remercions Satya Brink, Scott Murray et Urvashi Dhawan-Biswal de nous avoir conseillés lors de ce projet. Les méthodes de report graphique des résultats sociaux utilisés dans le présent projet se sont développées sur plusieurs années grâce à la collaboration du Conseil de recherches en sciences humaines, pour le poste de Willms à titre de chaire de recherche du Canada en développement humain à l'Université du Nouveau-Brunswick et pour le programme de recherche Raising and Levelling the Bar in Children's Cognitive, Behavioural and Health Outcomes. Nous remercions également Atlantic Networks for Prevention Research et l'Atlantic Rural Centre, qui ont fourni l'infrastructure ayant permis de réaliser ce genre de recherche au Canada atlantique.

Table des matières

1	La méthode de report graphique des résultats sociaux de l'ICRPS.....	4
2	Cartes illustrant la compréhension de textes suivis.....	11
3	Interprétation des cartes.....	13
4	Conclusions	22
	<i>Annexe A. Index des cartes.....</i>	<i>25</i>

Répartition géographique des compétences en littératie des adultes au Canada selon les estimations locales

1. La méthode de report graphique des résultats sociaux de l'ICRPS

Le présent rapport contient des cartes géographiques sur le niveau de littératie des adultes de chaque province et territoire et des trois plus grandes villes du Canada. Les données de ce projet proviennent de l'Enquête internationale sur l'alphabétisation et les compétences des adultes (*EIACA*) 2003 réalisée par Statistique Canada et l'Organisation de coopération et de développement économiques¹ et du Recensement du Canada 2001. Le projet s'appuie sur une technique de report graphique conçue par l'Institut canadien de recherche en politique sociale au cours des deux dernières années (Willms et Chan, 2005).

La technique de report graphique de l'ICRPS consiste à estimer la valeur d'une variable de résultat pour tous les citoyens canadiens, selon les meilleurs renseignements qui existent pour chaque personne, puis à afficher les résultats sur une carte locale ou provinciale. Cette méthode s'appuie sur les données du Recensement du Canada de 2001 et crée un fichier pour chaque province, celui-ci contenant un « pseudo-registre » pour chaque habitant de la province, d'après la répartition de la population selon le sexe et l'âge dans chaque aire de diffusion (AD). Ensuite, on produit une estimation des résultats (dans ce cas-ci la valeur du niveau de littératie) de chaque personne figurant dans le pseudo-registre à l'aide de techniques de régression multiple à plusieurs niveaux en s'appuyant sur les données suivantes : a) données individuelles d'une enquête de Statistique Canada (dans ce cas-ci l'*EIACA*) indiquant les résultats des autres personnes du même âge, du même sexe et de la même AD, et des autres AD de leur région, et

¹ Statistique Canada et Ressources humaines et Développement des compétences Canada (2005). *Apprentissage et réussite : premiers résultats de l'Enquête internationale sur la littératie et les compétences des adultes*. Ottawa, No 89-617-XIF au catalogue

b) renseignements à l'échelle des AD sur les résultats moyens et les caractéristiques démographiques de toutes les AD de la province. Ce travail s'effectue en plusieurs étapes.

1^{re} étape. Création de fichiers provinciaux des « pseudo-registres ».

Les données du recensement contiennent des variables disant combien d'hommes et de femmes d'un groupe d'âge précis se trouvent dans une AD. Par exemple, la variable V5 indique le nombre de garçons de 0 à 4 ans, la V6 le nombre de garçons de 5 à 9 ans, jusqu'à V22, qui représente le nombre d'hommes de 85 ans et plus. L'ensemble comparable de variables pour les femmes est placé à la suite. Ces données servent à créer un fichier *au plan individuel* pour toute la population de chaque province. Dans chaque AD, les registres individuels montrent la répartition selon l'âge et le sexe, comme dans le fichier du recensement au niveau des AD. Ces fichiers s'appellent les pseudo-fichiers provinciaux. Par exemple, à Terre-Neuve, on compte 1 220 AD, et la population totale est de 511 272 habitants. Le fichier du recensement indique que la première AD, numérotée 10010001, compte une population totale de 210 personnes, dont 0 garçon de 0 à 4 ans, 5 garçons de 5 à 9 ans, jusqu'à 5 femmes de 75 à 79 ans, 0 femme de 80 à 85 ans, et 0 femme de 85 ans et plus. Ainsi, le fichier sur le plan individuel montre 210 registres pour l'AD 1001001 et trois variables : l'AD, le sexe et l'âge. Le code de sexe est 0 pour les hommes, et 1 pour les femmes. La variable d'âge se voit attribuer la valeur représentant le milieu du groupe d'âge (c.-à-d. 2 pour les 0 à 4 ans, 7 pour les 5 à 9 ans, etc.). Le pseudo-fichier de Terre-Neuve compte un registre pour chaque personne, donc un total de 511 272 registres.

Au total, on compte 13 pseudo-fichiers, un pour chaque province et territoire. Chaque fichier contient le même nombre de registres que le nombre de personnes habitant dans la province, selon les données du recensement.

2^e étape. Création d'un fichier de contiguïté à trois niveaux.

Chaque AD compte une ou plusieurs AD contiguës (à l'exception des AD couvrant la totalité d'une petite île). Ces « voisins » sont des voisins de contiguïté de niveau 1. La première AD de Terre-Neuve, 10010001, n'a qu'un seul voisin contigu, 10010002. On détermine ensuite quels sont les voisins qui sont contigus aux voisins contigus de niveau 1. Dans ce cas-ci, il y en a trois : 10010003, 10010004 et 10010005. Ceux-ci s'appellent des voisins contigus de niveau 2. Si on va encore plus loin, on cerne les voisins contigus de niveau 3, et dans ce cas-ci, on en compte

quatre : 10010006, 10010010, 10010470 et 10010471. Ces neuf AD (l'AD cible et ses voisins) constituent ce qu'on appelle un groupe d'AD. Notre logiciel bâtit le groupe ou la grappe d'AD pour 10010001, puis procède au prochain groupe, 10010002, en répertoriant ses voisins contigus de niveaux 1, 2 et 3. Dans ce cas-ci, il y en a 4 au niveau 1, 4 au niveau 2, et 11 au niveau 3. Le logiciel de l'ICRPS exécute ce processus pour l'ensemble des 52 924 AD au Canada et crée ainsi notre « fichier de contiguïté principal » (la création de ce fichier nécessite plusieurs semaines de totalisations informatiques au moyen d'un ordinateur relativement rapide). Donc, le fichier de contiguïté principal compte X registres pour chaque AD, où X représente : a) le nombre de voisins de niveau 0 (c.-à-d. un, l'AD elle-même), plus b) le nombre de voisins de niveau 1, plus c) le nombre de voisins de niveau 2, plus d) le nombre de voisins de niveau 3. Dans chaque registre, on énumère l'AD, un voisin et le niveau de contiguïté (0, 1, 2 ou 3) de ce voisin. Enfin, il y a un groupe d'AD (l'AD et ses voisins) pour chaque AD.

3^e étape. Création de variables de recensement pour l'estimation de l'équation du niveau de l'AD.

Ici, on utilise les données du recensement pour créer les variables du niveau de l'AD qui servent à l'analyse finale où l'on estime la valeur du niveau de littératie. Plusieurs variables ont été créées, mais notre modèle utilisé pour la valeur du niveau de littératie des adultes comprenait sept variables du niveau de l'AD :

Années d'études : Le nombre moyen d'années d'études dans l'AD.

Migration : Le pourcentage de personnes qui avaient déménagé au cours des cinq années précédentes.

Revenu : Le niveau moyen de revenu familial.

Taux de chômage : Pourcentage de chômeurs dans l'AD.

Pourcentage dans les classes sociales 1 et 2 : Pourcentage de personnes de l'AD exerçant une profession dans le secteur professionnel ou semi-professionnel².

Pourcentage dans les classes sociales 4, 5 et 6 : Pourcentage de personnes exerçant un métier non spécialisé ou non classé.

² Les professions comprises sont celles décrites aux variables v989, v994, v1001, v1004, v1009, v1010, v990, v995, v1005, v1011, v1013, v1016, v1017 et v1027.

Pourcentage d'immigrants récents : Pourcentage de personnes qui avaient immigré au cours des cinq années précédentes.

Imputation des données manquantes à l'aide de renseignements locaux. Pour la plupart des variables, il manque certaines données dans le fichier du recensement au niveau des AD. Par exemple, bon nombre d'AD ne comptent pas d'estimation du revenu familial moyen. Pour toutes les AD, nous avons estimé une « valeur lissée ». La valeur lissée d'une AD représente la *moyenne pondérée des résultats de toutes les AD de son groupe d'AD correspondant*. La moyenne lissée d'une AD reçoit la pondération des valeurs des AD de son groupe d'AD selon la population totale de chaque AD et le niveau de contiguïté, le niveau 0 ayant une pondération de 1,0, le niveau 1 de 0,75, le niveau 2 de 0,50 et le niveau 3 de 0,25.

Le niveau de revenu pour une AD est ensuite réglé à la valeur qui lui est attribuée dans le recensement, le cas échéant. Par contre, s'il manque des données, le niveau de revenu est fixé à sa valeur lissée.

4^e étape. Création des variables de l'EIACA pour l'estimation de l'équation au niveau individuel.

Jusqu'ici, nous nous sommes uniquement servis du fichier du recensement du niveau des AD. Nous y ajoutons maintenant les données d'enquête de l'*EIACA* pour créer un fichier de niveau individuel comprenant des données sur les variables suivantes :

Province

AD

Conception de la pondération de l'enquête

Conception de la pondération normalisée de l'enquête. Il s'agit de la pondération de la conception de l'enquête sur la population, normalisée de façon à montrer une moyenne de 1,0 à l'échelle provinciale.

Pondérations selon les répétitions de la méthode du Jackknife. Il n'y en a pas dans la présente analyse descriptive, mais on pourrait s'en servir plus tard si on veut estimer l'erreur-type de chacune de nos estimations locales.

Niveau de littératie. On donne une « valeur plausible » pour les résultats en compréhension de textes suivis et de documents schématiques, en calcul et en résolution de problèmes.

Dans le présent rapport, seules les cinq valeurs plausibles pour la compréhension de textes suivis ont été utilisées.

Âge. Âge du répondant.

Sexe. Sexe du répondant (code 1.0 pour les femmes, et 0 pour les hommes).

Une version distincte du présent fichier est sauvegardée pour chaque province.

Pour chaque province, on crée également un fichier au niveau des AD qui indique combien de répondants à l'*EIACA* habitaient dans chaque AD.

5^e étape. Création du fichier de groupe d'AD de l'*EIACA*.

Cette étape consiste à créer un fichier contenant les registres de l'*EIACA* disponibles pour chaque groupe d'AD, c'est-à-dire pour chaque AD et chacune de ses AD voisines. Il est plus facile de décrire ce fichier si on ne tient compte que d'un seul groupe d'AD.

Par exemple, supposons une AD « cible » dans laquelle on compte quatre répondants à l'*EIACA*. Les quatre premiers registres du fichier du groupe d'AD pour cette AD cible comprendraient les données de notre fichier individuel de l'*EIACA* décrit à la 4^e étape. Si l'AD cible contenait trois voisins de contiguïté de niveau 1 et que ces AD voisines contenaient, respectivement, 3, 0 et 6 répondants à l'*EIACA*, notre fichier de groupe d'AD pour l'AD cible contiendrait donc neuf registres de plus, chacun contenant les données décrivant la valeur des niveaux de littératie et les variables contextuelles du fichier individuel de l'*EIACA*. On procède de la même façon pour le niveau 2 et le niveau 3, de façon à ce que le fichier contienne désormais des registres pour tous les répondants ayant servi de base dans les AD et le groupe d'AD.

Ce processus s'applique pour tous les groupes d'AD d'une province. Prenez note que, bien qu'il n'y ait pas de répondants à l'*EIACA* dans certaines AD, il se peut qu'il y en ait dans leurs AD voisines de niveau 1, et dans presque tous les cas, on dénombre certains répondants lorsqu'on arrive au troisième niveau de contiguïté des AS. Ainsi, presque tous les groupes d'AD comptent des répondants à l'*EIACA*. Par souci de confidentialité, les estimations qui ne comptent pas au moins cinq répondants dans un groupe d'AD ne sont pas déclarées.

6^e étape. Création du fichier principal pour les analyses HLM.

Ensuite, le fichier issu de la 5^e étape est fusionné avec les données du recensement extraites du fichier du niveau des AD. Ainsi, le fichier de microdonnées comprendra désormais, pour chaque groupe d'AD, les renseignements suivants :

- Les valeurs et les covariables pour tous les répondants à l'*EIACA* de cette AD;
- Les valeurs et les covariables de tous les répondants à l'*EIACA* des AD voisines;
- Les variables du niveau du recensement correspondant à chaque AD et annexées à chaque registre.

7^e étape. Sauvegarde d'un fichier personnel et de fichiers au niveau des groupes d'AD pour l'analyse du modèle hiérarchique.

En servant du fichier créé à la 6^e étape, on sauvegarde un fichier personnel comprenant des renseignements sur le niveau de littératie, l'âge et le sexe pour chaque groupe d'AD. À l'intérieur d'un groupe, on accorde une pondération aux personnes selon leur niveau de contiguïté. Dans ce cas-ci, la variable sexe a été fixée au point milieu (0,5), tout comme l'âge, dont la valeur médiane est de 45 ans. Cela signifie donc que les valeurs estimatives du niveau de littératie pour chaque AD constitueront la moyenne des valeurs prévues pour un homme et une femme de 45 ans.

À la 6^e étape, les variables du recensement ont été fusionnées avec les données de chaque personne. Le fichier au niveau du groupe d'AD compte des valeurs moyennes de ces variables du recensement pour ce groupe, pondérées de façon à ce que les personnes de l'AD cible reçoivent une pondération de 1,0 lorsqu'on totalise la moyenne du groupe d'AD, les voisins de niveau 1, 0,75, etc. Notre fichier de groupe d'AD compte un registre pour chaque AD.

8^e étape. Réalisation de l'analyse HLM.

À cette étape-ci, un modèle linéaire hiérarchique est associé aux données, et les personnes sont nichées à l'intérieur de groupes d'AD. Le modèle à l'intérieur du groupe d'AD établit par régression le niveau de compréhension de textes suivis d'après le sexe, l'âge et l'âge au carré pour chaque groupe d'AD. Le modèle au niveau du groupe d'AD établit par régression les coordonnées d'origine extraites du modèle à l'intérieur du groupe d'AD sur les sept variables de recensement (années d'études, migration, etc.). Cette analyse donne un ensemble distinct de coefficients comprenant une équation de prédiction pour chaque groupe d'AD. Dans le cas de

Terre-Neuve, par exemple, où l'on compte 1 220 AD, on compte aussi 1 220 équations de prévisions distinctes. Ces analyses s'effectuent une province à la fois.

9^e étape. Prévisions des valeurs pour toutes les personnes de chaque AD.

Le pseudo-fichier créé à la première étape est ensuite utilisé, et pour chaque personne, on estime la valeur prévue de son niveau de littératie en fonction des modèles de prévisions produits à la 8^e étape. Prenez note que nous avons estimé le modèle de prévisions à l'échelle locale en nous appuyant sur les données d'enquête pour toutes les personnes du groupe d'AD qui avaient effectivement rempli le sondage, ainsi que des renseignements plus généraux sur la façon dont les caractéristiques des AD se rattachent aux valeurs du niveau de littératie dans une AD. L'équation de prévision contient également une covariable normale aléatoire fondée sur les résultats du modèle de prévision. Celle-ci s'ajoute aux résultats de chaque personne.

Cette étape se caractérise également par un traitement informatique très intense, puisqu'on estime les résultats d'environ 30 millions de personnes en s'appuyant sur les résultats de près de 50 000 équations de régression distinctes. Dans la plupart des provinces, les analyses s'effectuent en quelques heures, sauf que les résultats de l'Ontario et du Québec prennent de deux à trois jours à traiter. À la fin, on obtient un fichier personnel montrant une valeur prévue du niveau de littératie pour chaque personne.

10^e étape. Report graphique des données.

À la dernière étape, on se sert des données pour créer une carte de chaque province et montrer les données de chaque AD. Le logiciel de report graphique utilisé, ARCVIEW (dont se sert également Statistique Canada), nécessite qu'on établisse des catégories fixes (p. ex., environ 8 à 10) pour l'affichage en différentes couleurs sur les cartes. Pour déterminer les points de découpage, nous avons examiné la répartition des résultats pour l'ensemble du Canada à l'aide du pseudo-fichier final de la population, et tenté de porter des jugements raisonnables d'une variable à l'autre.

Bien que cette méthode nécessite un traitement informatique très intense, l'une de ses forces, c'est que les données peuvent être résumées aux niveaux supérieurs, comme les SDR, les villes, etc. selon les résultats de notre pseudo-fichier sur la population.

2. Cartes illustrant la compréhension de textes suivis

Le présent rapport contient 180 cartes et deux « murales » pour tout le Canada. Les cartes sont organisées en groupes de dix, soit un groupe pour chacune des dix provinces, pour chacun des trois territoires, et chacune des trois plus grandes villes (Vancouver, Montréal et Toronto). De plus, on a produit des ensembles de cartes distincts pour le Sud-Est de l'Ontario et le Sud du Québec en raison de la forte densité de population de ces régions.

Les niveaux de littératie des adultes selon l'Enquête internationale sur l'alphabétisation et les compétences des adultes représentent cinq catégories, selon le niveau de compétences d'une personne. Les tâches requises pour chaque niveau sont résumées ci-dessous (voir Statistique Canada et OCDE, 2005, p. 17) :

- Niveau 1** « La plupart des tâches de ce niveau amènent le répondant à lire un texte relativement court pour repérer un élément d'information identique ou semblable à l'information donnée dans la question ou la directive.
- Niveau 2** Certaines tâches de ce niveau amènent le répondant à repérer un élément d'information dans le texte, qui peut contenir plusieurs éléments de distraction ou des éléments d'information plausibles mais incorrects; le répondant peut aussi devoir faire des déductions de faible niveau. D'autres tâches l'amènent à intégrer deux ou plusieurs éléments d'information ou à comparer et à mettre en opposition des éléments d'information facilement repérables en fonction d'un critère donné dans la question ou la directive.
- Niveau 3** Les tâches de ce niveau amènent habituellement le répondant à faire des adéquations littérales ou synonymiques entre le texte et l'information donnée dans la tâche, ou des adéquations nécessitant des déductions de faible niveau. D'autres tâches lui demandent d'intégrer des éléments d'information présents dans un texte dense ou long qui ne contient pas de rubriques ou d'intertitres. On peut aussi demander au répondant de fournir une réponse fondée sur une information facilement repérable dans le texte. Des éléments de distraction sont présents, mais ne sont pas situés près de la bonne information.
- Niveau 4** Ces tâches amènent le répondant à faire l'adéquation de plusieurs caractéristiques et à intégrer ou résumer des éléments d'information présents dans des passages complexes ou longs. À cette fin, il doit faire des déductions plus complexes. Les tâches de ce niveau contiennent souvent des renseignements conditionnels dont le répondant doit tenir compte.

Niveau 5 Certaines tâches de ce niveau amènent le répondant à chercher de l'information dans un texte dense contenant un certain nombre d'éléments de distraction plausibles. D'autres lui demandent de faire des déductions de haut niveau ou de faire appel à des connaissances spécialisées. Certaines tâches demandent au répondant de mettre en opposition des éléments d'information complexes. » (Statistique Canada et OCDE, 2005, p. 17)

Chaque ensemble de cartes comprend les cartes suivantes :

- 1 Pourcentage au niveau 1 en compréhension de textes suivis
- 2 Pourcentage au niveau 2 en compréhension de textes suivis
- 3 Pourcentage aux niveaux 1 et 2 en compréhension de textes suivis
- 4 Pourcentage aux niveaux 3, 4 et 5 en compréhension de textes suivis
- 5 Pourcentage aux niveaux 4 et 5 en compréhension de textes suivis
- 6 Résultats moyens des adultes aux niveaux 3, 4 et 5
- 7 Nombre d'adultes au niveau 1 en compréhension de textes suivis
- 8 Nombre d'adultes au niveau 2 en compréhension de textes suivis
- 9 Nombre d'adultes aux niveaux 1 et 2 en compréhension de textes suivis
- 10 Nombre d'adultes aux niveaux 4 et 5 en compréhension de textes suivis

3. Interprétation des cartes

Le présent rapport compte trois types de cartes fondamentaux : des cartes montrant le pourcentage d'adultes à des niveaux précis de compréhension de textes suivis, des cartes montrant le nombre d'adultes à chaque niveau de compréhension de textes suivis, et d'autres montrant les résultats moyens à l'épreuve de compréhension de textes suivis. Une personne qui sait comment interpréter ces trois cartes peut interpréter toutes les autres.

Cartes montrant le pourcentage d'adultes à des niveaux précis de compréhension de textes suivis

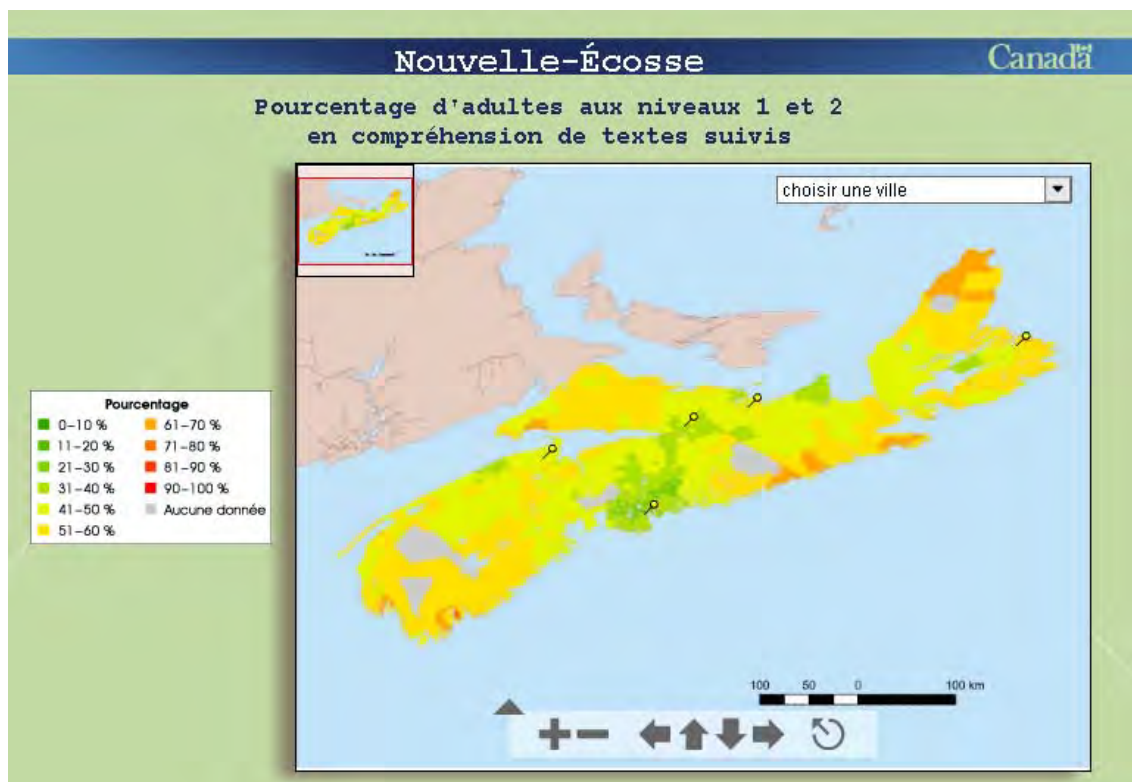


Figure 1. Pourcentage d'adultes aux niveaux 1 et 2 en compréhension de textes suivis en Nouvelle-Écosse

La figure 1 donne un exemple d'une carte montrant le pourcentage d'adultes aux niveaux 1 et 2 en compréhension de textes suivis. C'est la Nouvelle-Écosse qui sert d'exemple. Les fourchettes de pourcentage pour toutes les cartes montrant le pourcentage de personnes aux niveaux 1 et 2 sont des tranches de 10 points (c.-à-d. 0 % à 10 %, 10 % à 20 %, etc.). Dans ces

cartes, les couleurs varient de vert foncé (généralement un résultat jugé positif – dans ce cas-ci, un faible pourcentage d’adultes dans les deux plus faibles niveaux de littératie) à rouge vif. Les régions qui ne contenaient pas de données (ou des données insuffisantes pour afficher les résultats) apparaissent en gris pâle. Par exemple, lorsqu’on regarde l’extrémité nord-est de la province, le secteur est principalement orange pâle, ce qui indique qu’environ 61 % à 70 % des adultes de 16 à 65 ans de cette région se classent aux niveaux 1 et 2 de littératie.

Il faut interpréter cette carte avec prudence. On ne peut pas cibler une aire de diffusion précise et affirmer avec certitude qu’il y a, par exemple, de 60 % à 70 % des adultes de cette région qui se classent aux niveaux 1 et 2 en compréhension de textes suivis. La technique d’estimation permet d’estimer le résultat pour une aire de diffusion en fonction des résultats à l’*EIACA* des personnes de cette région qui ont répondu à l’enquête, et des résultats des personnes des aires de diffusion voisines. Elle s’appuie aussi sur plusieurs caractéristiques de la région, mesurées au niveau des AD, pour estimer le résultat. Par exemple, imaginez que l’on réalise uniquement une analyse de régression globale modélisant les résultats en compréhension de textes suivis selon les caractéristiques individuelles des adultes (c.-à-d. l’âge et le sexe) et les caractéristiques de leur aire de diffusion (c.-à-d. le nombre d’années d’études, la migration, le revenu, le taux de chômage, le pourcentage de personnes dans les classes sociales 1 et 2, et dans les classes sociales 4, 5 et 6) pour l’ensemble des données pondérées de l’*EIACA*. On pourrait ensuite utiliser ces résultats de régression (c.-à-d. les coefficients non normalisés) pour prédire le niveau de littératie de n’importe quel adulte au Canada, ou de n’importe quelle aire de diffusion au Canada, selon l’âge et le sexe d’une personne et la région où elle habite. Les résultats pour une personne ne seraient pas très précis, sauf que les résultats pour une AD seraient probablement assez fiables. La méthode qu’on a utilisée fonctionne essentiellement de cette façon, mais plutôt que de s’appuyer sur une régression globale pour l’ensemble du Canada, elle se sert de données locales pour chaque région afin d’estimer l’équation de régression. Ainsi, notre méthode est beaucoup plus précise, mais il faut garder en tête qu’il s’agit quand même d’une valeur estimée comportant un niveau d’incertitude attribuable à l’erreur d’échantillonnage et de mesure.

Pour visualiser la méthode, imaginez qu’on laisse tomber une petite goutte de peinture rouge sur un ciel bleu dans la création d’une aquarelle. La goutte rouge se mélangerait avec le bleu, ce qui créerait un cercle de couleurs passant graduellement du rouge au violet, puis au bleu (le ciel). Les données des 23 038 répondants à l’*EIACA* sont comme les gouttes de peinture rouge, mais

celles-ci varient en couleurs selon le résultat réel d'une personne à l'épreuve de littératie. De plus, les gouttes ne sont pas déposées sur un fond bleu, mais plutôt placées sur un fond dont la couleur reflète le contexte local de chaque personne, d'après les caractéristiques sociales et économiques de son voisinage telles que déterminées par les données extraites du recensement. Les couleurs s'amalgament et forment une image représentant la variation des niveaux de littératie dans toutes les provinces.

Cartes montrant le nombre d'adultes se situant à un niveau ou à des niveaux précis

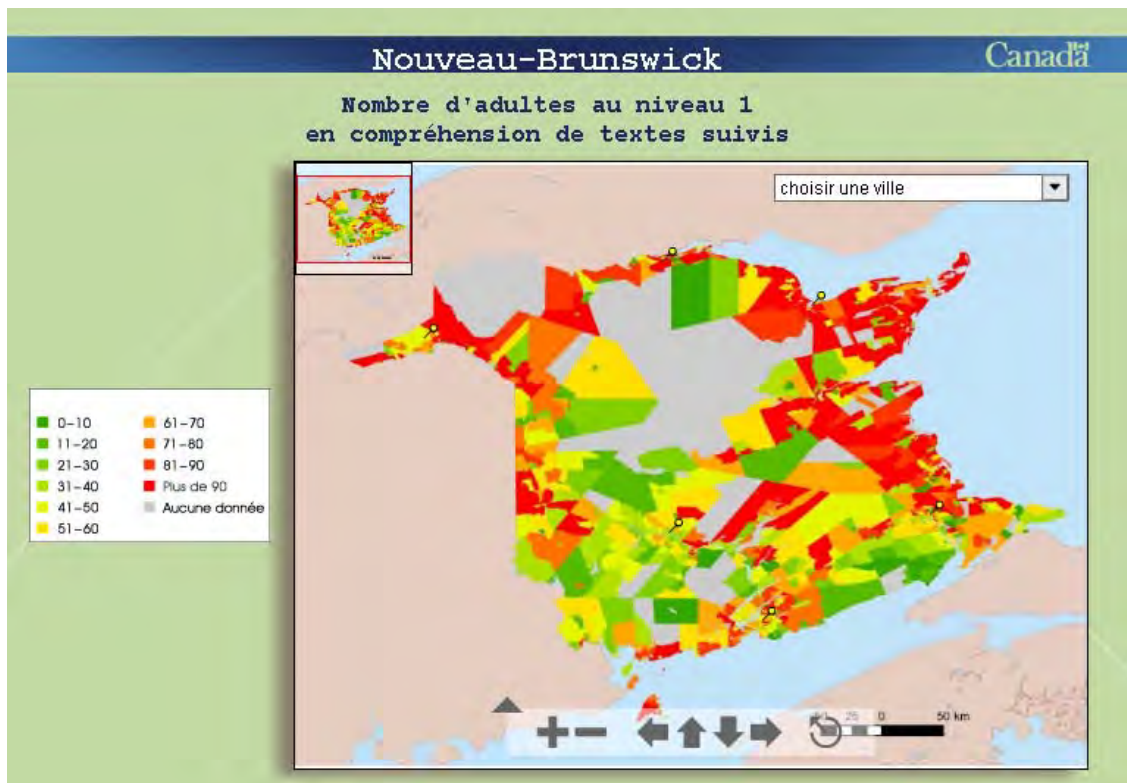


Figure 2. Nombre d'adultes au niveau 1 en compréhension de textes suivis au Nouveau-Brunswick

En préparant les cartes montrant les pourcentages, on a découvert que leur présentation graphique pouvait prêter quelque peu à confusion, car il se peut qu'une personne fasse un lien entre une région et un nombre d'habitants. Toutefois, les pourcentages se rapportent au pourcentage de personnes habitant dans une région géographique précise. Par exemple, imaginez une région où les personnes montrant un niveau élevé de littératie sont concentrées dans une grande ville, tandis que les personnes montrant le niveau le plus faible sont surreprésentées dans

les régions rurales. La carte serait en majeure partie rouge, puisque les sections vertes représenteraient principalement la grande ville, ce qui ne représenterait pas une grande superficie sur la carte. On aurait l'impression que la plupart des personnes de cette région montrent des niveaux relativement faibles de littératie, bien que la majorité des personnes, qui vivent surtout dans la grande ville, aient un niveau de littératie relativement élevé.

Par conséquent, on a également produit une série de cartes montrant le nombre d'adultes de la province ou de la ville dont les compétences en littératie se situent à un niveau particulier. La figure 2 en constitue un exemple. On y montre le nombre de personnes de chaque aire de diffusion du Nouveau-Brunswick qui ont enregistré un niveau de littératie 1 et 2. Tout comme la carte qui montre les pourcentages d'adultes ayant de faibles niveaux de littératie, celle-ci révèle qu'un grand nombre d'adultes des secteurs ruraux montrent de faibles niveaux de littératie. Toutefois, si on se concentre sur les trois grandes villes, Fredericton, St. John et Moncton, on remarque que, à l'intérieur de ces villes, il y a des secteurs où les gens montrent de faibles niveaux de littératie. Par ailleurs, certaines régions rurales ont une très faible densité de population, et celles-ci montrent un total relativement faible de personnes s'étant classées dans les plus faibles niveaux de littératie, bien qu'elles représentent un nombre relativement élevé en pourcentage. À de nombreux égards, ces cartes sont les plus utiles pour discerner à quel endroit il est préférable de cibler certains programmes.

Cartes montrant les niveaux moyens de littératie

On peut se servir exactement des mêmes techniques décrites précédemment pour produire le niveau moyen de littératie de chaque région, ou le niveau médian, ou toute statistique descriptive utile à des fins stratégiques. Ressources humaines et Développement des compétences Canada s'intéressait particulièrement aux résultats moyens des adultes qui s'étaient classés aux niveaux 3, 4 et 5. À titre d'exemple, voici, à la figure 3, la carte de Toronto.

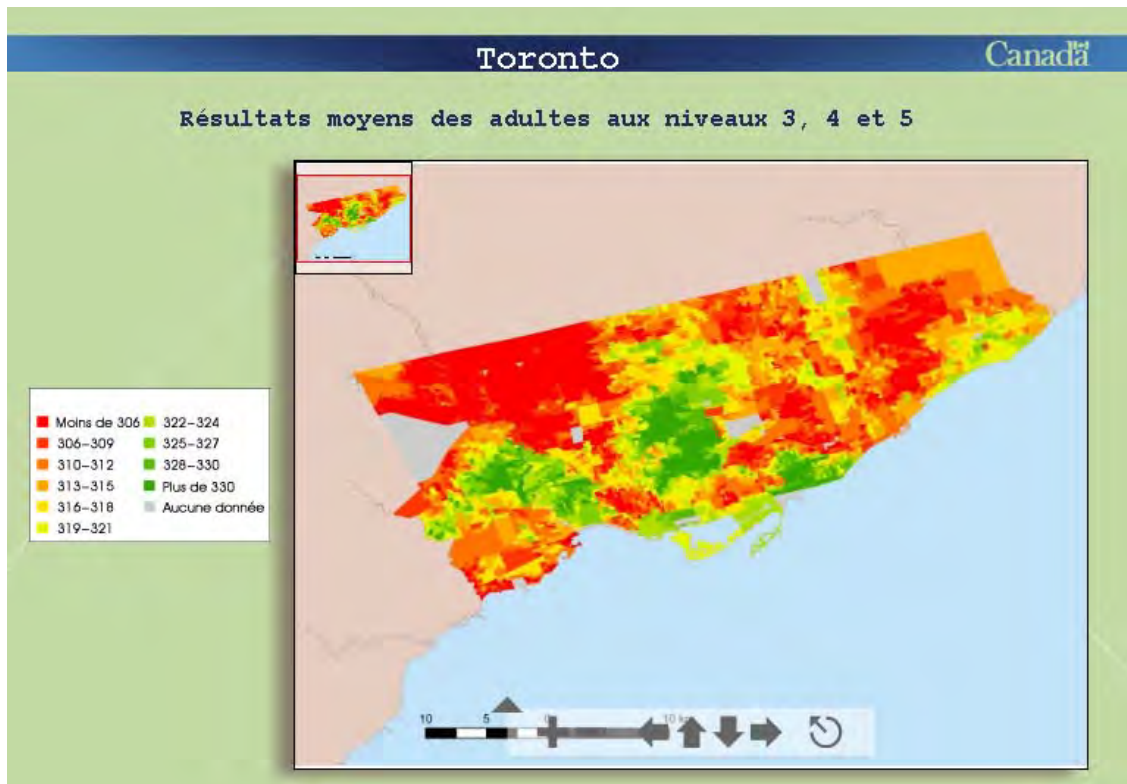


Figure 3. Niveaux moyens de compréhension de textes suivis pour les adultes s'étant classés aux niveaux 3, 4 et 5 à Toronto.

Dans ce cas-ci, les sections rouges sont les endroits où le résultat moyen chez les adultes des trois premiers niveaux est relativement faible, c'est-à-dire inférieur à 306. Les sections vert foncé sont les endroits où l'on a observé des résultats moyens élevés chez les adultes qui se sont classés dans les trois premiers niveaux. Puisque cette statistique est en grande partie attribuable à la répartition des résultats des personnes constituant le haut de l'échelle, on pourrait penser que les sections vert foncé sont les endroits où l'on observe une prépondérance de personnes dont le niveau de littératie est très élevé.

Statistiques de l'ensemble du Canada

Un indicateur du rendement peut généralement prendre son sens de trois façons : par comparaison avec une norme fixe ou à une série de normes, par comparaison avec les tendances au fil du temps, ou par comparaison entre les provinces et territoires. L'*EIACA* permet d'appliquer ces trois interprétations. D'une part, le cadre de mise à l'essai constitue un moyen de classer les résultats des personnes selon des niveaux de compétence (c.-à-d. les niveaux 1 à 5)

ayant un sens puisqu'il se rattache à des types de compétences précises. Les interprétations dont il est question précédemment se fondent sur les normes inhérentes à cette classification et exigent que la personne interprète une carte pour se familiariser avec la signification des résultats à chaque niveau de compétence. La deuxième interprétation nécessite des données longitudinales sur les personnes ou des données décrivant les échantillons de la même population à deux ou plusieurs périodes. L'*EIACA* s'appuyait sur les mêmes domaines et les mêmes questionnaires (à quelques modifications mineures près) pour qu'on puisse comparer les résultats d'une période avec ceux d'une autre période³. Bien que ce ne soit pas ce genre d'analyse qu'on utilise ici, il serait possible de le faire en s'appuyant sur des données sur les aires de diffusion pour les données sur la littératie des adultes de 1994 et de 2003. La troisième méthode nécessite la compréhension de l'amplitude des variations entre les provinces et les territoires. Dans une carte, cette distinction se fait naturellement puisqu'on peut voir l'étendue d'une variation à l'intérieur d'une province grâce à l'agencement des couleurs sur la carte. Toutefois, il est également utile de voir les résultats dans le contexte de l'ensemble du Canada. Le tableau 1 montre la variation entre les provinces pour chacun des dix indicateurs présentés dans une série de cartes. Cette analyse est réalisée au niveau des AD, car ce sont les AD qui sont représentées sur les cartes. Toutefois, les résultats sont pondérés selon le nombre d'adultes dans chaque AD, car ce sont ces comparaisons qui sont les plus pertinentes pour la politique sociale.

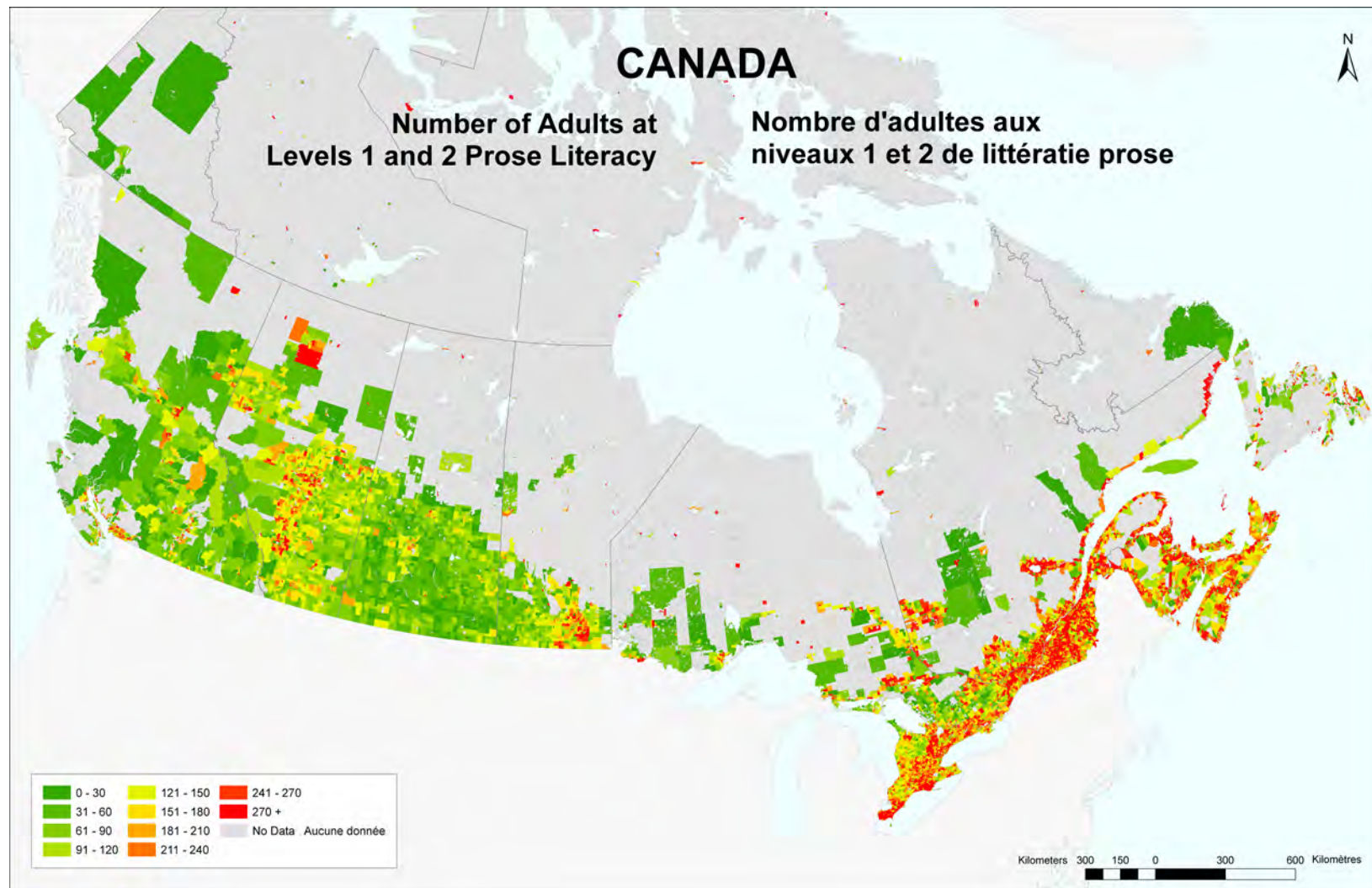
³ Par exemple, voir :

Willms, J.D. (2005). Les compétences, le niveau de scolarité des parents et la pratique de la littératie dans la vie courante. Dans Statistique Canada et OCDE (éd.), *Apprentissage et réussite : Premiers résultats de l'Enquête internationale sur la littératie et les compétences des adultes* (p. 229 à 246). Ottawa et Paris : Statistique Canada et OCDE.

Tableau 2. Statistiques descriptives sur la compréhension de textes suivis des adultes dans les provinces et les territoires

Province	Nombre d'AD	Taille de la population (en milliers)	% niveau 1		% niveau 2		% niveaux 1 et 2		% niveaux 3, 4 et 5		% niveaux 4 et 5		Résultat moyen niveaux 3, 4 et 5		Nombre d'adultes au niveau 1		Nombre d'adultes au niveau 2		Nombre d'adultes aux niveaux 1 et 2		Nombre d'adultes aux niveaux 4 et 5	
			Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT	Moy.	ÉT
Terre-Neuve	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
Île-du-Prince-Édouard	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
Nouvelle-Écosse	1 276	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
Nouveau-Brunswick	1 223	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
Québec	11 558	7 228	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	17 745	11 378	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	2 024	1 117	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	2 387	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	4 789	2 970	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
Colombie-Britannique	6 647	3 897	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
Territoires du Nord-Ouest	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	48 954	29 936	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. Nombre d'adultes aux niveaux 1 et 2 en compréhension de textes suivis au Canada



La figure 4 montre le nombre d'adultes aux niveaux 1 et 3 en compréhension de textes suivis pour l'ensemble du Canada. Cette carte a été conçue sous forme de « murale » à haute résolution et doit servir d'introduction à la technique de report graphique de l'ICRPS et d'outil stratégique pour Ressources humaines et Développement des compétences Canada. La projection utilisée dans cette carte est la « projection conique conforme de Lambert », qui permet d'exagérer la taille des régions les plus au sud. On s'en sert souvent dans les cartes du Canada puisque la majorité de la population se situe relativement près des frontières sud.

4. Conclusions

Pour autant que nous sachions, ce travail représente la première tentative de report graphique des résultats sociaux à grande échelle à l'aide des données d'enquête de Statistique Canada. Ce projet respecte la confidentialité des personnes étant donné que les estimations pour les petites régions se fondent sur les échantillons de régions locales relativement vastes. Ce travail crée plusieurs nouvelles possibilités de report graphique des résultats sociaux, comme les résultats de santé d'après l'Enquête sur la santé dans les collectivités canadiennes ou les résultats de la petite enfance d'après l'Enquête longitudinale nationale sur les enfants et les jeunes. Par ailleurs, on peut utiliser cette technique avec les données de n'importe quelle enquête; le critère essentiel, ce sont des données décrivant l'âge et le sexe d'une personne et l'aire de diffusion où elle habite. La précision des estimations dépend de la densité des données dans la région à l'étude.

Les techniques faisant l'objet du présent document peuvent être améliorées de plusieurs façons :

- 1) Si le pseudo-fichier de l'ensemble du Canada comprenait des renseignements sur le niveau de scolarité des personnes, ainsi que leur sexe et leur âge, le modèle individuel prévoyant des résultats en littératie serait plus solide. Statistique Canada serait tenu de produire une totalisation croisée du « niveau de scolarité selon le sexe et l'âge » pour chaque aire de diffusion. Dans les autres enquêtes, il faudrait probablement obtenir la répartition des autres caractéristiques des AD, comme le « revenu selon l'âge et le sexe ». Ces données rendraient plus solides les équations de prévision au niveau microéconomique.
- 2) Cette analyse s'est servie des valeurs plausibles inhérentes aux résultats de l'*EIACA*, mais pas des pondérations multiples du plan d'échantillonnage de la méthode du Jackknife. Ces pondérations permettraient de produire une erreur-type pour chacune des statistiques sur les petites régions. Sur le plan informatique, cette méthode est relativement simple : il suffit plus ou moins de répéter nos totalisations une trentaine de fois et de regrouper les résultats. Par contre, de façon pratique, le programme de l'ICRPS se sert de SPSS pour la plupart de ses calculs, et tel qu'indiqué précédemment, certaines phases peuvent prendre de deux à trois jours. Si on devait concevoir le logiciel de production d'erreur-type, celui-ci fonctionnerait à l'aide d'un programme comme S-Plus, ou on créerait peut-être un logiciel spécialisé ne s'appuyant pas autant sur des

méthodes statistiques routinières préfabriquées. Il pourrait également être intéressant de reporter graphiquement l'indicateur et son erreur-type sur la même carte.

3) Cela pourrait valoir la peine de trouver une façon de présenter des données qui normalisent essentiellement la taille de l'unité géographique. Nous avons remarqué la différence en tentant de reporter le pourcentage de personnes ayant un résultat particulier par rapport au nombre de personnes. Le problème, c'est que les AD varient non seulement en taille selon le nombre de personnes dans chacune d'elles, mais également en superficie géographique. Par exemple, la taille moyenne d'une AD est d'environ 600 personnes, sauf que ce nombre varie considérablement, de 40 à 11 657, et l'écart-type est de 356. De plus, la répartition est asymétrique (asymétrie = 3,97), 90 % des AD ayant une population totale de 261 à 1 035 personnes. Ce qu'il faut, c'est une technique qui permettrait de produire les estimations des statistiques pour les unités d'espace dont chacune couvrirait la même superficie géographique (p. ex., en hectares). Ce défi est beaucoup plus difficile à relever que les deux autres modifications proposées précédemment.

Si l'on tient compte de l'*EIACA* précisément, cette étude ouvre plusieurs possibilités. La majeure partie du travail à l'ICRPS porte sur les gradients socioéconomiques, c'est-à-dire la relation entre des résultats sociaux particuliers (p. ex., les niveaux de littératie, l'obésité chez les jeunes, les résultats de la petite enfance, la réussite des études secondaires) et le statut socioéconomique. Il serait possible de modifier le travail présenté ici pour déterminer si les gradients socioéconomiques varient à l'échelle locale. Selon l'hypothèse, dans les secteurs où l'on observe de hauts niveaux de littératie, les écarts seraient moins grands entre les personnes d'une bonne et d'une mauvaise situation socioéconomique⁴. On pourrait appliquer les mêmes idées et les mêmes méthodes à l'examen des écarts du niveau de littératie entre les hommes et les femmes, ou entre les immigrants et les non-immigrants, puis en faire un report graphique pour chacune des grandes villes du Canada. Dans cette recherche, il serait également intéressant de voir si les écarts entre les personnes de situations différentes varient en fonction de la mesure dans laquelle ils sont isolés dans certains secteurs des grandes villes du Canada. Par exemple, on pourrait présumer que l'écart des niveaux de littératie entre les immigrants et les non-immigrants

⁴ Par exemple, voir : Willms, J.D. (2003). Literacy proficiency of youth: Evidence of converging socioeconomic gradients. *International Journal of Educational Research* 39(3), 247-252.

soit moindre dans les villes où les immigrants ne sont pas aussi isolés dans des quartiers en particulier.