



L'Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

**Fichier de microdonnées à
grande diffusion**

Manuel de l'utilisateur



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	7
2.0	Aperçu de l'enquête	9
3.0	Objectifs de l'enquête	11
4.0	Concepts et définitions.....	13
4.1	Définir et mesurer les compétences : aperçu.....	13
4.2	Comprendre ce qui a été mesuré lors de l'Enquête internationale sur l'alphabétisation et les compétences des adultes	14
4.2.1	Introduction	14
4.2.2	Échelonner les tâches de littératie, de numératie et de résolution de problèmes dans le cadre de l'EIACA.....	15
4.3	Mesurer la compréhension de textes suivis et de textes schématiques dans le cadre de l'EIACA	16
4.3.1	Définir les caractéristiques des tâches.....	18
4.3.2	Type d'appariement	20
4.3.3	Type d'information demandée.....	20
4.3.4	Plausibilité des éléments de distraction	21
4.3.5	Caractériser les tâches de compréhension de textes suivis	21
4.3.6	Caractériser les tâches de compréhension de textes schématiques	26
4.4	Mesurer la numératie dans le cadre de l'EIACA.....	30
4.4.1	Définir les caractéristiques des tâches.....	32
4.4.2	Vie quotidienne	32
4.4.3	Travail	32
4.4.4	Vie sociale ou collective	33
4.4.5	Formation complémentaire	33
4.4.6	Quantité et nombre	34
4.4.7	Dimension et forme	34
4.4.8	Structure, fonctions et relations.....	35
4.4.9	Données et hasard.....	35
4.4.10	Variation	36
4.4.11	Caractériser les tâches de numératie	36
4.5	Mesurer la résolution de problèmes dans le cadre de l'EIACA	40
4.5.1	Définir les caractéristiques des tâches.....	42
4.5.2	Caractériser les tâches de résolution de problèmes	44
4.6	Conclusion.....	48
4.6.1	Quelques critères analytiques.....	49
5.0	Méthodologie de l'enquête	55
5.1	Plan d'évaluation	55
5.2	Couverture de la population	55
5.3	Base de sondage	56
5.4	Plan d'échantillonnage	57
5.5	Précision recherchée.....	58
5.6	Taille de l'échantillon	58
5.7	Collecte des données	60

Table des matières (suite)

5.8	Notation des tests.....	61
5.9	Réponse à l'enquête et pondération	62
5.9.1	Poids de population.....	62
5.9.2	Poids jackknife	65
6.0	Méthodes d'enquête et traitement des données	69
6.1	Introduction.....	69
6.2	Modèles de méthodes, de manuels et d'instruments	69
6.2.1	Questions de base	70
6.2.2	Items de tâche.....	70
6.2.3	Codage normalisé des non-réponses	71
6.3	Notation	73
6.3.1	Deuxième notation à l'intérieur d'un pays	73
6.3.2	Deuxième notation entre les pays.....	74
6.4	Saisie, traitement et codage des données	76
6.4.1	Traitement postcollecte (étape 1).....	78
6.4.2	Épuration des données (étape 2).....	79
6.4.3	Vérification préliminaire (étape 3)	81
6.4.4	Vérifications de cheminement (étape 4).....	85
6.4.5	Vérifications de la cohérence (étape 5).....	87
6.4.6	Imputations et variables dérivées (étape 6)	90
6.4.7	Création du fichier principal (étape 7)	134
6.4.8	Création du fichier national de microdonnées à grande diffusion (étape 8)	140
6.4.9	Création du FMGD international (étape 9)	142
7.0	Totalisation et analyse des données	143
7.1	Utilisation de coefficients de pondération dans la totalisation	143
7.2	Définitions de l'estimation catégorique et de l'estimation quantitative	143
7.2.1	Totalisation des estimations catégoriques	144
7.2.2	Totalisation des estimations quantitatives.....	145
7.3	Estimations des niveaux de compétence	145
7.4	Arrondissement des estimations	146
8.0	Qualité des données	149
8.1	Erreurs d'échantillonnage.....	149
8.1.1	Lignes directrices en matière de diffusion du c.v.	150
8.1.2	Utilisation des valeurs plausibles et des poids de rééchantillonnage dans le calcul de l'erreur d'échantillonnage.....	151
8.1.3	Estimation de la variance d'erreur dans le cadre de l'EIACA.....	158
8.1.4	Analyse des données de l'EIACA sur SPSS	166
8.1.5	Analyse des données de l'EIACA sur SAS	178
8.2	Erreurs non liées à l'échantillonnage	184
8.2.1	Base de sondage	184
8.2.2	Non-réponse	184
8.2.3	Erreur dans la réponse.....	185
8.2.4	Notation.....	185

Table des matières (suite)

9.0	Questionnaire	187
9.1	Seuils pour la diffusion des estimations pour l'enquête sur l'alphabétisation et les compétences des adultes.....	187
9.2	Tables de variabilité d'échantillonnage approximative	187
9.3	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	188
9.4	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques.....	190
9.5	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance	193
9.6	Comment utiliser les tables de coefficients de variation pour effectuer un test t	195
9.7	Coefficients de variation pour des estimations quantitatives.....	195
9.8	Coefficients de variation pour les estimations du niveau de compétence.....	196
9.9	Tables des coefficients de variation	198
10.0	Dispositions d'enregistrement et variables uniques	205
11.0	Personnes-ressources.....	207

Table des tableaux

Tableau 1.1	Répartition géographique des répondants à l'EIACA, Canada, provinces et territoires, personnes de 16 ans et plus, 2003	8
Tableau A1	Résolution de problèmes : étapes et mesures à prendre	44
Tableau 4.1	Probabilités moyennes de réussite, échelle des textes suivis	50
Tableau 4.2	Probabilités moyennes de réussite, échelle des textes schématiques	50
Tableau 5.1	Taux de couverture estimatif par province ou territoire, EIACA de 2003	56
Tableau 5.2a	Composition prévue des échantillons avant la collecte, EIACA de 2003	59
Tableau 5.2b	Répartition finale des répondants, EIACA de 2003.....	59
Tableau 5.3	Variables de référence par province ou territoire	64
Tableau 5.4	Échantillon obtenu et taux de réponse par province	65
Tableau 6.1	Fiabilité de la notation à l'intérieur d'un pays par domaine, en pourcentage	74
Tableau 6.2	Fiabilité de la deuxième notation entre les pays	75
Tableau 8.1	Lignes directrices en matière de niveau de qualité	150

1.0 Introduction

Qu'est-ce que l'EIACA?

Réalisée en 2003, l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA) est le volet canadien de l'Enquête sur la littératie et les compétences des adultes (ELCA). L'ELCA est une « [...] étude conjointe à grande échelle menée par des gouvernements, des organismes statistiques nationaux, des établissements de recherche et des organismes multilatéraux » qui produit des mesures comparables à l'échelle internationale dans quatre domaines : la compréhension de textes suivis et de textes schématiques (qui constituent les deux domaines de la littératie), la numératie et la résolution de problèmes¹.

L'Enquête sur la littératie et les compétences des adultes (ELCA)

L'élaboration et la gestion de l'ELCA ont été coordonnées par Statistique Canada et Educational Testing Service (ETS, Princeton, États-Unis) en collaboration avec le National Center for Education Statistics (NCES) du ministère de l'Éducation des États-Unis, l'Organisation de coopération et de développement économiques (OCDE) et l'Institut de statistiques (IS) de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO).

Dans le cadre de l'ELCA, qui a été réalisée au cours de la première moitié de 2003, tous les pays participants² devaient recueillir des données auprès d'un échantillon national représentatif d'au moins 3 000 répondants âgés de 16 à 65 ans pour chaque langue sondée, soit l'anglais et le français pour le Canada. Le Canada a excédé les exigences minimales établies pour les échantillons de l'ELCA, grâce à la contribution de plusieurs organismes fédéraux et gouvernements provinciaux qui a permis la collecte de données supplémentaires afin d'assurer une fiabilité élevée dans l'estimation des valeurs des données pour les petits ensembles d'une population. En outre, contrairement à l'Enquête internationale sur l'alphabétisation des adultes

D'un nom à l'autre

ELCA, EIACA : pourquoi des noms différents? On nous pose souvent cette question. Il existe un certain nombre de raisons pour lesquelles le nom de l'enquête diffère au Canada.

D'une part, un sondage a révélé qu'au Canada, le terme « connaissances utiles à la vie » avait une connotation particulière qui ne cadrerait pas avec les compétences mesurées dans cette enquête. Or, cette connotation risquait d'influer sur les taux de réponse et de biaiser les résultats de l'enquête.

D'autre part, il fallait enregistrer le titre de l'enquête menée au Canada avant que le titre du programme international (Enquête sur la littératie et les compétences des adultes) n'ait été officiellement accepté.

Le titre canadien « Enquête internationale sur l'alphabétisation et les compétences des adultes » fait également ressortir le lien analytique et conceptuel entre ce nouvel ensemble de données et celui de l'EIACA de 1994.

À cela s'ajoute l'avantage de pouvoir distinguer de manière beaucoup plus précise la base de données utilisée (celle de l'EIACA ou celle de l'ELCA).

¹ OECD et Statistique Canada (2005), p. 15.

² Les pays participants comprenaient les Bermudes, le Canada, les États-Unis, l'Italie, la Norvège, la Suisse et l'État mexicain de Nuevo León.

(EIAA) de 1994³, l'EIACA de 2003 a également bénéficié de l'apport des administrations territoriales et d'autres organismes non gouvernementaux. Résultat : le nombre de répondants est suffisant pour établir des estimations précises pour le Yukon, les Territoires du Nord-Ouest et le Nunavut. Enfin, à l'instar de l'EIAA de 1994, l'échantillon de l'EIACA de 2003 comprenait des Canadiens âgés de plus de 65 ans. Au Canada, la collecte de données a commencé en mars 2003 et a pris fin en août. Au terme de cette période de six mois, plus de 23 000 personnes de tout le Canada avaient consacré en moyenne deux heures pour répondre à l'EIACA, contribuant ainsi à la valeur de l'enquête. Le tableau 1.1 montre les répartitions réelle et pondérée des répondants de l'ensemble du Canada.

Tableau 1.1 Répartition géographique des répondants à l'EIACA, Canada, provinces et territoires, personnes de 16 ans et plus, 2003

Province ou territoire	Nombre réel de répondants	Nombre pondéré de répondants
Terre Neuve et Labrador	1 299	431 647
Île du Prince Édouard	645	111 274
Nouvelle-Écosse	1 272	747 446
Nouveau-Brunswick	1 466	599 680
Québec	4 166	5 994 043
Ontario	4 946	9 621 290
Manitoba	2 267	852 805
Saskatchewan	1 234	741 828
Alberta	1 307	2 428 843
Colombie-Britannique	1 849	3 313 116
Yukon	1 092	20 738
Territoires du Nord-Ouest	818	26 541
Nunavut	677	12 592
Canada	23 038	24 901 843

Veillez prendre note : Le fichier de microdonnées à grande diffusion (FMGD) de l'EIACA contient les résultats canadiens de l'ELCA, dont un certain nombre de variables concernant la langue et la scolarité qui n'ont été recueillies dans aucun autre pays participant.

Par contre, le FMGD de l'ELCA contient les résultats du Canada et d'autres pays participants, **mais seulement pour les répondants âgés de 16 à 65 ans** et seulement pour un **sous-ensemble** des variables présentes dans l'ensemble de données de l'EIACA. Autrement dit, le FMGD international contient des données d'autres pays, mais aussi des variables supplémentaires pour le Canada et des répondants supplémentaires (âgés de plus de 65 ans) dans le fichier de données de l'EIACA.

En fait, le volet canadien du FMGD de l'ELCA constitue un sous-ensemble de ce FMGD canadien.

Selon le type d'analyse à effectuer, on peut avoir besoin de l'un ou de l'autre, voire des deux fichiers.

³ L'Enquête internationale sur l'alphabétisation des adultes (EIAA) a été réalisée entre 1994 et 1998 dans 22 pays. Les données canadiennes ont été recueillies en 1994. Les résultats sont présentés dans trois volumes : OCDE et Statistique Canada (1995), OCDE et DRHC (1997) et OCDE et Statistique Canada (2000).

2.0 Aperçu de l'enquête

On a d'abord remis à chaque répondant à l'EIACA un questionnaire demandant des renseignements sur certaines caractéristiques et variables démographiques comme le niveau de scolarité, l'emploi, le revenu ainsi que la participation à des activités d'apprentissage pour adultes et à des activités communautaires.

On a ensuite remis aux répondants un questionnaire psychométrique⁴ validé à l'échelle internationale et conçu pour mesurer les compétences dans quatre domaines :

❖ **Compréhension de textes suivis** – Les connaissances et les compétences requises pour comprendre et utiliser l'information de certains textes, notamment des éditoriaux, des nouvelles, des brochures et des manuels d'instruction.

❖ **Compréhension de textes schématiques** – Les connaissances et les compétences requises pour situer et utiliser l'information contenue dans des documents de divers formats, notamment des demandes d'emploi, des formulaires de paie, des horaires de services de transport, des cartes, des tableaux et des graphiques.

❖ **Numératie** – Les connaissances et les compétences requises pour gérer efficacement les exigences relatives aux notions de calcul de diverses situations.

❖ **Résolution de problèmes** – La réflexion et l'action orientées vers un but dans des situations pour lesquelles aucune solution de routine n'existe. La personne qui cherche à résoudre un problème a défini un objectif de façon plus ou moins précise, mais ne sait pas exactement comment l'atteindre. L'incompatibilité des objectifs et des opérateurs admissibles constitue un problème. La compréhension du problème et sa transformation par étapes, fondée sur la planification et le raisonnement, constituent le processus de résolution du problème.


Connaissances en matière de santé

Dans le cadre de l'EIACA, les domaines de la compréhension de textes suivis, de textes schématiques et de la numératie comprenaient chacun un certain nombre de tâches, dont celle de « la bouteille d'aspirine », qui contenaient de l'information liée à la santé. En recyclant ces items et en les regroupant en un « domaine » cohérent, on obtient un nouvel ensemble de valeurs plausibles qui, théoriquement et empiriquement, mesurent la capacité d'utiliser et de comprendre l'information contenue dans des textes, des diagrammes et des graphiques comprenant de l'information et des tâches liées à la santé.

L'ensemble de données de l'EIACA comprend ces mesures, qu'on peut utiliser et interpréter de la même manière que les domaines de la compréhension de textes suivis, de textes schématiques et de la numératie.

L'EIAA de 1994 et l'EIACA de 2003 ont utilisé les mêmes échelles d'évaluation de la compréhension de textes suivis et de textes schématiques. Pour ces deux domaines, on a lié les échelles de compétence des deux enquêtes en intégrant un sous-ensemble d'items de test utilisés en 1994. Par conséquent, pour plusieurs pays, dont le Canada, on peut comparer la répartition actuelle des niveaux de compréhension de textes suivis et de textes schématiques à celle de 1994 pour en connaître l'évolution.

⁴ La psychométrie est la branche de la psychologie qui traite de la conception, de l'administration et de l'interprétation des tests quantitatifs visant la mesure de variables comme l'intelligence, l'aptitude et les traits de personnalité.



L'échelle de numératie de l'EIACA de 2003 a été établie en fonction du domaine des textes au contenu quantitatif mesuré en 1994, produisant ainsi une mesure plus vaste et englobante des compétences mathématiques et de la connaissance des concepts mathématiques. Cette échelle étendue mesure plus que la simple capacité de réaliser des opérations mathématiques sur des nombres intégrés dans un texte : elle comprend de nombreuses tâches qui requièrent peu ou pas de lecture.

Enfin, l'EIACA comportait des tâches visant à évaluer la capacité de résoudre des problèmes. Ce nouveau domaine, validé dans le cadre d'essais rigoureux, présente des caractéristiques uniques que l'on ne trouve pas dans les autres domaines. Dans une certaine mesure, ce domaine exige l'intégration des connaissances et des compétences mesurées par les domaines de la littératie et de la numératie ainsi que leur application à de nouvelles situations. Il mesure également la capacité d'utiliser des outils logiques de base qui procureront des stratégies de solutions efficaces aux problèmes quotidiens. Ces outils comprennent la capacité de classer, d'évaluer et de prioriser une série de facteurs ainsi que d'établir une distinction, de planifier, d'analyser et de raisonner par rapport à une variété de choix afin d'en arriver à une solution efficace à un problème donné.

Le chapitre qui suit présente la conceptualisation et la définition des quatre domaines, ainsi que des exemples d'items de test utilisés dans les évaluations. Les lecteurs qui ont besoin de renseignements techniques supplémentaires sur les aspects psychométriques de l'Enquête sont priés de consulter la source Statistique Canada (2004).

3.0 Objectifs de l'enquête

L'établissement de l'EIACA avait deux objectifs fondamentaux :

- 1) Le premier objectif consistait à approfondir l'évaluation des compétences mesurées dans le cadre de l'EIAA de 1994 en ajoutant de nouveaux domaines d'évaluation avec des cadres théoriques robustes et des paramètres d'item stables d'un pays à l'autre et d'une langue à l'autre. Par le fait même, cet objectif consistait aussi à lier directement l'EIAA à l'EIACA selon les deux domaines de littératie afin de permettre une comparaison des profils de compréhension de textes suivis et de textes schématiques mesurés en 1994, puis en 2003.
- 2) Le deuxième objectif consistait à mener une analyse internationale, nationale et infra-nationale des corrélats et des antécédents possibles des compétences mesurées au Canada en rassemblant une banque de données assez importante à partir d'un nombre suffisant de répondants.

L'enquête reposait avant tout sur l'évaluation directe des compétences des répondants en littératie, en numératie et en résolution de problèmes au moyen de tâches courantes présentant divers degrés de difficulté et relevant de diverses sphères de connaissances et de compétences. Cette information était soutenue par la collecte de renseignements de base sur les répondants. De plus, le questionnaire de référence comprenait des questions sur l'auto-évaluation des compétences des répondants en littératie et en numératie, sur la formation suivie au cours de l'année précédant l'enquête et sur les obstacles perçus à l'amélioration des niveaux de compétence en littératie ou en numératie.

4.0 Concepts et définitions

4.1 Définir et mesurer les compétences : aperçu

Dans le cadre de l'EIACA, chaque échelle de compétence va de zéro jusqu'à un maximum théorique de 500 points. Une personne dont les connaissances la situent à un échelon donné de l'échelle a 80 % de probabilité de réussir une tâche qui comporte ce même niveau de difficulté. Par exemple, une personne dont le rendement a été évalué à 250 points a 80 % de probabilité de réaliser correctement une tâche dont le niveau de difficulté a été estimé à 250. Cette même personne aurait « 80 % et plus » de probabilité de réaliser correctement une tâche plus simple (environ 95 % pour une tâche dont le niveau de complexité a été établi à 200) et moins de probabilité (moins de 80 %) de réaliser une tâche plus difficile (environ 40 % pour une tâche dont le niveau de complexité a été établi à 300)⁵.

En vue de faciliter l'analyse, on a regroupé ces notes continues en cinq niveaux de compétence (seulement quatre niveaux ont été définis pour l'échelle de la résolution de problèmes), le niveau 1 étant le niveau le plus faible. Les niveaux de compétence utilisés dans l'EIACA sont utiles pour résumer les résultats, mais comportent également des limites. D'abord, la proportion relativement faible de répondants qui atteignent le niveau 5 ne permet pas toujours d'établir des constatations exactes. C'est pourquoi, chaque fois que les résultats sont présentés par niveau de capacité, les niveaux 4 et 5 sont combinés. Ensuite, les niveaux indiquent des ensembles particuliers de compétences et, par conséquent, les seuils de ces niveaux ne sont pas équidistants. Par conséquent, les fourchettes de résultats pour chaque niveau ne sont pas identiques. En fait, pour les quatre domaines, le niveau 1 englobe presque la moitié de l'échelle. Les seuils du domaine de la résolution de problèmes sont établis de façon quelque peu différente, et le niveau 1 couvre exactement la moitié de l'échelle. Ce niveau comprend toutes les compétences de base requises pour atteindre des niveaux plus élevés. En d'autres mots, la capacité de lire peut se situer quelque part au niveau 1, mais la capacité de comprendre et d'utiliser ce qui a été lu affiche un accroissement de complexité allant du niveau 1 au niveau 5. Les fourchettes relativement larges des résultats du niveau 1 pour chacune des échelles entraînent donc des sous-niveaux multiples de compétences à ce niveau. Ainsi, la fourchette comprend les personnes qui peuvent à peine lire de même que celles qui lisent mal ou avec un manque d'attention⁶.

Littératie et analphabétisme

Il est intéressant de noter que, si la probabilité d'une bonne réponse se rapproche de zéro à mesure que la tâche devient plus difficile, cette probabilité ne peut jamais vraiment atteindre ce niveau parce qu'il y a toujours des chances, même minimes, de répondre correctement, quel que soit le niveau de capacité.

C'est pourquoi les résultats de l'EIACA mesurent le rendement selon un continuum de compétences. Les échelles ne mesurent pas l'absence d'une compétence et ne distinguent pas les personnes qui possèdent une compétence donnée de celles qui ne la possèdent pas.

Le présent chapitre propose un bref aperçu des cadres de travail adoptés pour mettre au point et interpréter les échelles qui ont servi à mesurer la compréhension de textes suivis et de textes

⁵ Kirsch, Jungeblut et Campbell (1992), p. 14-15.

⁶ L'Enquête internationale sur les compétences en matière de lecture est un suivi de l'EIACA de 2003 qui fournira davantage de renseignements sur les répondants du niveau 1. Les résultats sont attendus pour 2006.

schématiques, la numératie et la résolution de problèmes lors de l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA). L'élaboration d'un cadre de travail semble revêtir une importance capitale dans une approche conceptuelle de la mesure. Entre autres éléments, le cadre de travail doit comprendre une définition convenue de ce qu'il faut mesurer et le choix des caractéristiques à retenir dans la conception et l'interprétation des tâches. Dans le présent chapitre, nous décrivons ces caractéristiques pour chaque mesure, mais nous donnons également des exemples d'item et précisons les caractéristiques qui, à la lumière des résultats, contribuent à la difficulté des items. Ensemble, ces renseignements fournissent un moyen pour passer d'une interprétation des résultats d'enquête centrée sur des tâches prises isolément ou sur un seul nombre, à une interprétation permettant de définir des niveaux de capacité assez généralisés pour être valables à l'égard de l'ensemble des évaluations et des groupes.

4.2 Comprendre ce qui a été mesuré lors de l'Enquête internationale sur l'alphabétisation et les compétences des adultes

4.2.1 Introduction

En 1992, l'Organisation de coopération et de développement économiques (OCDE) a conclu que la faiblesse des niveaux de littératie constituait, à l'échelle internationale, une grave menace pour les résultats économiques et la cohésion sociale (OCDE, 1992). Or, l'absence de données internationales comparables a compliqué l'analyse des problèmes de littératie observés dans les pays industrialisés, et empêché les décideurs d'en tirer des leçons. Statistique Canada et Educational Testing Service (ETS) ont collaboré pour mettre au point et mener une étude comparative internationale de la littératie.

L'Enquête internationale sur l'alphabétisation des adultes (EIAA) était la première enquête comparative menée auprès d'adultes pour étudier la répartition de la littératie parmi les pays participants. En 2000, on a publié dans un rapport final (OCDE et Statistique Canada, 2000) les résultats de trois cycles d'évaluations auxquels avaient participé quelque 23 pays ou groupes linguistiques représentant un peu plus de 50 % du PIB mondial. Si l'EIAA a jeté une base importante pour les enquêtes comparatives internationales menées auprès des adultes, on a cependant exprimé le besoin d'ajouter des éléments à mesurer. Les administrations publiques et les décideurs tenaient de plus en plus à savoir quelles autres compétences étaient nécessaires à une personne pour participer pleinement et fructueusement à une société moderne, et à la société pour relever les défis d'un monde en évolution rapide. Un projet conçu à cette fin, intitulé *Définition et sélection des compétences* (DeSeCo), a été mené sous l'égide de la Suisse. Son objectif consistait à définir, d'un point de vue théorique, un ensemble de compétences clés qui sont essentielles à une vie réussie et à la bonne marche de la société (Rychen et Salganik, 2003).

En réponse au souci d'élargir la mesure des compétences, les responsables de la mise au point de l'EIACA ont entrepris de définir des cadres de travail qui serviraient à effectuer de nouvelles mesures lors des évaluations comparatives des adultes. Ils étaient conscients du fait que la conception d'un instrument valide et fiable reposait avant tout sur des bases théoriques solides, représentées par un cadre de travail qui reflète la pensée courante dans le domaine. Selon Messick (1994), un cadre de travail qui adopte une approche conceptuelle du plan d'une

évaluation doit commencer par une définition ou un énoncé de principe général, qui précise la raison d'être de l'enquête et ce qu'elle vise à mesurer en fait de connaissances, de compétences ou d'autres attributs, puis définir divers rendements ou comportements qui reflètent ces concepts et, enfin, définir diverses caractéristiques des tâches et indiquer comment on utilisera ces caractéristiques pour concevoir les tâches qui feront ressortir ces comportements.

4.2.2 Échelonner les tâches de littératie, de numératie et de résolution de problèmes dans le cadre de l'EIACA

Les résultats de l'EIACA sont présentés sur quatre échelles – deux échelles de littératie (textes suivis et textes schématiques), une échelle de numératie et une échelle de la résolution de problèmes –, chacune allant de 0 à 500 points. On peut imaginer ces tâches disposées le long de leur échelle respective en fonction de leur difficulté pour les adultes et du niveau de capacité nécessaire pour accomplir correctement chaque tâche. Dans l'EIACA, la méthode employée pour modéliser ces continuums de difficulté et de capacité est la théorie de la réponse d'item (TRI). La TRI est un modèle mathématique servant à estimer la probabilité, pour une personne donnée, d'accomplir correctement une tâche donnée tirée d'une banque de tâches (Murray, Kirsch et Jenkins, 1998).

Sur l'échelle, la valeur attribuée à chaque item est liée au rendement d'échantillons représentatifs d'adultes des pays participants à l'égard de cet item. Elle repose sur la théorie selon laquelle une personne qui se situe à un point donné de l'échelle est capable d'accomplir, avec la même compétence, toutes les tâches qui se situent à ce point de l'échelle. Pour les besoins de l'EIACA, comme pour ceux de l'EIAA, on a déterminé qu'une personne qui se situe à un point donné de l'échelle de capacités avait 80 % des chances de répondre correctement aux items qui se situent à ce point.

Tout comme les adultes au sein de chaque pays participant à l'EIACA sont échantillonnés à partir de la population d'adultes vivant dans des ménages, chaque tâche conçue et utilisée dans l'évaluation représente un type de tâche échantillonnée à partir du domaine ou du concept défini ici. Elle est donc représentative d'un type donné de tâche de littératie, de numératie ou de résolution de problèmes qui est lié à des contextes pour adultes.

Lorsqu'on regarde la répartition des tâches le long de chacune des échelles, une question évidente se pose : qu'est-ce qui distingue les tâches qui se situent au bas de chaque échelle de celles qui se situent au milieu et au sommet de l'échelle? Les tâches qui se situent à peu près au même endroit sur chaque échelle ont-elles en commun un ensemble de caractéristiques qui leur confère un niveau de difficulté semblable? Même un examen superficiel des items révèle que les tâches qui se situent au bas de chaque échelle diffèrent de celles qui se situent au sommet.

Et les connaissances en matière de santé?

On peut également analyser selon ce même continuum une cinquième échelle, étroitement liée à la compréhension de textes suivis et de textes schématiques, celle des connaissances en matière de santé. Toutefois, comme elle a été créée après la collecte de données, son cadre de travail n'intègre pas encore systématiquement les détails empruntés à chacun des domaines donneurs. Comme point de départ, on entend par connaissances en matière de santé la capacité d'utiliser et de comprendre l'information imprimée liée à la santé et provenant de diverses sources. Par conséquent, les facteurs de difficulté des items de compréhension de textes suivis, de textes schématiques ou de numératie réutilisés dans cette nouvelle échelle contribuent également à sa difficulté.

Afin de représenter cette progression de la complexité et de la difficulté, on a divisé chaque échelle de capacités en niveaux. Les échelles de littératie et de numératie comportent cinq niveaux de capacités allant du niveau 1 (le plus faible) au niveau 5 (le plus élevé). Ces niveaux sont définis comme suit : niveau 1 (0 à 225), niveau 2 (226 à 275), niveau 3 (276 à 325), niveau 4 (326 à 375) et niveau 5 (376 à 500). L'échelle de la résolution de problèmes comporte quatre niveaux de capacités allant du niveau 1 (le plus faible) au niveau 4 (le plus élevé). Ces quatre niveaux sont définis comme suit : niveau 1 (0 à 250), niveau 2 (251 à 300), niveau 3 (301 à 350) et niveau 4 (351 à 500).

Comme chaque niveau représente une progression des connaissances et des compétences, les personnes qui se situent à un niveau donné possèdent non seulement les connaissances et les compétences liées à ce niveau, mais aussi les compétences liées aux niveaux inférieurs. En pratique, cela signifie que les personnes dont le rendement est de 250 (le milieu du niveau 2 sur l'une des échelles de littératie ou de numératie) devraient pouvoir accomplir les tâches moyennes des niveaux 1 et 2 avec un degré de compétence élevé. Sur l'échelle de la résolution de problèmes, un point comparable serait 275. Dans l'EIACA, comme dans l'EIAA, on définit un degré de compétence élevé en fonction d'une probabilité de réponse de 80 (PR80). Cela signifie que les personnes qui se situeraient à un niveau précis de l'échelle réaliseraient les tâches à ce niveau avec une probabilité de réussite de 80 %. Cela signifie également que la probabilité qu'elles exécutent les tâches situées à un niveau de capacité plus faible sur l'échelle est supérieure à 80 %, mais pas que les personnes possédant un niveau de capacité donné ne pourront jamais effectuer des tâches dont le niveau de difficulté est plus élevé; elles pourraient réussir de temps à autre. Cela signifie que la probabilité de succès est « relativement » faible, c'est-à-dire que plus le degré de difficulté de la tâche est élevé par rapport à leur niveau de capacité, plus la probabilité de bonne réponse est faible.

On peut expliquer ce principe à l'aide d'une analogie. La relation qui existe entre la difficulté de la tâche et la compétence de la personne est comparable à l'épreuve du saut en hauteur dans une compétition d'athlétisme : l'athlète essaie de sauter par-dessus une barre placée de plus en plus haut. Chaque athlète est compétent à une hauteur particulière. À cette hauteur, la probabilité de réussite est élevée et il peut aussi sauter presque tout le temps par-dessus la barre placée à des niveaux plus bas. Cependant, lorsque la barre est placée à un niveau plus élevé que celui où l'athlète est compétent, on ne s'attend pas à ce qu'il puisse réussir de manière constante.

4.3 Mesurer la compréhension de textes suivis et de textes schématisés dans le cadre de l'EIACA

Financée par le National Center for Education Statistics (NCES) dans le cadre de son programme global d'évaluation de la littératie des adultes, la National Adult Literacy Survey (NALS) est l'étude la plus vaste et la plus complète de la littératie des adultes jamais menée aux États-Unis (Kirsch *et coll.*, 1993). Comme toutes les évaluations à grande échelle financées par le NCES, la NALS a été conçue par un comité composé d'universitaires, de praticiens et d'administrateurs reconnus à l'échelle nationale, qui ont adopté la définition suivante de la littératie :

« La littératie est la capacité d'utiliser des imprimés et des écrits nécessaires pour fonctionner dans la société, atteindre ses objectifs, parfaire ses connaissances et accroître son potentiel. »

Cette définition résulte du travail initial du comité d'orientation de l'évaluation et peut servir de base pour créer d'autres aspects du cadre de travail à aborder. Approuvée et adoptée par les pays qui ont participé au premier cycle de l'EIAA, elle a également été retenue pour les besoins de l'EIACA. Elle comporte certaines hypothèses formulées par les membres du comité; il importe donc d'envisager tour à tour les divers éléments de cette définition.

Tout d'abord, « **La littératie est...** » : ici, on emploie le terme « littératie », de préférence à « lecture », parce qu'il est susceptible de communiquer plus précisément à un public profane ce que l'enquête mesure. Par « lecture », on entend souvent le simple fait de décoder ou lire à voix haute, alors que les enquêtes menées auprès des adultes visent à mesurer une réalité plus vaste et plus profonde. Les chercheurs qui étudient la littératie dans certains contextes ont observé que, selon les cultures et les groupes, on semblait privilégier différentes sortes de pratiques en matière de littératie (Sticht, 1975; Heath, 1980; Szwed, 1981). Heath, par exemple, observe que la pratique de la lecture peut constituer une fin en soi ou servir de mode d'interaction sociale, de moyen d'information, d'aide-mémoire, de substitut de messages oraux, de moyen d'archivage ou de confirmation personnelle. Le fait qu'on lise différents documents dans des buts différents suppose un éventail de capacités dont ne témoigne pas nécessairement le fait de signer son nom, de compter un certain nombre d'années de scolarité ou d'obtenir un résultat du niveau de la 8^e année lors d'un test scolaire de compréhension de lecture.

L'expression « ... **la capacité d'utiliser des imprimés et des écrits** » attire l'attention sur le fait que les membres du comité ne considèrent pas la littératie comme un ensemble de capacités isolées liées à la lecture et à l'écriture, mais – ce qui est plus important – comme l'application de ces capacités à des fins précises dans des contextes précis. Lorsqu'on étudie la littératie dans des contextes divers, elle devient marquée au sceau de la diversité. Premièrement, les gens pratiquent la littératie pour répondre à diverses fins ou à divers besoins (Sticht, 1978; Heath, 1980); Cook-Gumperz et Gumperz, 1981; Mikulecky, 1982). Ces utilisations varient selon les contextes (Heath, 1980; Venezky, 1983) et selon les personnes à l'intérieur du même contexte (Kirsch et Guthrie, 1984a). Cette variation de l'utilisation amène les gens à lire un vaste éventail de documents ayant des formes linguistiques qualitativement très différentes (Diehl, 1980; Jacob, 1982; Miller, 1982). Dans certains cas, on a lié ces différents types de tâches de littératie à différentes stratégies cognitives ou habitudes de lecture (Sticht, 1978, 1982; Crandall, 1981; Scribner et Cole, 1981; Kirsch et Guthrie, 1984b).

L'expression « ... **pour fonctionner dans la société, atteindre ses objectifs, parfaire ses connaissances et accroître son potentiel** » vise à englober toutes les situations dans lesquelles la littératie joue un rôle dans la vie des adultes, en privé et en public, de l'école au travail, puis à l'acquisition continue du savoir et à la participation active à la vie de la collectivité. « Atteindre ses objectifs, parfaire ses connaissances et accroître son potentiel » exprime le point de vue selon lequel la littératie permet de combler les aspirations personnelles – celles qui sont définies, comme l'obtention d'un diplôme ou d'un emploi, et celles qui sont moins définies et moins immédiates, qui enrichissent la vie d'une personne. L'expression « dans la société » tient compte du fait que la littératie offre à chacun le moyen de contribuer à la vie de la société tout en en tirant des avantages. On s'accorde à reconnaître que les compétences en littératie sont importantes pour permettre à un pays de maintenir ou d'améliorer son niveau de vie et de s'imposer sur un marché concurrentiel de plus en plus mondial. Pourtant, elles sont tout aussi importantes pour la participation individuelle à la vie d'une société caractérisée par l'évolution

technologique avec ses institutions structurées, un système juridique complexe et de vastes programmes gouvernementaux.

4.3.1 Définir les caractéristiques des tâches

Les caractéristiques des tâches représentent des variables qu'on peut utiliser de diverses façons pour mettre au point une évaluation et en interpréter les résultats. Selon Almond et Mislevy (1998), les variables peuvent jouer l'un des cinq rôles suivants : limiter la portée de l'évaluation, définir les caractéristiques à utiliser pour concevoir les tâches, contrôler l'assemblage des tâches en livrets ou en formulaires de test, caractériser le rendement des répondants ou leurs réponses aux tâches ou caractériser les aspects des compétences ou des capacités. L'EIAA s'est concentrée sur des variables pouvant servir à concevoir des tâches ainsi qu'à caractériser le rendement selon une ou plusieurs échelles de capacités.

Chaque tâche de l'évaluation témoigne d'un aspect de la littératie d'une personne (Mislevy, 2000). Si le but de l'évaluation consiste à broser un tableau aussi fidèle que possible des compétences et des capacités d'une personne, le test ne peut pas comprendre un nombre infini de tâches et l'on ne peut pas manipuler un nombre infini de caractéristiques de ces tâches. Il faut donc faire des choix au sujet des caractéristiques à inclure dans le processus de mise au point du test. Pour les besoins de la construction des tâches de l'EIAA, on a retenu les trois caractéristiques suivantes :

Contextes ou contenu pour adultes. Comme les adultes ne lisent pas d'écrits ou d'imprimés au hasard, mais plutôt dans un contexte donné ou pour répondre à un besoin particulier, on choisit, en vue de l'évaluation de la littératie, des documents qui représentent divers contextes et divers contenus. On s'assure ainsi qu'aucun groupe d'adultes n'est avantagé ou désavantagé en raison du contexte ou du contenu retenu pour l'évaluation. On a retenu les six catégories suivantes de contexte ou de contenu pour adultes :

- ❖ **Vie familiale** : documents portant sur les relations interpersonnelles, les finances personnelles, le logement et l'assurance.
- ❖ **Santé et sécurité** : documents portant sur les drogues et l'alcool, la prévention et le traitement des maladies, la sécurité et la prévention des accidents, les premiers soins, les urgences et les mesures à prendre pour rester en santé.
- ❖ **Vie sociale et collective** : documents portant sur les ressources communautaires et les moyens d'information.
- ❖ **Économie de la consommation** : documents portant sur le crédit et les opérations bancaires, l'épargne, la publicité, les achats et les biens personnels.
- ❖ **Travail** : documents portant sur diverses professions en général (mais pas sur des cas particuliers), la recherche d'un emploi, les finances et la vie professionnelle.
- ❖ **Loisirs et détente** : documents portant sur les voyages, les activités récréatives et les restaurants.

Documents ou textes. S'il ne fait aucun doute qu'une évaluation de la littératie doit porter sur des documents divers, la diversité et les caractéristiques spécifiques des textes qui entrent dans la conception des tâches sont essentielles à la conception et à l'interprétation des résultats qui sont produits. Pour les besoins de l'EIACA, on a établi une distinction essentielle entre les textes continus et les textes non continus. Ordinairement, les textes continus sont

formés de phrases organisées en paragraphes. Dans ces textes, l'organisation tient à la disposition des paragraphes, aux alinéas et à la ventilation du texte en une hiérarchie signalée par des rubriques qui permettent au lecteur de reconnaître l'organisation du texte. Les textes sont classés le plus souvent selon le but de l'auteur ou le type de texte. Pour l'EIAA, ces catégories sont les suivantes : exposition, description, argumentation et instructions.

Les textes non continus sont organisés différemment des textes continus, de sorte qu'ils permettent au lecteur d'employer des stratégies différentes pour y accéder et en extraire de l'information. En surface, ces textes semblent présenter de nombreux principes d'organisation ou formats différents – tableaux, horaires, diagrammes, graphiques, cartes géographiques, formulaires, etc. Toutefois, on dit que le principe d'organisation de ces types de texte, que Mosenthal et Kirsch (1998) appellent textes schématiques, présente l'une des quatre structures de base suivantes : liste simple, liste combinée, liste croisée ou liste imbriquée. Ensemble, ces quatre types de document constituent ce que les auteurs appellent des documents matriciels, soit des textes non continus présentant des rangées et des colonnes bien définies. Ils sont aussi étroitement liés à d'autres textes non continus que ces auteurs appellent documents graphiques, documents localifs et documents à remplir.

Deux des trois échelles de littératie utilisées dans l'EIAA reposaient sur la distinction entre textes continus et textes non continus. Les tâches situées le long de l'échelle des textes suivis étaient axées sur des textes continus, alors que celles situées le long de l'échelle des textes schématiques étaient axées sur des textes non continus. L'échelle des textes au contenu quantitatif comprenait à la fois des textes continus et des textes non continus. La caractéristique distinctive de cette échelle tenait au fait que les répondants devaient cerner et exécuter une ou plusieurs opérations arithmétiques fondées sur l'information contenue dans les textes. Dans l'EIACA, on a remplacé cette échelle par l'échelle de la numératie, dont nous traiterons en détail plus loin dans la présente annexe.

Processus ou stratégies. Cette caractéristique des tâches a trait à la façon dont le répondant traite le texte pour répondre correctement à une question ou à une directive. Elle comprend les processus utilisés pour rapprocher l'information énoncée dans la question (l'information donnée) de l'information nécessaire contenue dans le texte (la nouvelle information), ainsi que les processus nécessaires pour repérer ou formuler la bonne réponse à partir de l'information disponible. Nous envisageons ici trois variables retenues pour examiner les tâches de lecture ou de littératie utilisées dans les enquêtes nationales et internationales : le type d'appariement, le type d'information demandée et la plausibilité des éléments de distraction.

4.3.2 Type d'appariement

On distingue quatre types de stratégie d'appariement : le repérage, le regroupement des caractéristiques, l'intégration et la formulation. Pour les tâches de **repérage**, le répondant doit appairer un ou plusieurs éléments d'information énoncés dans la question à des renseignements identiques ou synonymiques présents dans le texte. Les tâches de **regroupement des caractéristiques** invitent également le répondant à appairer un ou plusieurs éléments d'information; elles diffèrent cependant des tâches de repérage du fait que le répondant doit appairer une série de caractéristiques pour satisfaire aux conditions énoncées dans la question.

Les tâches **d'intégration** demandent au répondant de rapprocher deux ou plusieurs éléments d'information tirés du texte selon un type de relation donnée. Par exemple, cette relation peut amener le répondant à repérer des similitudes (établir une comparaison), des différences (mettre en opposition), un degré (trouver plus petit ou plus grand) ou des relations de cause à effet. Cette information peut se trouver dans un seul paragraphe ou figurer dans différents paragraphes ou différentes sections du texte. Lorsqu'il intègre l'information, le répondant s'inspire des catégories d'information données dans une question pour repérer l'information correspondante dans le texte. Il peut alors rapprocher l'information énoncée dans le texte à ces différentes catégories en fonction du terme de relation précisé dans la question. Dans certains cas, toutefois, le répondant doit **formuler** ces catégories ou relations avant d'intégrer l'information énoncée dans le texte.

Outre qu'il demande au répondant d'appliquer l'une de ces quatre stratégies, le type d'appariement entre une question et le texte est influencé par plusieurs autres conditions de traitement qui contribuent à la difficulté globale d'une tâche. La première de ces conditions est le nombre d'expressions à utiliser dans la recherche. La difficulté d'une tâche augmente avec la quantité d'information énoncée dans la question et que le répondant doit chercher dans le texte. Par exemple, les questions qui comportent une seule proposition indépendante sont habituellement plus simples, en moyenne, que celles qui contiennent plusieurs propositions, indépendantes ou non. La difficulté augmente aussi avec le nombre de réponses qu'on demande au répondant de donner. Les questions qui demandent une seule réponse sont plus simples que celles qui en demandent trois ou plus. En outre, les questions qui précisent le nombre de réponses demandées sont habituellement plus simples que celles qui ne le font pas. Par exemple, une question qui énonce « Énumérez les trois raisons... » est plus simple qu'une question qui énonce « Énumérez les raisons... ». Les tâches sont également influencées par la mesure dans laquelle le répondant doit faire des déductions pour appairer l'information donnée dans une question à l'information correspondante dans le texte et pour trouver l'information demandée.

4.3.3 Type d'information demandée

Il s'agit de la sorte d'information que le lecteur doit repérer pour répondre correctement à une question du test. Plus l'information demandée est concrète, plus la tâche est jugée facile. Dans les travaux de recherche antérieurs portant sur les évaluations à grande échelle de la littératie des adultes et des enfants (Kirsch et Mosenthal, 1994; Kirsch, Jungeblut et Mosenthal, 1998), la variable « type d'information » était notée sur une échelle à cinq points. La cote 1

représentait l'information la plus concrète, donc la plus simple à traiter, et la cote 5, l'information la plus abstraite, donc la plus difficile à traiter.

Par exemple, les questions où le répondant devait trouver une personne, un animal ou une chose (soit des noms imaginables) demandaient une information très concrète; on leur attribuait donc une valeur de 1. Celles où le répondant devait trouver des objectifs, des conditions ou des buts demandaient des types d'information plus abstraits. Comme elles étaient jugées plus difficiles, on leur attribuait une valeur de 3. Les questions qui demandaient au répondant de trouver un « équivalent » étaient jugées les plus abstraites; on leur attribuait donc une valeur de 5. Dans ce dernier cas, l'équivalent était habituellement un terme ou une expression que le répondant ne connaissait pas et pour lequel il devait déduire une définition ou une interprétation à partir du texte.

4.3.4 Plausibilité des éléments de distraction

Il s'agit de la mesure dans laquelle l'information contenue dans le texte possède une ou plusieurs caractéristiques en commun avec l'information demandée dans la question, mais sans satisfaire entièrement aux exigences de la question. Les tâches sont jugées le plus simples lorsque le texte ne comporte aucun élément de distraction. Elles ont tendance à devenir plus difficiles à mesure que le nombre d'éléments de distraction augmente, que les éléments de distraction possèdent plus de caractéristiques en commun avec la bonne réponse et qu'ils figurent plus près de la bonne réponse. Par exemple, les tâches sont habituellement jugées plus difficiles lorsqu'un ou plusieurs éléments de distraction répondent à une partie, mais non à la totalité, des conditions énoncées dans la question et qu'ils figurent dans un paragraphe ou une section du texte autre que celui ou celle qui contient la bonne réponse. Les tâches sont jugées le plus difficiles lorsque deux ou plusieurs éléments de distraction possèdent la plupart des caractéristiques de la bonne réponse et qu'ils figurent dans le même paragraphe ou bloc d'information que la bonne réponse.

4.3.5 Caractériser les tâches de compréhension de textes suivis

L'échelle de compréhension de textes suivis comporte 55 tâches ordonnées le long de l'échelle de 500 points; de ce nombre, 19 tâches sont tirées de l'EIAA et 36 nouvelles tâches ont été conçues et mises au point pour les besoins de l'EIACA. L'indice de difficulté de ces tâches varie de 169 à 439. L'une des tâches les plus simples (indice de difficulté : 188; niveau 1) consiste, pour le répondant, à lire une étiquette de médicament afin de déterminer le nombre maximal de jours de prise du médicament. En fonction des variables retenues, le type d'appariement est jugé simple parce que le lecteur devait repérer un seul élément d'information qui est énoncé textuellement sur l'étiquette du médicament. L'étiquette ne fait qu'une seule référence au nombre de jours, et ce renseignement figure sous la rubrique « Posologie ». Le type d'information est jugé simple parce qu'on demandait un certain nombre de jours et la plausibilité des éléments de distraction est jugée simple parce qu'il n'y a pas d'autre mention d'un nombre de jours sur l'étiquette du médicament.

ASPIRIN MEDCO

500

INDICATIONS THÉRAPEUTIQUES : Maux de tête, douleurs musculaires et rhumatismales, maux de dents et d'oreilles. SOULAGE LES SYMPTÔMES COMMUNS DU RHUME.

POSOLOGIE : ORALE. Prendre 1 ou 2 comprimés toutes les 6 heures, de préférence en mangeant, pendant au plus 7 jours. Conserver en un endroit frais et sec.

MISE EN GARDE : Ne pas prendre en cas de gastrite ou d'ulcère gastro-duodéal. Éviter de prendre en même temps qu'un anticoagulant, ou en cas de maladie du foie ou d'asthme bronchique grave. Si ce médicament est pris à grosse dose pendant une période prolongée, il peut affecter les reins. Avant d'administrer à un enfant atteint de varicelle ou de grippe, consulter un médecin au sujet du syndrome de Reyes, maladie rare mais grave. Les femmes enceintes et celles qui allaitent doivent consulter leur médecin avant d'utiliser ce produit, surtout durant le troisième trimestre de la grossesse. En cas de symptômes persistants ou de surdose accidentelle, consulter un médecin. Garder hors de la portée des enfants.

INGRÉDIENTS : Chaque comprimé contient 500 mg d'acide acétylsalicylique.
Excipient c.b.p. 1 comprimé.
N° d'enreg. 88246

Fabriqué au Canada par STERLING PRODUCTS, INC.
1600, boul. Industriel, Montréal (Québec) H2J 3P1

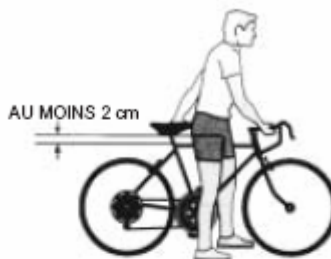
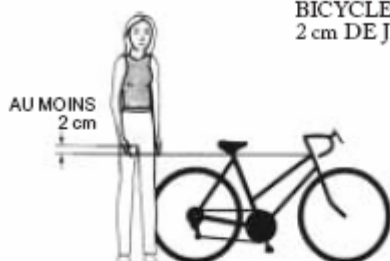


Reproduit avec autorisation

Une deuxième tâche de compréhension de textes suivis demande au lecteur de lire un article sur les impatientes. Cette tâche se situe au milieu du niveau 2 et comporte un indice de difficulté de 254. Elle demande au lecteur ce qu'il peut déduire de la surface lisse des feuilles et des tiges de cette plante. Encore une fois, la tâche amenait le lecteur à repérer une information contenue dans le texte; elle est donc jugée simple en ce qui concerne le type d'information. La dernière phrase du deuxième paragraphe, sous la rubrique *Apparence*, énonce : « La surface lisse des feuilles et la nature des tiges montrent que cette plante a grand besoin d'eau. » Le type d'information est jugé moyen parce que la tâche demandait au lecteur de trouver un état. La plausibilité des éléments de distraction est également jugée moyenne parce que le même paragraphe contenait une phrase qui pouvait distraire un certain nombre de lecteurs. Cette phrase énonce : « Les tiges (...) sont ramifiées et très juteuses, ce qui, en raison de l'origine tropicale de cette plante, la rend très fragile au froid. »

CHOIX D'UN CADRE APPROPRIÉ

LE CYCLISTE DOIT POUVOIR ENFOURCHER SA BICYCLETTE ET, EN POSITION DEBOUT, AVOIR AU MOINS 2 cm DE JEU



NOTA : Pour les femmes, déterminer la mesure à partir d'un modèle pour hommes.

GRANDEUR APPROPRIÉE DE LA BICYCLETTE

GRANDEUR DU CADRE	LONGUEUR DE JAMBE DU CYCLISTE
430mm	660mm-760mm
460mm	690mm-790mm
480mm	710mm-790mm
530mm	760mm-840mm
560mm	790mm-860mm
580mm	810mm-890mm
635mm	860mm-940mm

AU-DESSUS DE LA TIGE HORIZONTALE.

RESPONSABILITÉS DU PROPRIÉTAIRE

1. Choix et achat de la bicyclette : Voir si la bicyclette convient au futur propriétaire. Les bicyclettes sont de grosseur variable. Pour assurer une sécurité et un confort optimaux, la selle et les guidons doivent être ajustés. Comme les bicyclettes sont dotées d'une grande variété d'équipement et d'accessoires ..., il faut veiller à ce que le cycliste sache comment s'en servir.

2. Assemblage : Suivre attentivement les directives d'assemblage. Veiller à ce que tous les écrous, boulons et vis soient bien serrés.

3. Réglage de la bicyclette : Pour que le cycliste puisse rouler en toute sécurité et de façon confortable, la bicyclette doit être bien ajustée. La hauteur de la selle doit être réglée de manière que, lorsqu'il a le pied posé à plat sur la pédale au plus bas de sa course, le cycliste ait le genou légèrement plié.

Nota : Le tableau de gauche permet de déterminer la grandeur de cadre appropriée.

Le fabricant n'est pas responsable des défaillances, blessures ou dommages causés par un assemblage incomplet ou par un mauvais entretien après expédition.

Les tâches qui se situent à des niveaux plus élevés sur l'échelle présentent au lecteur des exigences plus variées en fonction du type d'appariement demandé ainsi que du nombre et de la nature des éléments de distraction présents dans le texte. L'une de ces tâches (indice de difficulté : 281; début du niveau 3) demande au lecteur de consulter une page d'un manuel d'entretien de bicyclette et de trouver comment s'assurer que le siège est dans la bonne position. Le type d'information est jugé moyen parce que le lecteur devait trouver et énoncer par écrit deux conditions à respecter. De plus, on ne précisait pas combien de caractéristiques il devait trouver parmi celles qui étaient énoncées. Le type d'information est également jugé moyen parce qu'il s'agissait de trouver une condition; enfin, on a attribué à la plausibilité des éléments de distraction une cote indiquant qu'elle était relativement simple.

Une tâche un peu plus difficile (318), qui se situe près du sommet du niveau 3, porte sur un article concernant les couches en coton. On demande au lecteur d'énumérer trois raisons pour lesquelles l'auteur préfère les couches en coton aux couches jetables. Cette tâche est relativement difficile en raison de plusieurs variables. Premièrement, le type d'appariement est jugé difficile parce que le lecteur devait donner plusieurs réponses en faisant des déductions à

partir du texte. L'auteur ne dit nulle part dans le texte : « Je préfère les couches en coton parce que... ». Ces déductions sont un peu plus difficiles parce que le type d'information demandée est une « raison » plutôt qu'une réponse concrète. Cette variable est également jugée difficile à cause de son côté abstrait. Enfin, la plausibilité des éléments de distraction est jugée moyenne parce que le texte contient de l'information pouvant distraire le lecteur.

Une autre tâche, qui se situe au niveau 4 sur l'échelle des textes suivis (338), demande au lecteur d'utiliser l'information contenue dans un dépliant sur les entrevues d'emploi et de décrire dans ses mots une différence entre l'entrevue par jury et l'entrevue en groupe. Ici, la difficulté ne tient pas au repérage de l'information dans le texte. Au lieu de simplement repérer un fait concernant chaque type d'entrevue, le lecteur doit intégrer ce qu'il a lu pour trouver une caractéristique qui distingue les deux types d'entrevue. L'expérience retenue d'autres évaluations de ce genre révèle que les tâches demandant au lecteur de mettre en opposition des éléments d'information sont plus difficiles, en moyenne, que celles pour lesquelles on lui demande de trouver des similitudes. Le type d'appariement est donc jugé complexe et difficile. Le type d'information est également jugé difficile parce qu'on demandait au lecteur de trouver une différence. Les différences sont habituellement plus abstraites puisqu'il s'agit de repérer des caractéristiques distinctives ou opposées liées, dans ce cas, à un processus d'entrevue. La plausibilité des éléments de distraction est jugée simple parce que le texte ne contenait aucun élément de distraction. On n'estimait donc pas que cette variable contribuait à la difficulté globale de la tâche.

L'entrevue d'emploi

Avant l'entrevue

Essayez de bien vous renseigner sur l'entreprise. Quels produits fabrique-t-elle ou quels services offre-t-elle? Quelles méthodes ou procédés utilise-t-elle? Ces renseignements peuvent être trouvés dans des annuaires professionnels, des répertoires de chambre de commerce ou d'industrie, ou à votre bureau d'emploi local.

Renseignez-vous aussi sur le poste. S'agit-il de remplacer une personne ou de combler un poste nouvellement créé? Dans quels services ou ateliers travailleriez-vous? Vous pouvez obtenir de la plupart des bureaux d'emploi locaux des conventions collectives qui décrivent divers postes et fonctions uniformisés. Vous pouvez aussi communiquer avec l'organisation syndicale appropriée.

À l'entrevue

Posez des questions au sujet du poste et de l'entreprise. Répondez clairement et avec précision à toutes les questions qui vous sont posées. Apportez un bloc-notes ainsi que vos documents de travail et de formation.

Les genres d'entrevue les plus courants

Individuelle : Se passe d'explication.

Par jury : Un certain nombre de personnes vous posent des questions et comparent ensuite leurs opinions sur votre candidature.

En groupe : Après avoir écouté un exposé sur le poste et les fonctions en compagnie d'autres candidats, vous participez à une discussion en groupe.

Après l'entrevue

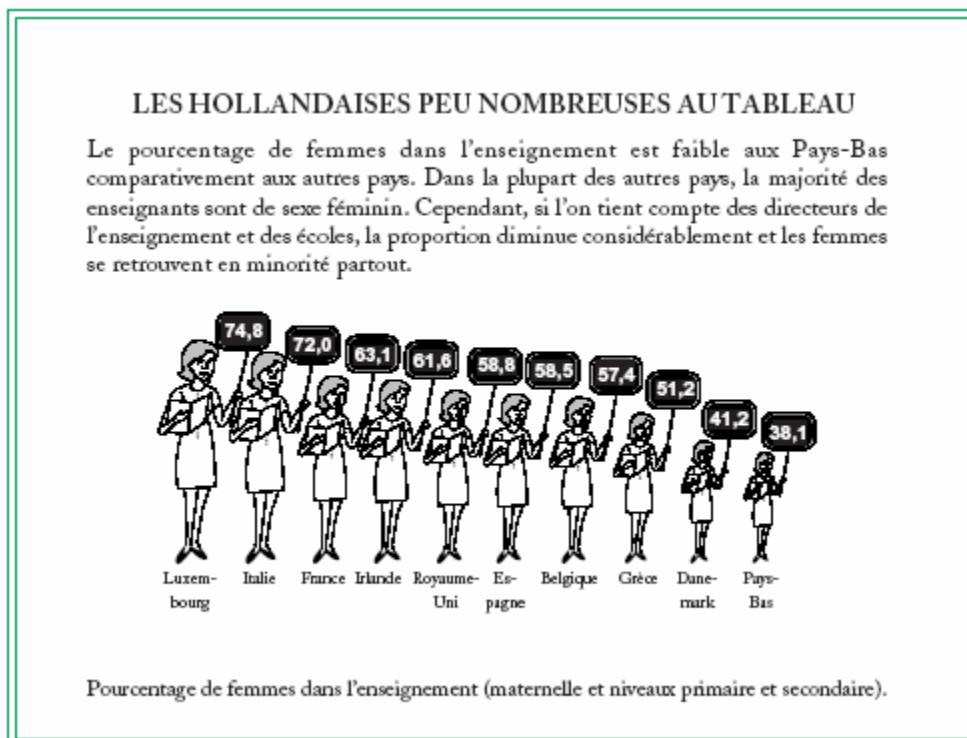
Notez les principaux points qui ont été discutés. Comparez les questions qui vous ont causé des difficultés et celles qui vous ont permis de faire valoir vos points forts. Une telle revue vous aidera à vous préparer à de futures entrevues. Si vous le désirez, vous pouvez en discuter avec l'agent de placement ou l'orienteur professionnel de votre bureau d'emploi local.

La tâche la plus difficile sur l'échelle des textes suivis (377) se situe au bas du niveau 5 et demande au lecteur de lire une annonce publiée par un service du personnel et de citer deux façons dont le CIEM (un groupe d'aide aux salariés d'une entreprise) peut aider les personnes qui perdent leur emploi par suite d'une restructuration de leur service. Le type d'appariement est jugé difficile parce que la question contenait plusieurs éléments d'information que le lecteur devait garder à l'esprit en lisant le texte. De plus, le lecteur devait donner plusieurs réponses et faire des déductions de faible niveau à partir du texte. Le type d'information est jugé moyen parce que le lecteur cherchait un but ou une fonction; la plausibilité des éléments de distraction est jugée relativement difficile. Cette tâche est un peu plus difficile du fait que l'annonce est

centrée sur des renseignements qui sont différents de ceux qui sont demandés dans la question. Ainsi, bien que les renseignements corrects se situent dans une seule phrase, l'information est cachée sous une série de rubriques décrivant les activités du CIEM pour les employés à la recherche d'un autre emploi. Cette liste de rubriques représente un excellent élément de distraction pour le lecteur qui ne cherche pas ou ne repère pas le renseignement conditionnel énoncé dans la question, soit celui qui concerne les personnes qui perdent leur emploi en raison d'une réorganisation de leur service.

4.3.6 Caractériser les tâches de compréhension de textes schématiques

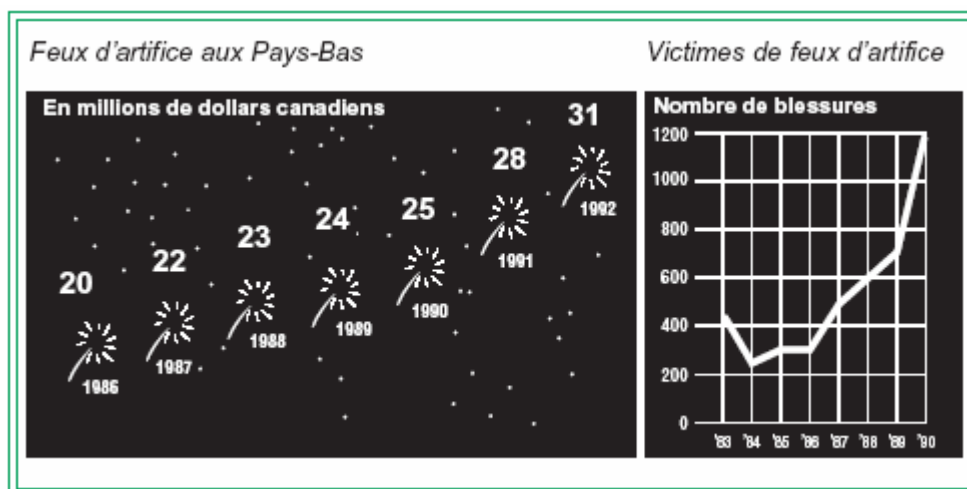
L'échelle de compréhension de textes schématiques comporte 54 tâches ordonnées le long de l'échelle de 500 points. Ces 54 tâches comprennent 19 items tirés de l'EIAA et 35 nouvelles tâches mises au point pour les besoins de l'EIACA. L'indice de difficulté de ces tâches varie de 157 à 444. L'une des tâches de niveau 1 (indice de difficulté : 188) demande au lecteur de déterminer, à partir d'un diagramme, le pourcentage d'enseignants en Grèce qui sont des femmes. Le diagramme présente le pourcentage d'enseignants de divers pays qui sont des femmes. En fonction des variables retenues, le type d'appariement est jugé simple parce que le lecteur devait repérer un seul élément d'information qui était énoncé littéralement dans le diagramme; le type d'information est jugé relativement simple parce qu'il s'agissait d'un montant; enfin, la plausibilité des éléments de distraction est aussi jugée relativement simple parce que des éléments de distraction entourent l'information demandée.



Une deuxième tâche de compréhension de textes schématiques relative à ce même diagramme demande au lecteur de trouver le pays autre que les Pays-Bas dans lequel les femmes sont en

minorité parmi les enseignants (indice de difficulté : 234; milieu du niveau 2). Cette tâche était un peu plus difficile que la première car, au lieu de chercher un pays et de repérer un pourcentage, le lecteur devait savoir qu'une minorité signifie moins de 50 %. Puis, il devait passer en revue les pourcentages pour trouver les pays dans lesquels le pourcentage de femmes parmi les enseignants était inférieur à 50 %. De plus, il devait se rappeler la condition « autre que les Pays-Bas »; autrement, il aurait pu choisir ce pays au lieu de la bonne réponse. Par conséquent, le type d'appariement est jugé moyennement difficile; le type d'information, simple (parce que l'information demandée est un pays ou un lieu); et la plausibilité des éléments de distraction, relativement simple (parce que des éléments de distraction entourent l'information demandée).

Une tâche un peu plus difficile (indice de difficulté : 295; milieu du niveau 3) demande au lecteur de regarder des diagrammes concernant les feux d'artifice aux Pays-Bas et de rédiger une brève description du lien qui existe entre les ventes et les blessures à partir de l'information fournie. Ici, le lecteur doit examiner et comparer l'information contenue dans les deux diagrammes et l'intégrer en faisant une déduction concernant la relation entre les deux ensembles d'information. La tâche est jugée relativement difficile en fonction du type d'appariement. Le type d'information est aussi jugé relativement difficile parce que l'information demandée est une tendance ou une similitude dans les données. La plausibilité des éléments de distraction est jugée moyennement difficile, principalement parce que l'information donnée et l'information demandée sont présentes dans la tâche. Par exemple, l'un des facteurs pouvant avoir contribué à la difficulté de cette tâche est le fait que le graphique des ventes va de 1986 à 1992, alors que celui des blessures va de 1983 à 1990. Le lecteur devait comparer l'information des deux diagrammes pour la période comparable.



Un autre ensemble de tâches de difficulté variable sur l'échelle de compréhension de textes schématiques concerne un document plutôt compliqué tiré d'une revue de consommateurs dans laquelle on évalue les radios-réveils. La plus simple des trois tâches (indice de difficulté : 287; niveau 3) demande au lecteur de trouver deux caractéristiques qui n'appartiennent à aucun radio-réveil de base. Le lecteur doit parcourir le document pour trouver la liste des radios-réveils de base, puis déterminer qu'un tiret représente l'absence d'une caractéristique. Il doit alors repérer les deux caractéristiques indiquées par l'ensemble de tirets. Le type d'appariement est donc jugé relativement difficile parce qu'il s'agit d'une tâche de regroupement des caractéristiques nécessitant plusieurs réponses avec une condition ou une déduction de faible niveau. Le type d'information est jugé relativement simple parce qu'il s'agit de trouver des

caractéristiques du radio-réveil, et la plausibilité des éléments de distraction est aussi relativement simple parce que certaines caractéristiques ne sont pas liées aux autres radios-réveils.

Une tâche un peu plus difficile liée à ce document se situe au bas du niveau 4 et comporte un indice de difficulté de 327. Elle demande au lecteur : « Quel radio-réveil possédant toutes les caractéristiques est le mieux coté à l'égard du rendement? » Ici, le lecteur doit apparier trois caractéristiques (« possédant toutes les caractéristiques », « le mieux coté » et « rendement »), dont l'une l'amène à traiter un renseignement conditionnel. Il est possible, par exemple, qu'un lecteur ait trouvé les radios-réveils possédant toutes les caractéristiques et la colonne « Rendement », mais qu'il ait choisi le premier radio-réveil de la liste en supposant qu'il était le mieux coté. Dans ce cas, il n'a pas compris le renseignement conditionnel, qui est une légende énonçant la signification des symboles. Un autre lecteur peut être allé à la colonne « Cote globale » et, ayant trouvé le nombre le plus élevé, avoir choisi le radio-réveil correspondant. Pour ces raisons, la plausibilité des éléments de distraction est jugée moyennement difficile. Le type d'information est jugé simple parce que l'information demandée est une chose concrète.

La tâche la plus difficile relative à ce document (indice de difficulté : 408; niveau 5) demande au lecteur de trouver le prix moyen annoncé pour le radio-réveil de base ayant obtenu la meilleure cote globale. Cette tâche est plus difficile parce que le lecteur devait apparier quatre caractéristiques au lieu de trois; il devait aussi traiter un renseignement conditionnel; enfin, il y avait un élément de distraction hautement plausible dans le même bloc que la bonne réponse. En raison de ces facteurs, le type d'appariement est jugé relativement difficile et le type d'information, relativement simple, tandis que la plausibilité des éléments de distraction présente le plus haut niveau de difficulté.

ÉVALUATIONS



Radios-réveils

Présentée par catégories et, à l'intérieur de chacune, par ordre de cote globale. Les écarts de 4 points ou moins au niveau de la cote ont été jugés significatifs.

- 1** Marque et modèle. Si vous ne pouvez trouver un modèle donné, please l'appeler la fiche. Les numéros de téléphone des fabricants figurent à la page 798.
- 2** Prix. Prix de vente conseillé par le fabricant ou prix de détail approximatif, si le prix moyen annoncé.
- 3** Dimensions. Au centimètre près.
- 4** Cote globale. Réfère les résultats de tous les tests et évaluations. Une radio « parfaite » aurait une cote de 100 points.

- 5** Commodité. L'évaluation de cet aspect complexe tient compte d'éléments tels la facilité de l'affichage, la facilité de synchronisation, avantages ainsi que présence ou absence de sensibilité aux interférences.
- 6** Rendement. Appréciation globale reflétant les résultats des tests de sensibilité et de sélectivité; de la facilité de synchronisation; du taux de coulage; capacité de recevoir la station qui a la plus grande puissance de diffusion entendue; qui a le même niveau de fréquence du rejet d'image; capacité d'écarter les signaux provenant de la bande voisine; et de la résistance aux interférences provenant de signaux réfléchis par des avions ou par d'autres appareils similaires.
- 7** Sensibilité. Qualité de la réception d'une même station, par rapport aux autres radios, sans trop d'interférences.

- 8** Sélectivité. Qualité de la réception, par toutes les radios, d'une station faible, voisine d'une forte sur la bande.
- 9** Qualité du son. Fondée principalement sur l'absence d'interférences des sons produits par le haut-parleur et sur des tests d'écoute de musique enregistrée sur disques compacts. Aucun modèle n'a donné de son haut-fidélité.
- 10** Réglage par avance ou recul. Cette caractéristique facilite le réglage de l'horloge et du réveil, car si on se réveille, il suffit de reculer.
- 11** Double commande. Permet de sélectionner deux heures de réveil distinctes.

	1	2	3	4	5	6	7	8	9	10	11	Avantages	Inconvénients	Observations	
	Marque et modèle	Prix	Dimensions (HxLxP) en cm	Cote globale	Rendement	Sensibilité	Sélectivité	Qualité du son	Réglage par avance ou recul	Double commande	Garantie, en mois				
Radios-réveils possédant toutes les caractéristiques															
RCA RP-36	50\$/40\$	8x25x18	86	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	A, B, D, H, J, L, O, T, U	A		
Sony ICF-C303	50\$/45	5x20x15	84	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	C, E, F, I, N, T	C		
Panasonic RC-X220	50\$/45	10x28x13	82	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	A, G, K, M, O, S, T, U	b, c	A	
Realistic 272	50\$/30	5x26x15	79	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	3	A, G, H, K, O, T	D		
Magnavox AJ3900	65/—	15x35x13	78	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	3	D, G, K, M, O, R, T	b, g	B	
Emerson AK2745	39\$/20	8x26x15	70	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	3	G, O	g	K	
Soundesign 3753	20\$/20	8x25x13	62	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	3	J, Q	d, h	J	
Radios-réveils de base															
Realistic 263	28\$/18	10x20x10	74	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	A, D, H, O, P, U	h	—	
Soundesign 3622	12\$/10	5x20x13	68	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	U	d	L	
Panasonic RC-6064	18\$/15	5x20x13	67	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	12	—	b, c	—	
General Electric 7-4612	13\$/10	5x20x13	66	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	12	A, D	a, g	—	
Lloyds CR001	20\$/15	5x18x13	64	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	U	—	—	
Sony ICF-C240	15\$/13	5x18x15	63	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	12	—	f, g	—	
Emerson AK2720	19\$/10	5x20x13	61	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	O, T	e	K	
Gran Prix D507	15\$/10	5x18x10	54	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	—	d	—	
Radios-réveils avec lecteur de cassettes															
General Electric 7-4965	60\$/50	10x30x15	85	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	A, D, G, H, K, O, S, T	—	B, E	
Panasonic RC-X250	111	10x33x13	78	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	A, G, K, O, R, U	b, c	A, H	
Sony ICF-CS650	75\$/65	10x28x15	74	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	✓	✓	12	G, R, T, U	a, f, i	A, F, H	
Soundesign 3844MGY	40\$/30	13x30x13	62	●●●●●	●●●●●	●●●●●	●●●●●	●●●●●	—	—	3	G, K, J, S, U	F, G, I, M		

11 Disponible et remplacé par RC-X260, prix courant de 79 \$ et prix de vente moyen annoncé de 60 \$.

Caractéristiques communes
 A - Prix de vente conseillé.
 L - Chronomètre de quartz.
 M - Entrée audio pour lecteur de cassettes ou de CD.
 N - Affichage de la date et de l'heure.
 O - Affichage à deux intensités lumineuses.
 P - Affichage à chiffres plus gros que le plupart des radios.
 Q - Volume ajustable suivant le claquement de la pièce.
 R - Réajustement des heures.
 S - Rendement des signaux.
 T - Synchronisation plus facile que la plupart des radios.
 U - Réglage d'image multiple que la plupart des radios.

Légende des inconvénients
 a - Possibilité de charger l'heure par accident.
 b - Les boutons de réglage de l'heure ou le réveil sont placés à des endroits peu commodes - en dessous ou à l'arrière.
 c - Affichage plus faible que la plupart dans une pièce à éclairage intense.
 d - Le volume doit être un peu bas pour que la sonnerie se fasse entendre.

Légende des observations
 A - Affichage des chiffres en vert.
 B - Affichage des chiffres en bleu.
 C - Affichage à cristaux liquides.
 D - Borne d'antenne extérieure.
 E - Égalisateur graphique à trois positions.
 F - Le lecteur de cassettes ne permet pas d'assourdir.
 G - Le lecteur de cassettes ne permet pas de rebriquer.
 H - Réveil possible par le lecteur de cassettes.
 I - Le lecteur de cassettes est plus bruyant que la plupart.
 J - Frais de réparation de 3 \$ pour les réparations garanties.
 K - Frais de réparation de 3,50 \$ pour les réparations garanties.
 L - Frais de réparation de 5 \$ pour les réparations garanties.
 M - Frais de réparation de 10 \$ pour les réparations garanties.

4.4 Mesurer la numératie dans le cadre de l'EIACA

La conception de la numératie adoptée pour les besoins de l'EIACA repose sur des travaux de recherche récents, menés dans plusieurs pays, sur les exigences fonctionnelles de différents contextes de vie, sur la nature des connaissances et des compétences mathématiques et statistiques des adultes, ainsi que sur l'application ou l'utilisation de ces compétences dans différentes circonstances. Vu l'intention générale de l'EIACA de produire des renseignements sur un ensemble diversifié de compétences utiles à la vie, le présent cadre de travail définit la numératie comme suit :

Connaissances et compétences nécessaires pour gérer efficacement les exigences mathématiques de diverses situations.

Cette définition suppose que le concept de la numératie est plus large que celui de la compréhension de textes au contenu quantitatif, défini par l'EIAA. En outre, la numératie des adultes doit être considérée comme différente d'une « connaissance scolaire des mathématiques ». S'il n'existe pas de définition universellement reconnue de la « numératie » (Baker et Street, 1994), un examen de certaines perspectives concernant le sens de la numératie des adultes montre qu'elles renferment plusieurs points communs. Voici deux exemples, tous deux tirés de travaux menés en Australie :

La numératie consiste à utiliser les mathématiques pour fonctionner efficacement au sein d'un groupe et d'une collectivité, ainsi qu'à utiliser ces compétences pour favoriser l'évolution d'une personne et celle de sa collectivité (Beazley, 1984). [Traduction libre]

La numératie consiste à interpréter, appliquer et communiquer une information mathématique dans des situations courantes afin de jouer pleinement, de manière essentielle et efficace, un vaste éventail de rôles dans la vie (Département de l'Éducation du Queensland, 1994). [Traduction libre]

Sur le plan de la portée générale, toutes ces définitions sont très semblables à celles de la compréhension de textes suivis et de la compréhension de textes schématiques formulées pour les besoins de l'EIACA et présentées dans une section antérieure. Bon nombre de conceptions de la numératie mettent l'accent sur l'application et l'utilisation pratiques ou fonctionnelles des connaissances et des compétences mathématiques pour composer avec la présence d'éléments mathématiques dans des situations réelles. Les adultes sont censés maîtriser plusieurs façons de réagir avec souplesse à une situation mathématique en vue d'atteindre un objectif, selon les besoins et les intérêts d'une personne dans un contexte donné (maison, collectivité, travail, etc.) ainsi que son attitude et ses croyances à l'égard de la numératie (Gal, 2000; Coben, O'Donoghue et FitzSimons, 2000).

La numératie ne se limite donc pas à l'application des compétences arithmétiques à l'information contenue dans des imprimés, qui était évaluée dans le cadre de l'EIAA. La numératie des adultes englobe également le sens des nombres, ainsi que des compétences en estimation, en mesure et en statistique. Étant donné le rôle envahissant de la numératie dans le monde moderne, ce ne sont pas nécessairement que les situations courantes qui font appel aux pratiques de calcul, mais aussi de **nouvelles** situations.

Un autre élément important de la définition de la numératie tient au rôle des processus de communication. La numératie englobe non seulement la capacité d'une personne d'utiliser et d'appliquer ses compétences mathématiques de manière efficace et critique, mais aussi la

capacité d'interpréter des messages textuels ou symboliques, ainsi que de communiquer au sujet de l'information mathématique et des processus de raisonnement (Marr et Tout, 1997; Gal, 1997).

Les définitions de la numératie énoncent explicitement que la numératie ne consiste pas seulement à effectuer des opérations avec des nombres, comme le terme le laisse entendre, surtout pour ceux qui connaissent bien les conceptions de la numératie des enfants, mais qu'elle englobe une foule de compétences et de connaissances mathématiques. En outre, au cours des dernières années, il a beaucoup été question de la relation entre les mathématiques et la numératie et du caractère « essentiel » de la numératie (Frankenstein, 1989; Steen, 2001). Johnston, par exemple, soutient que :

Savoir calculer, ce n'est pas qu'être capable de manipuler des nombres, ni même « posséder » les mathématiques apprises à l'école ou à l'université. La numératie suppose une connaissance essentielle qui constitue un lien entre les mathématiques et le monde réel, dans toute sa diversité (Johnston, 1994).
[Traduction libre]

Bon nombre d'auteurs soutiennent qu'en abordant les compétences fonctionnelles, on doit aussi tenir compte d'attitudes et de croyances « habilitantes ». Dans le domaine des compétences mathématiques des adultes, on s'attend à l'« aisance » avec les nombres ou à la « confiance » dans les compétences mathématiques, car ces caractéristiques ont une incidence sur la mise en pratique effective des compétences et des connaissances (Cockroft, 1982; Tobias, 1993).

À la brève définition de la numératie formulée pour les besoins de l'EIACA et présentée plus haut s'ajoute une définition plus large des **pratiques de calcul**, formulée par l'équipe d'évaluation de la numératie de l'EIACA pour servir de base à la mise au point des items de numératie de l'EIACA :

On met en œuvre les pratiques de calcul lorsqu'on gère une situation ou qu'on résout un problème dans un contexte réel; il s'agit de réagir à l'information sur des notions mathématiques pouvant être représentées de diverses façons et de mettre en œuvre une gamme de connaissances, de facteurs et de processus habilitants.

Cette conception des pratiques de calcul suppose que pour évaluer la numératie des adultes, il est nécessaire de produire des tâches et des items qui varient selon les contextes, les réactions demandées, la nature de l'information mathématique en question et les représentations de cette information. Ces caractéristiques des tâches sont définies ci-dessous. Cette conception est beaucoup plus large que la définition de la compréhension de textes au contenu quantitatif formulée pour les besoins de l'EIAA. Ses concepts clés concernent, de façon générale, la gestion d'une situation et un éventail de réactions efficaces (et pas seulement l'application des compétences arithmétiques). Elle englobe une vaste gamme de compétences et de connaissances (et pas seulement les opérations de calcul) et une foule de situations qui présentent différents types d'information mathématique (et pas seulement ceux qui comprennent des **nombres** inclus dans des textes **imprimés**).

La mise au point des items visait à assurer qu'une certaine proportion de la banque d'items imposerait aux répondants un fardeau de lecture minimal, c'est-à-dire que certains stimuli comporteraient peu ou pas de texte, ce qui permettrait même aux répondants possédant une

maîtrise limitée de la langue du test de comprendre la situation décrite. D'autres sections de la banque d'items comprenaient des items comportant des textes essentiels plus ou moins longs, selon les exigences de la situation représentée par l'item.

Comme le laissent entendre la documentation et les notions évoquées plus haut, la nature des réactions d'une personne aux exigences mathématiques et autres d'une situation dépend essentiellement de la mise en application de diverses bases de connaissances habilitantes (compréhension du contexte, connaissances et compétences dans les domaines des mathématiques, de la statistique et de la littératie), de processus de raisonnement, ainsi que de son attitude et de ses croyances à l'égard de la numératie. Les pratiques de calcul dépendent en outre de l'intégration des connaissances et des compétences mathématiques avec des compétences et des stratégies plus vastes en littératie et en résolution de problèmes ainsi qu'avec l'expérience et les comportements acquis qu'un adulte apporte à chaque situation. Il est clair que les pratiques de calcul consistent à tenter d'accomplir une tâche de manière autonome, sans la déléguer ni la gérer en faisant volontairement abstraction de son contenu mathématique.

4.4.1 Définir les caractéristiques des tâches

Pour mettre au point et représenter les tâches de numératie conçues pour les besoins de l'EIACA, on a retenu quatre caractéristiques clés des pratiques de calcul : le type d'objet ou de contexte, le type de réaction, le type d'information mathématique ou statistique et le type de représentation de cette information. Chacune d'elles est décrite ci-dessous.

Type d'objet ou de contexte. Une personne tente de gérer une situation de numératie ou d'y réagir parce qu'elle veut satisfaire un besoin ou atteindre un objectif. Quatre types de buts et d'objectifs sont décrits ci-dessous. Force est de constater qu'ils ne sont pas mutuellement exclusifs et qu'ils peuvent correspondre aux mêmes thèmes mathématiques sous-jacents.

4.4.2 Vie quotidienne

Les tâches de numératie qui se présentent dans des situations courantes sont souvent des tâches de gestion auxquelles une personne fait face dans sa vie personnelle et familiale. D'autres sont liées à des passe-temps, ainsi qu'au perfectionnement et aux intérêts personnels. Il peut s'agir des tâches suivantes : régler des questions d'argent et de budget, comparer des prix, gérer son temps personnel, prendre des décisions en matière de déplacements, planifier des vacances, faire des calculs dans le cadre de passe-temps comme la courtepoinette ou le travail du bois, jouer à des jeux de hasard, comprendre des statistiques et des résultats sportifs, lire des cartes géographiques et utiliser des mesures dans le cadre d'activités ménagères comme la cuisine ou le bricolage.

4.4.3 Travail

Au travail, on fait face à des situations à composante quantitative qui sont souvent plus spécialisées que celles qui se présentent dans la vie quotidienne. Dans ce contexte, une personne peut acquérir de bonnes compétences pour gérer des situations plus « pointues » dans l'application de thèmes mathématiques. Il peut s'agir des tâches suivantes : remplir des bons de commande, totaliser des recettes, calculer la monnaie, gérer des horaires, utiliser des feuilles de calcul, disposer et emballer des articles de formes différentes, remplir et interpréter

des cartes de contrôle, effectuer et consigner des mesures, lire des plans, contrôler des dépenses, prévoir des coûts et appliquer des formules.

4.4.4 Vie sociale ou collective

Les adultes ont besoin de connaître les processus qui existent dans le monde qui les entoure, par exemple les tendances de la criminalité, des salaires et de l'emploi, de la pollution et des risques pour la santé ou l'environnement. Ils sont parfois appelés à prendre part à des activités sociales, communautaires ou politiques. Ils doivent donc savoir lire et interpréter l'information quantitative présentée dans les médias, y compris les messages et graphiques statistiques. Ils peuvent être appelés à gérer les situations suivantes : organiser une campagne de financement, calculer l'incidence budgétaire d'un programme communautaire ou interpréter les résultats d'une étude sur la plus récente tendance en matière de santé.

4.4.5 Formation complémentaire

Les compétences en numératie permettent à une personne de participer à des études complémentaires, que ce soit dans un but scolaire ou dans le cadre d'une formation professionnelle au niveau des métiers. Dans un cas comme dans l'autre, il importe de connaître certains aspects formels des mathématiques – symboles, règles et formules – et de comprendre certaines conventions régissant l'application des règles et des principes mathématiques.

Type de réaction. Dans différents types de situation réelle, une personne peut avoir à réagir de l'une ou de plusieurs des façons suivantes (la première va pratiquement de soi; les autres dépendent de l'interaction entre les exigences de la situation et les objectifs, les compétences, les dispositions et les connaissances acquises de la personne) :

Identifier ou repérer, dans la tâche ou la situation qui se présente, une information mathématique qui est pertinente à l'égard d'un besoin ou d'un objectif.

Réagir à l'information présente dans la situation. Bishop (1988), par exemple, distingue six modes d'actions mathématiques qui sont communs à toutes les cultures : compter, repérer, mesurer, concevoir, jouer et expliquer. D'autres types d'actions ou de réactions peuvent entrer en jeu, comme faire des calculs (mentalement ou à l'aide d'une calculatrice), ordonner ou trier, estimer, mesurer ou modéliser (par exemple, en utilisant une formule).

Interpréter l'information comprise dans la situation (et les résultats d'une intervention antérieure) et comprendre ce qu'elle signifie ou suppose. Il peut s'agir de porter un jugement sur la façon dont une information mathématique ou des faits connus s'appliquent à la situation ou au contexte. Il peut être nécessaire de faire preuve de jugement pour déterminer qu'une réponse a du sens ou non dans le contexte donné (savoir, par exemple, qu'un résultat de « 2,35 voitures » n'est pas une solution valide lorsqu'on se demande combien de voitures sont nécessaires pour transporter un groupe). La situation peut aussi comporter un aspect critique si la personne remet en question l'objet de la tâche, la validité des données ou de l'information présentée, ainsi que le sens et la portée des résultats, pour elle-même sur le plan individuel et, peut-être, pour la collectivité dans son ensemble.

Communiquer à quelqu'un d'autre au sujet de l'information mathématique donnée ou des résultats d'une intervention ou d'une interprétation. On peut le faire oralement ou par écrit (il

peut s'agir simplement d'un nombre ou d'un mot ou encore d'une explication ou d'une analyse détaillée) ou au moyen d'un dessin (diagramme, carte géographique, graphique).

Type d'information mathématique ou statistique. On peut classer l'information mathématique d'un certain nombre de façons et selon différents niveaux d'abstraction. Une démarche consiste à se reporter aux « grands principes » fondamentaux du domaine des mathématiques. Steen (1990), par exemple, distingue les six grandes catégories suivantes : quantité, dimension, structure, forme, incertitude et variation. Rutherford et Ahlgren (1990) décrivent des réseaux de notions connexes : nombres, formes, incertitude, données récapitulatives, échantillonnage et raisonnement. Dossey (1997) catégorise comme suit les comportements mathématiques en matière de compréhension de textes au contenu quantitatif : représentation et interprétation de données, sens des nombres et des opérations, mesure, variables et relations, formes géométriques et visualisation spatiale, et hasard. L'équipe d'évaluation de la numératie de l'EIACA s'est inspirée de ces catégorisations et d'autres catégorisations étroitement liées (par ex., National Council of Teachers of Mathematics, 2000) pour en arriver à un ensemble de cinq notions fondamentales qui, selon elle, caractérisent les exigences mathématiques que les adultes doivent satisfaire dans diverses situations en ce début du XXI^e siècle.

4.4.6 Quantité et nombre

Selon Fey (1990), la notion de **quantité** procède du besoin de quantifier le monde qui nous entoure en utilisant des attributs comme les suivants : longueur, surface et volume des cours d'eau ou des masses continentales; température, humidité et pression atmosphérique; populations et taux de croissance des espèces; amplitudes des marées; revenus ou bénéfices d'entreprises.

La notion de **nombre** est fondamentale pour la quantification, et différents types de nombres déterminent la quantification de diverses façons : les nombres entiers servent de compteurs ou d'estimateurs; les fractions, décimales et pourcentages expriment des quantités plus précises ou représentent les parties d'un tout, ce qui permet de comparer des proportions. Les nombres positifs et négatifs servent d'indicateurs directionnels. En plus de la quantification, on utilise les nombres pour ordonner et pour identifier (par exemple, dans le cas de numéros de téléphone ou de codes postaux). L'aisance avec la quantité, les nombres et les opérations sur les nombres exige un bon « sens » de la grandeur et de la signification de nombres très grands ou très petits, et parfois le sens de la grandeur relative de différentes proportions.

La gestion de l'argent et du temps, aspect omniprésent des mathématiques dans la vie des adultes, dépend d'un sens aigu des nombres et des quantités. On doit faire preuve de jugement contextuel pour déterminer dans quelle mesure il faut être précis ou quel outil utiliser (calculatrice, calcul mental, ordinateur). Une tâche de numératie de base peut consister à calculer le coût d'une boîte de soupe en sachant que quatre boîtes coûtent 2,00 \$; une tâche dont l'exigence cognitive est élevée peut nécessiter des nombres plus « pointus », par exemple si l'on calcule le coût d'un kilo de fromage en sachant que 0,783 kg coûte 12,95 euros.

4.4.7 Dimension et forme

La **dimension** repose sur les « grands principes » liés à une, deux et trois dimensions des « choses ». On doit faire appel à la connaissance des dimensions lorsqu'il s'agit de lire ou de

produire des descriptions spatiales ou numériques d'objets, de faire des projections ou de travailler avec des longueurs, des périmètres, des plans, des surfaces, un emplacement, etc. L'aisance avec chaque dimension exige un sens des mesures « repères », des mesures directes et des mesures estimatives.

La **forme** est une catégorie qui décrit des images et des entités réelles ou imaginaires qu'on peut visualiser (maisons, immeubles, motifs d'art ou d'artisanat, enseignes de sécurité, emballages, nœuds, cristaux, silhouettes, plantes, etc.). L'orientation et l'emplacement sont des qualités fondamentales auxquelles on a recours pour lire ou tracer des cartes géographiques et des diagrammes. Une tâche de numérisation de base liée à cet aspect fondamental pourrait consister à identifier des formes, et une tâche complexe, à décrire la variation de la taille ou du volume d'un objet dont une dimension est modifiée, par exemple lorsqu'on choisit entre différentes boîtes pour emballer certains objets.

4.4.8 Structure, fonctions et relations

On écrit souvent que les mathématiques sont l'étude des structures et des relations. La structure apparaît comme un vaste concept qui englobe les structures qu'on trouve tout autour de nous, par exemple sous forme musicale, dans la nature et dans les mouvements de la circulation. Senechal (1990) soutient que notre capacité de reconnaître, d'interpréter et de créer des structures est essentielle pour appréhender le monde qui nous entoure. La capacité humaine d'identifier des relations et d'analyser sous-tend le raisonnement mathématique. L'algèbre – au-delà de la manipulation symbolique – offre un outil servant à représenter des relations entre des montants au moyen de tableaux, de graphiques, de symboles et de mots. La capacité de généraliser et de caractériser des fonctions et des relations entre des variables est essentielle à la compréhension des analyses économiques, politiques ou sociales, même les plus simples. Une tâche de reconnaissance de structure relativement simple pourrait consister à demander à quelqu'un de décrire la structure d'une suite de nombres ou de formes donnés ou, dans un contexte fonctionnel, à comprendre la relation entre des listes de variables (par ex., le poids et le volume de certains objets); élaborer une formule pour un tableur imposerait à la personne un niveau d'exigence plus élevé.

4.4.9 Données et hasard

Les données et le hasard recouvrent deux sujets connexes mais distincts. Les **données** reposent sur des « grands principes » tels que la variabilité, l'échantillonnage, l'erreur ou la prévision et sur des sujets statistiques connexes tels que la collecte et l'analyse de données, les mesures courantes d'une valeur centrale ou d'une fourchette de valeurs, ou la notion d'une inférence statistique. La société moderne exige des adultes qu'ils interprètent (et parfois même qu'ils produisent) des tableaux de fréquence, des diagrammes et des graphiques de base, des renseignements sur des moyennes et des médianes, et qu'ils reconnaissent des affirmations statistiques douteuses (Gal, 2002).

Le **hasard** repose sur des « grands principes » liés à la probabilité et aux concepts et outils statistiques pertinents. Peu de choses au monde sont absolument certaines; la capacité d'associer un nombre à la probabilité d'un événement (et de ses risques ou effets secondaires) constitue donc un outil précieux, qu'il s'agisse du temps qu'il fait, de la bourse des valeurs ou de la décision de prendre un certain médicament. Dans cette catégorie, une tâche de numérisation simple pourrait consister à interpréter un diagramme circulaire simple ou à comprendre un

énoncé relatif à une moyenne; une tâche plus complexe consisterait à déduire la probabilité d'un événement d'après l'information fournie.

4.4.10 Variation

Il s'agit des mathématiques de l'évolution du monde autour de nous. Les organismes croissent, les populations varient, les prix fluctuent, les objets en mouvement accélèrent et ralentissent. La variation et les taux de variation nous aident à décrire l'état du monde au fil du temps. Les structures additives, multiplicatives et exponentielles de la variation peuvent caractériser des tendances stables; les variations périodiques dénotent des cycles, et les courbes de variation irrégulières évoquent la théorie du chaos. Décrire une perte de poids constitue une tâche relativement simple, alors que calculer un intérêt composé est une tâche relativement complexe.

Type de représentation de l'information mathématique. L'information mathématique inhérente à une activité ou à une situation peut exister ou être représentée sous plusieurs formes. Il peut s'agir d'objets concrets à compter (moutons, personnes, immeubles, voitures, etc.) ou d'illustrations de ces objets. Elle peut être représentée au moyen d'une notation symbolique (chiffres, lettres et signes d'opération). Parfois, l'information mathématique figure dans des formules, qui constituent un modèle de relations entre des entités ou des variables.

En outre, l'information mathématique peut être codée de manière visuelle, comme dans un **diagramme**; on peut utiliser des **graphiques** et des **tableaux** pour présenter une information statistique ou quantitative agrégée. De même, une **carte géographique** d'une entité réelle (par exemple, le plan d'une ville ou d'un chantier) peut contenir des données numériques, mais aussi une information qu'on peut quantifier ou mathématiser.

Enfin, une personne peut être appelée à extraire une information mathématique de divers types de textes suivis ou schématisés de formes précises (par exemple, des déclarations de revenus). Les tâches de numératie fonctionnelle portent sur deux sortes différentes de texte. La première comporte une information mathématique représentée sous forme textuelle, c'est-à-dire accompagnée de mots ou d'expressions ayant un sens mathématique. On peut citer comme exemples l'utilisation de chiffres en toutes lettres (par exemple, « cinq » au lieu de « 5 »), de termes mathématiques de base (par exemple, fraction, multiplication, pourcentage, moyenne, proportion) ou d'expressions plus complexes (par exemple, « le taux de criminalité a chuté de moitié ») qu'il faut interpréter. Dans le deuxième type de texte, l'information mathématique est exprimée en notations ou symboles réguliers (nombres, signes plus ou moins, symboles d'unités de mesure, etc.), mais elle est entourée d'un texte qui, malgré sa nature non mathématique, doit aussi être interprété afin de fournir une information et un contexte supplémentaires. On peut citer l'exemple d'un bordereau de dépôt bancaire comportant un texte et des directives dans lesquels sont inclus des nombres correspondant à des montants monétaires.

4.4.11 Caractériser les tâches de numératie

Pour les besoins de l'EIACA, on a retenu au total 40 tâches de numératie. Ces tâches s'échelonnent de 174 à 380 le long de l'échelle de la numératie, leur emplacement étant déterminé par la façon dont les adultes des pays participants ont réagi à chaque tâche. On trouvera ci-dessous quelques exemples de tâches qui reflètent certains aspects conceptuels de

la numératie et les principes de conception de l'échelle décrits plus haut, soit le calcul, le raisonnement spatial et proportionnel, la mesure et les connaissances statistiques.

La tâche la plus simple sur l'échelle de la numératie (174) présentait aux adultes une photographie montrant deux cartons de bouteilles de Coca-Cola. On leur demandait de trouver le nombre total de bouteilles que contenaient les deux caisses pleines. Si cette tâche était simple, c'est d'abord parce qu'elle était empruntée à la vie quotidienne et que les objets en question étaient relativement familiers pour la plupart des gens. Ensuite, ce qu'on demandait aux répondants était évident et explicite : cette tâche faisait appel à une photographie représentant des objets concrets et ne comportait aucun texte à traiter. Un troisième facteur tenait au fait que les répondants pouvaient accomplir la tâche de diverses façons plus ou moins complexes, par exemple en multipliant des rangées par des colonnes, mais aussi par simple comptage. Cette tâche amène le répondant à faire une conjecture, car la totalité des bouteilles de la caisse du dessous n'est pas visible, mais, comme l'indique le faible niveau de difficulté de la tâche, cette caractéristique ne présentait pas un problème pour la vaste majorité des adultes de tous les pays participants.

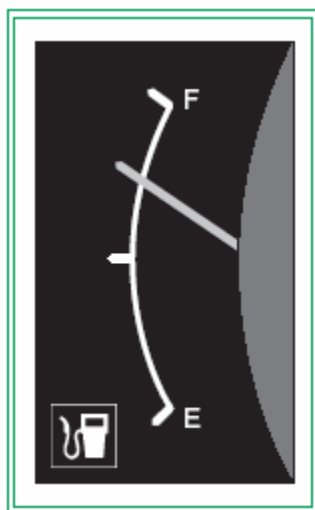


Une deuxième tâche, également très simple, exigeait des adultes qu'ils lisent un court texte rapportant les résultats d'une élection qui opposait trois candidats et de déterminer le nombre total de votes exprimés. Cette tâche (indice de difficulté : 192) se situe au niveau 1 de l'échelle de la numératie. Encore une fois, on mettait les répondants dans un type de situation réaliste en leur présentant une information numérique simple, sous forme de colonnes montrant le nom de chaque candidat et le nombre de votes obtenus par chacun. La tâche ne comportait aucune autre information numérique pouvant constituer un élément de distraction. Pour trouver le nombre total de votes exprimés lors de l'élection, il suffit d'effectuer une seule addition, ce que la question rend explicite par l'emploi du mot-clé « total », et le calcul porte sur des nombres entiers relativement peu élevés.

Une tâche de numératie plus complexe, située au milieu du niveau 2 (indice de difficulté : 248) amène les adultes à regarder un indicateur de niveau d'essence. Cet indicateur comporte trois

lignes ou coches dont l'une est marquée « F » pour « full » (plein), une autre est marquée « E » pour « empty » (vide) et la troisième se situe au milieu, entre les deux. Une ligne représentant l'aiguille de l'indicateur montre un niveau qui se trouve à peu près à mi-chemin entre la coche du milieu et la coche marquée « F », montrant ainsi que le réservoir est environ aux trois quarts plein. La directive énonce que le réservoir peut contenir 48 gallons et demande au répondant de déterminer combien de gallons il reste dans le réservoir. Cette tâche est empruntée à un contexte quotidien et amène le répondant à interpréter une présentation comportant une information quantitative, mais pratiquement pas de texte ni de nombres. Aucune information mathématique n'est présente, à part celle qui est fournie dans la question.

Ce qui rend cette tâche plus difficile que les deux précédentes est le fait que le répondant doit d'abord estimer le niveau d'essence restant dans le réservoir en convertissant l'emplacement de l'aiguille en une fraction. Puis, il doit déterminer combien de gallons cette fraction représente par rapport à la capacité de 48 gallons énoncée dans la question ou la directive. Cette tâche l'amène donc à effectuer plus d'une opération pour arriver à une bonne réponse, sans préciser le type d'opération. Néanmoins, comme bien d'autres tâches de numératie courantes, cette tâche n'exige pas qu'on effectue un calcul exact, mais une approximation qui devrait se situer à l'intérieur de limites raisonnables.

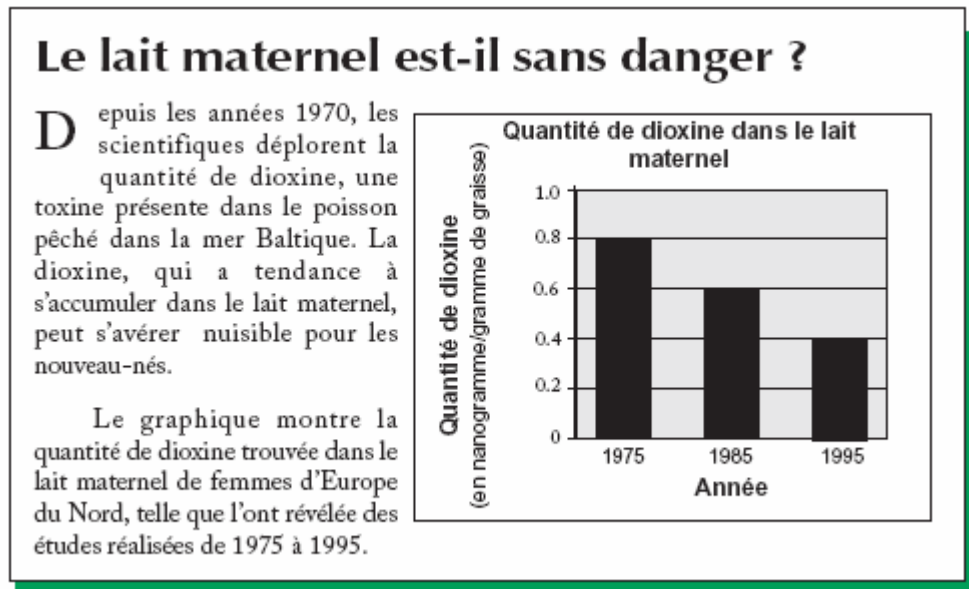


Une tâche de numératie un peu plus difficile, située au sommet du niveau 2 (indice de difficulté : 275) présente aux adultes le diagramme d'un conteneur sur lequel se trouvent quatre marques ou lignes; on demande aux répondants de tracer sur le conteneur une ligne à la hauteur du tiers. La ligne du haut est marquée « 1 », alors que celle du milieu est marquée « $\frac{1}{2}$ ». Les deux autres lignes sont sans marque; l'une est située à mi-chemin entre « 1 » et « $\frac{1}{2}$ » et l'autre, à mi-chemin entre la ligne marquée « $\frac{1}{2}$ » et le fond du conteneur. Pour répondre correctement, le répondant doit tracer sur le conteneur une ligne entre la ligne marquée « $\frac{1}{2}$ » et la ligne au-dessous à la hauteur du quart (mais cette ligne n'est pas marquée « $\frac{1}{4}$ » – il faut le déduire). Ici, le contexte peut être moins familier pour le répondant mais, encore une fois, l'image visuelle est simple et réaliste et ne comporte pratiquement pas de texte. Il ne s'agit pas d'écrire un symbole ou un texte, mais simplement de tracer une ligne à un certain endroit sur le dessin du conteneur. Pour accomplir correctement cette tâche, le répondant doit posséder une certaine connaissance pratique des fractions et le sens des proportions : il doit connaître les symboles

« ½ » et « ⅓ », savoir ordonner les fractions en fonction de leur taille relative et les faire correspondre aux marques existantes sur le conteneur.

On a mis au point quelques tâches de numératie axées sur un bref article de journal intitulé « Le lait maternel est-il sans danger? », au sujet des risques environnementaux et de la salubrité des aliments. L'article comportait deux brefs paragraphes concernant une toxine, la dioxine, qu'on trouve dans le poisson de la mer Baltique, ainsi qu'un graphique à barres indiquant les niveaux de dioxine observés à trois moments – 1975, 1985 et 1995 – dans le lait maternel des femmes de l'Europe du Nord. Une question demandait au répondant de décrire comment la quantité de dioxine avait varié de 1975 à 1995, c'est-à-dire d'interpréter directement les données présentées dans le graphique. On ne lui demandait pas de calculer la variation exacte observée au cours de chacune des périodes, mais simplement de décrire en ses propres mots la variation des niveaux de dioxine (avaient-ils diminué, augmenté ou étaient-ils restés inchangés?).

On a attribué à cette tâche un indice de difficulté de 280, au bas du niveau 3. Le graphique montre clairement que la quantité de dioxine a diminué au cours de chacune des trois périodes; pourtant, certains adultes ont du mal à accomplir cette tâche, qui est fondée sur un stimulus dont la structure est courante dans les journaux, soit un texte bref accompagné d'un graphique. Le niveau de difficulté accru de cet item est sans doute attribuable en partie au fait que le répondant doit formuler sa propre description, au degré moyen de dépendance au texte pour comprendre le contexte auquel se rapporte le graphique, ou à la nécessité de comprendre l'orientation des valeurs décimales sur l'axe vertical (mode courant de représentation des concentrations de produits chimiques polluants).



Dans une deuxième tâche axée sur le même stimulus, mais plus difficile, on demandait aux adultes de comparer le pourcentage de variation du niveau de dioxine de 1975 à 1985 à celui de 1985 à 1995, de déterminer lequel était le plus élevé et d'expliquer leur réponse. Cette tâche s'étant avérée beaucoup plus difficile pour les adultes des pays participants, on lui a attribué un indice de difficulté de 377 sur l'échelle de la numératie. Ici, l'information nécessaire est intégrée

au graphique et exige une certaine part de transformation et d'interprétation. Pour arriver à une bonne réponse, le répondant doit prendre en compte le taux de variation exprimé en pourcentage, et pas seulement la valeur absolue de la variation. En outre, il doit composer avec des pourcentages d'entités inférieures à un (c'est-à-dire les valeurs décimales sur l'axe vertical) et savoir que la base de calcul de la variation en pourcentage varie d'une paire à l'autre. Il semble que la nécessité de prendre en compte ces caractéristiques de la tâche, d'employer des méthodes mathématiques formelles ou de composer avec la notion abstraite de taux de variation ajoute considérablement à la difficulté de ce genre de tâche.

La tâche de numératie la plus difficile de cette évaluation (indice de difficulté : 380; niveau 5) présentait aux adultes une publicité selon laquelle il est possible, pour un investisseur, de doubler en sept ans un montant investi au taux d'intérêt fixe de 10 % par année. On demandait aux répondants s'il était possible de doubler en sept ans un montant de 1 000 \$ investi à ce taux et d'étayer leur réponse par leurs calculs. On a accepté différentes réponses à condition qu'elles soient accompagnées d'une justification raisonnable et des calculs pertinents. Les répondants étaient libres d'effectuer le calcul à leur manière, mais pouvaient aussi utiliser un « tuyau financier » qui accompagnait la publicité et présentait une formule pour estimer la valeur d'un investissement après un certain nombre d'années. Ceux qui employaient la formule devaient assimiler l'information énoncée dans le texte aux variables de la formule (principal, taux d'intérêt et période), puis effectuer les calculs nécessaires et comparer le résultat au montant qu'on obtiendrait si le montant de 1 000 \$ était doublé.

Tous les répondants pouvaient utiliser une calculatrice de poche fournie au moment de l'évaluation. Cette tâche s'est avérée difficile parce qu'elle portait sur des pourcentages et que le calcul, avec ou sans la formule, exigeait qu'on intègre plusieurs étapes et plusieurs types d'opération. Pour effectuer les calculs sans la formule, il était nécessaire de comprendre la méthode de calcul des intérêts composés. Cette tâche permettait aux répondants de recourir à différentes stratégies de raisonnement, y compris des méthodes non structurées ou de leur cru. Pourtant, à l'instar de la tâche qui consistait à comparer des taux de variation, elle exigeait le recours à une information mathématique formelle et à une compréhension approfondie de méthodes de calcul peu courantes, ce qui n'était peut-être pas familier ou accessible à bien des adultes.

4.5 Mesurer la résolution de problèmes dans le cadre de l'EIACA

La résolution de problèmes fait depuis longtemps l'objet de recherches psychologiques universitaires et de recherches appliquées en ressources humaines. Voici une définition très générale de la résolution de problèmes telle qu'on l'entend habituellement dans le domaine de la recherche psychologique (Hunt, 1994; Mayer, 1992; Mayer et Wittrock, 1996; Smith, 1991) :

La résolution de problèmes correspond à la pensée et à l'action orientées vers les buts dans une situation où il n'existe aucune procédure courante de résolution. La personne qui résout des problèmes a un but plus ou moins bien défini, mais ne sait pas immédiatement comment l'atteindre. La non-congruence des buts et des opérateurs admissibles constitue un problème. La compréhension de la situation problématique et sa transformation progressive fondée sur la planification et le raisonnement constituent le processus de la résolution de problèmes.

Lorsqu'on élabore un cadre de travail de la résolution de problèmes en vue d'une enquête comme l'EIACA, l'une des principales difficultés consiste à adapter le mieux possible la documentation psychologique aux contraintes imposées par une étude comparative internationale à grande échelle. À cette fin, on a décidé de mettre l'accent sur un sous-ensemble essentiel de résolution de problèmes : la résolution de problèmes analytiques. Il faut éviter de confondre notre notion de la résolution de problèmes analytiques avec le sens courant et intuitif de l'expression et avec la notion clinico-psychologique selon laquelle la résolution de problèmes est associée à la résolution de conflits sociaux et émotifs. Pourtant, le contexte social est tout à fait pertinent pour notre définition de la résolution de problèmes analytiques, par exemple lorsqu'il faut aborder un problème interactivement et le résoudre par la collaboration. Des facteurs motivants comme l'intérêt pour le sujet et l'orientation des tâches influencent également le processus de résolution des problèmes. En revanche, la qualité de la résolution est déterminée avant tout par la compréhension de la situation problématique, les processus de pensée utilisés pour aborder le problème et l'à-propos de la solution retenue.

On peut faire appel à différents aspects pour caractériser le **problème** lui-même :

- ❖ Le **contexte** peut refléter divers domaines, de nature théorique ou pratique, rattachés à une situation intellectuelle ou au monde réel. Dans ces domaines, un problème peut s'avérer plus ou moins authentique.
- ❖ **L'ampleur** d'un problème est variable; une personne peut s'en tenir à des aspects limités et concrets d'une tâche, planifier et exécuter des actions complexes ou évaluer des séquences d'actions multiples.
- ❖ Le problème peut comporter un but bien ou mal défini, des contraintes transparentes (explicites) ou non, un nombre restreint d'éléments indépendants ou un grand nombre d'éléments interdépendants. Ces caractéristiques définissent la **complexité** du problème.

La connaissance du contexte que possède la population cible, le fait que le problème porte sur des tâches concrètes ou des actions complexes, la mesure dans laquelle le but est bien défini, la transparence des contraintes, le nombre d'éléments que la personne qui résout des problèmes doit prendre en compte et l'ampleur de leur interdépendance – voilà autant de caractéristiques qui déterminent le niveau de compétence nécessaire pour résoudre un problème donné. La difficulté empirique, c'est-à-dire la probabilité de trouver une solution correcte, dépend de la relation entre ces caractéristiques du problème, d'une part, et le niveau de compétence des sujets, de l'autre.

Les **processus cognitifs** actionnés au cours de la résolution d'un problème sont divers et complexes, et sont plutôt organisés de manière non linéaire. Ces processus comportent notamment les cinq éléments suivants :

1. Chercher l'information, puis la structurer et l'intégrer en une représentation intellectuelle du problème (« modèle situationnel »).
2. Raisonner à la lumière du modèle situationnel.
3. Planifier des actions et d'autres étapes de la solution.
4. Exécuter et évaluer les étapes de la solution.
5. Traiter l'information externe et la rétroaction sur une base soutenue.

Baxter et Glaser (1997) présentent une liste semblable d'activités cognitives, désignées les « composantes générales de la compétence en résolution de problèmes » : représentation du problème, stratégies de résolution, autorégulation et explications. La résolution de problèmes analytiques dans un contexte courant, mesurée par l'instrument de résolution de problèmes de l'EIACA, met l'accent sur les trois premières composantes (et, dans une certaine mesure, sur la quatrième).

L'un des constats les plus importants des récents travaux de recherche en psychologie cognitive est que la résolution de problèmes difficiles exige au moins une connaissance élémentaire du domaine en question. La notion d'un espace de problème dans lequel la personne qui résout des problèmes évolue au moyen de stratégies de recherche indépendantes du domaine (Newell et Simon, 1972) s'est avérée trop simple pour décrire la manière d'appréhender une situation problématique et la recherche d'une solution. Les efforts en vue d'identifier une compétence générale, indépendante du domaine et susceptible de guider les systèmes dynamiques (l'intelligence opérative) au sein du cadre de la recherche en résolution de problèmes complexes ont également été infructueux; seule une partie du rendement de ces systèmes peut être transférée à un autre système (Funke, 1991). En revanche, des recherches réalisées auprès des élèves de la troisième à la douzième années ont révélé que l'aptitude à résoudre des problèmes s'améliorait manifestement dans des conditions de formation bien rodées et qu'il était possible de réaliser un transfert appréciable entre différents problèmes (Reeff *et coll.*, 1989, 1992, 1993; Regenwetter, 1992; Regenwetter et Müller, 1992; Stirner, 1993).

La résolution de problèmes dépend de la connaissance des notions et des faits (connaissance déclarative) ainsi que des règles et des stratégies (connaissance procédurale) dans un domaine donné. S'il est évident, à la lumière des recherches passées, que la connaissance déclarative du domaine problème peut favoriser considérablement la réussite des stratégies de résolution de problèmes, la connaissance procédurale est tout aussi essentielle. La quantité de connaissances préalables pertinentes dont on dispose peut également expliquer le rapport entre l'intelligence et le rendement en résolution de problèmes, présenté dans les travaux de Raaheim (1988) et Leutner (1999). Les personnes n'ayant absolument aucune connaissance préalable pertinente sont incapables d'approfondir la situation problématique et de planifier systématiquement une solution; elles doivent plutôt procéder par tâtonnements. Celles qui connaissent déjà très bien la tâche sont en mesure de l'aborder de manière systématique. L'aptitude mentale générale, mesurée par des tâches de raisonnement, n'intervient dans aucun de ces cas. Lorsque la tâche est moyennement familière à la personne qui résout des problèmes, celle-ci peut mettre en œuvre des stratégies de raisonnement analytiques. L'approche adoptée pour l'évaluation de la résolution de problèmes dans le cadre de l'EIACA est fondée sur la notion de tâches (moyennement) familières. Dans un contexte assez familier, les problèmes à résoudre sont suffisamment peu « explicites » pour ne pas être perçus comme des tâches purement routinières. En revanche, les connaissances préalables spécifiques au domaine sont assez limitées pour faire des techniques de raisonnement analytiques le principal outil cognitif de résolution des problèmes.

4.5.1 Définir les caractéristiques des tâches

Comment définir des problèmes réels contextualisés et les transformer en un ensemble de tâches d'évaluation? Après avoir passé en revue les diverses approches adoptées lors de recherches antérieures pour mesurer la résolution de problèmes, les concepteurs de l'EIACA ont opté pour une approche par projet. L'approche par projet peut s'avérer un instrument

puissant lorsqu'il s'agit d'évaluer les compétences en résolution de problèmes analytiques dans des contextes réels et courants, et ce, pour plusieurs raisons. Il est à la fois important et pertinent, pour la vie professionnelle et privée de l'adulte, que celui-ci puisse résoudre des problèmes dans un contexte de projet. De plus, l'approche par projet a été appliquée avec succès dans d'autres évaluations à grande échelle et elle peut se réaliser en tant qu'outil de type crayon-papier, ce qui a une importance déterminante pour les enquêtes contemporaines à grande échelle. En outre, l'approche par projet utilise différentes étapes de la résolution de problèmes pour orienter la mise au point des items de test proprement dits. Dans la foulée de Pólya (1945, 1980), le processus de résolution de problèmes a souvent été réduit aux étapes suivantes :

- ❖ Définir le but.
- ❖ Analyser la situation et construire une représentation mentale.
- ❖ Concevoir une stratégie et planifier les mesures à prendre.
- ❖ Exécuter le plan, en y incorporant le contrôle et – s'il y a lieu – la modification de la stratégie.
- ❖ Évaluer le résultat.

Les diverses étapes définissent le cours de l'action d'un projet « courant ». À chacune de ces étapes correspondent un ou plusieurs items ou tâches. Ainsi, les répondants exécutent les tâches individuelles qui ont été identifiées comme des étapes à franchir dans le cadre de leur projet (par exemple, « planifier une réunion de famille » ou « rénover un pavillon »). En incorporant les tâches individuelles dans le contexte d'une action, on obtient une grande authenticité contextuelle. Même si elles font toutes partie d'un projet cohérent et exhaustif, les tâches individuelles sont conçues pour être résolues indépendamment les unes des autres et sont appelées à varier en complexité et en difficulté globale pour les adultes.

Comme l'évaluation des compétences en résolution de problèmes a des origines relativement récentes, il convient de rendre compte en détail du processus de construction. Le tableau A1 présente un aperçu des étapes de la résolution de problèmes qui correspondent aux diverses étapes susmentionnées. Il énumère divers aspects et composantes de chacune des étapes de la résolution de problèmes.

Tableau A1 Résolution de problèmes : étapes et mesures à prendre

Définir les buts	<ul style="list-style-type: none">• Fixer les buts.• Reconnaître les buts à atteindre et préciser les motifs importants de la décision.• Reconnaître les buts/souhaits contradictoires et ceux qui sont compatibles.• Attribuer des priorités aux buts/souhaits.
Analyser la situation	<ul style="list-style-type: none">• Sélectionner, obtenir et évaluer l'information.<ul style="list-style-type: none">⇒ De quelle information a-t-on besoin, quelle information est déjà disponible, laquelle manque toujours et laquelle est superflue?⇒ Où et quand peut-on obtenir l'information?⇒ Comment faut-il interpréter l'information?• Identifier les personnes (possédant les connaissances et les compétences nécessaires) qui doivent participer à la résolution du problème.• Choisir les outils à utiliser.• Reconnaître les conditions (contraintes temporelles, par exemple) dont il faut tenir compte.
Planifier la solution	<ul style="list-style-type: none">• Reconnaître les étapes à suivre.• Décider de la séquence des étapes (par exemple, les questions à l'ordre du jour).• Coordonner le travail et les délais.• Procéder à une analyse comparative de différents plans (reconnaître le plan qui convient à la réalisation des buts).• Adapter le plan aux changements de conditions.• Sélectionner un plan.
Exécuter le plan	<ul style="list-style-type: none">• Réaliser chacune des étapes (par exemple, rédiger une lettre, remplir un formulaire, faire des calculs).
Évaluer les résultats	<ul style="list-style-type: none">• Évaluer si, et dans quelle mesure, on a atteint la cible.• Reconnaître les erreurs.• Identifier pourquoi les erreurs ont été commises.• Évaluer les conséquences des erreurs.

La construction d'une banque de tâches d'évaluation qui correspondraient à ces cinq étapes comportait elle-même plusieurs étapes. Premièrement, il fallait choisir des projets pertinents qui conviendraient aux adultes ayant des antécédents scolaires variables et qui s'appliqueraient au plus grand nombre possible de personnes au sein du groupe cible. Puis, les concepteurs devaient établir et définir la situation problématique et une séquence d'actions correspondant au modèle. Troisièmement, ils devaient constituer une banque d'items correspondant à chaque étape et faisant appel à certains processus; il fallait notamment mettre au point les bonnes réponses et les éléments de distraction pertinents pour les items à choix multiple et établir une clé de correction et des guides de notation pour les tâches ouvertes.

4.5.2 Caractériser les tâches de résolution de problèmes

L'EIACA comportait au total quatre projets, dont vingt tâches consacrées à l'évaluation de la résolution de problèmes. On a ainsi produit 19 items à noter qui s'échelonnaient de 199 à 394 le long de l'échelle et, à l'instar des tâches de littératie et de numératie, leur emplacement était déterminé par les tendances des bonnes et des mauvaises réponses chez les adultes des pays participants. Au lieu de publier l'un des quatre projets qui ont servi à l'EIACA, nous allons décrire une échelle hypothétique de compétences en résolution des problèmes analytiques qui a été testée à partir des données de l'étude pilote et présenter un exemple, tiré de l'étude pilote, qui n'a pas servi à l'évaluation principale. On a décrit des modèles semblables dans le cadre de

travail d'autres évaluations à grande échelle des compétences en résolution de problèmes, comme le test projet pour Hambourg, en Allemagne (Ebach, Klieme et Hensgen, 2000) et l'évaluation de l'étude PISA 2003 de la résolution de problèmes transdisciplinaires (OCDE, sous presse).

Pour les besoins de l'EIACA, quatre niveaux de compétence en résolution de problèmes sont postulés :

Niveau 1

À un niveau très élémentaire, une personne peut maîtriser des tâches concrètes et limitées en appliquant un raisonnement pratique axé sur le contenu. À ce niveau, la personne fait appel à des schémas spécifiques axés sur le contenu afin de résoudre les problèmes.

Niveau 2

Le deuxième niveau exige au minimum un niveau rudimentaire de raisonnement systématique. Les problèmes à ce niveau se caractérisent par des buts unidimensionnels bien définis; ils appellent l'évaluation de certaines solutions de rechange au regard de contraintes transparentes et explicites. À ce niveau, la personne fait appel à des opérations logiques et concrètes.

Niveau 3

Au troisième niveau de compétence en résolution de problèmes, la personne est en mesure de faire appel à des opérations formelles (l'ordonnancement, par exemple) afin d'intégrer des buts multidimensionnels ou mal définis et de composer avec des contraintes dépendantes multiples ou non transparentes.

Niveau 4

Au niveau supérieur de compétence, la personne est à même de comprendre, dans son ensemble, un système d'états des problèmes et des solutions envisageables. Ainsi, la cohérence de certains critères, l'interdépendance de séquences multiples d'intervention et d'autres « métacaractéristiques » d'une situation problématique peuvent être considérées systématiquement. C'est également à ce stade que la personne est en mesure d'expliquer comment et pourquoi elle est arrivée à une solution donnée. Ce niveau de compétence exige une pensée critique et une certaine part de métacognition.

L'exemple ci-dessous illustre la réalisation concrète d'un projet. Le projet en question, exclu de l'instrument final de l'EIACA, consiste à exécuter une tâche typique de la résolution de problèmes. Il s'agit de planifier un voyage et une réunion de famille.

Dans l'introduction au projet, le répondant prend connaissance du résumé suivant, où sont décrits le scénario et le problème global :

« Imaginez que vous habitez la ville A. Vos parents sont éparpillés aux quatre coins du pays, et vous aimeriez organiser une réunion de famille, qui durera une journée. Vous décidez de vous réunir dans la ville B, facilement accessible pour tous. Comme vous et vos parents aimez tous la randonnée pédestre, vous décidez de prévoir une longue randonnée dans un parc provincial près de la ville B. Vous avez accepté de prendre en charge la plus grande partie de l'organisation. »

Le répondant reçoit ensuite la liste suivante d'étapes à franchir :

- Fixez la date de la réunion
- Évaluez les recommandations de vos parents en vue de la randonnée
- Planifiez la marche à suivre avant de réserver votre billet d'avion
- Répondez aux questions de vos parents concernant le voyage en avion
- Réservez votre billet d'avion
- Veillez à ce que votre billet ne comporte aucune erreur
- Planifiez le déplacement entre la ville B et l'aéroport

La première tâche de ce projet, soit « Fixez la date de la réunion », constitue un bon exemple d'une tâche de résolution de problèmes, et nous la présentons maintenant comme elle figurerait dans un cahier de test.

Exemple de tâche : Fixez la date de la réunion

La réunion de famille devrait se tenir au mois de juillet.

Vous avez demandé à tous vos parents de vous faire part des dates qui leur conviendraient. Après leur avoir parlé, vous avez dressé la liste des engagements de vos parents pendant le mois de juillet. Votre propre agenda est ouvert devant vous. Vous constatez que certaines personnes devront arriver un jour avant la réunion et qu'elles ne pourront rentrer chez elles que le lendemain de la réunion.

Veillez consulter la liste des engagements de vos parents et votre propre agenda.

Liste des engagements de vos parents en juillet 1999

Henri	Karine	Pierre	Janette	Anne	François
Vacances à E à compter du 26 juillet	N'importe quel jour de la semaine sauf le jeudi et le 16 juillet	Rendez-vous d'affaires les 2 et 13 juillet et entre les 27 et 29 juillet	Aucun engagement	Ne peut se libérer les 5, 20 et 24 juillet	Doit faire un voyage d'affaires la première semaine complète de juillet, mais n'en connaîtra les dates exactes que la veille de son départ

Henri, Karine et Pierre pourraient arriver le jour prévu pour la réunion, alors que Janette, Anne et François ne peuvent arriver que la veille, en après-midi, et ne rentrer chez eux que le lendemain de la réunion.

Exemple de tâche (suite)

Votre agenda – juillet 1999

Juillet 1999

Jeudi	1	Rendez-vous avec David
Vendredi	2	
Samedi	3	
Dimanche	4	
Lundi	5	
Mardi	6	
Mercredi	7	
Jeudi	8	
Vendredi	9	
Samedi	10	Randonnée pédestre à C
Dimanche	11	
Lundi	12	
Mardi	13	
Mercredi	14	
Jeudi	15	
Vendredi	16	
Samedi	17	
Dimanche	18	
Lundi	19	
Mardi	20	
Mercredi	21	
Jeudi	22	
Vendredi	23	
Samedi	24	
Dimanche	25	
Lundi	26	
Mardi	27	
Mercredi	28	Vacances
Jeudi	29	Vacances
Vendredi	30	Vacances
Samedi	31	

Question 1. Lesquelles des dates suivantes conviendraient à la réunion de famille?

Veillez cocher toutes les dates possibles.

- a 4 juillet
- b 7 juillet
- c 14 juillet
- d 18 juillet
- e 25 juillet
- f 29 juillet

Ce projet illustre bien comment la logique des actions requises est effectivement « traduite » en un cheminement concret d'actions thématiques. L'intrigue sous-jacente – la planification d'un voyage et d'une réunion de famille – constitue une action courante très représentative dans laquelle une vaste majorité des habitants de différents pays peuvent vraisemblablement se reconnaître. Les actions requises elles-mêmes et l'ordre dans lequel elles sont accomplies peuvent s'écarter du modèle normatif d'action complète, comme c'est le cas ici. Le modèle normatif sert de balise, qui est adaptée à chaque contexte. En l'occurrence, par exemple, la tâche intitulée « Évaluez les recommandations de vos parents en vue de la randonnée » correspond à peu près à l'étape « Analyser la situation », la tâche intitulée « Planifiez la marche à suivre avant de réserver votre billet d'avion », à l'étape « Planifier la solution » et la tâche « Réservez votre billet d'avion », à l'étape « Exécuter le plan ».

L'exemple de tâche donne une première idée des structures et des formats des items. La tâche s'amorce habituellement par une brève introduction à la situation, suivie de divers types et quantités de renseignements à assimiler. Dans l'exemple, le répondant qui veut fixer la date de la réunion de famille doit traiter, comparer et incorporer l'information fournie dans la liste des engagements de ses parents, y compris les renseignements complémentaires, ainsi que son propre agenda. En l'occurrence, l'information est surtout textuelle et sous forme de tableaux. Il faut donner plus d'une réponse à la question à choix multiple, bien que le nombre de bonnes réponses ne soit pas précisé.

4.6 Conclusion

Le présent chapitre propose un bref aperçu des cadres de travail adoptés pour mettre au point les tâches qui ont servi à mesurer la compréhension de textes suivis et de textes schématiques, la numératie et la résolution de problèmes lors de l'EIACA ainsi qu'à comprendre le sens des résultats à l'égard des compétences en littératie comparatives des adultes. Ces cadres tiennent compte d'une série de variables qui, à la lumière des résultats, sous-tendent l'accomplissement fructueux d'une vaste gamme de tâches. Ensemble, elles fournissent un moyen pour passer d'une interprétation des résultats d'enquête centrée sur des tâches prises isolément ou sur un seul nombre, à une interprétation permettant de définir des niveaux de capacités assez généralisés pour valoir à l'égard de l'ensemble des évaluations et des groupes. Au fur et à mesure qu'on s'intéresse moins aux comportements discrets ou aux observations isolées et davantage à des interprétations significatives du rendement, on obtient un niveau de mesure supérieur (Messick, 1989).

4.6.1 Quelques critères analytiques

Les niveaux de compétence présentés dans l'ensemble de données de l'EIACA permettent non seulement d'étudier la progression des exigences de l'information à traiter sur chacune des échelles, mais aussi d'expliquer comment les compétences démontrées par les répondants reflètent leur probabilité de répondre correctement au vaste éventail de tâches utilisées dans cette évaluation ainsi qu'à d'autres tâches présentant les mêmes caractéristiques. En pratique, cela signifie que les personnes dont le rendement est de 250 sur chaque échelle devraient pouvoir accomplir les tâches moyennes des niveaux 1 et 2 avec un degré de compétence élevé, c'est-à-dire avec une probabilité moyenne d'au moins 80 % de répondre correctement. Cela ne signifie pas qu'elles sont incapables d'effectuer les tâches des niveaux 3 et plus. On s'attendrait à ce qu'elles y arrivent parfois, mais pas toujours.

Où sont passés les analphabètes?

À l'instar de l'enquête antérieure, l'EIACA de 2003 conceptualise les compétences selon un continuum qui indique dans quelle mesure les adultes utilisent l'information afin de fonctionner dans la société et dans l'économie. Rappelons que l'EIACA ne mesure pas l'absence de compétences, mais plutôt les connaissances et les compétences liées aux quatre domaines par rapport à une gamme étendue de capacités. Par conséquent, les résultats ne peuvent servir à classer les groupes de population en « personnes alphabétisées » et en « analphabètes ».

D'après les résultats de l'EIAA de 1994 pour les deux échelles communes à l'EIACA de 2003, les tableaux 4.1 et 4.2 montrent la probabilité que des personnes dont les résultats se situent à certains points des échelles de compréhension de textes suivis et de textes schématiques répondent correctement à des tâches de difficulté variable. Par exemple, un lecteur dont la note en compréhension de textes suivis est de 150 a moins de 50 % des chances de répondre correctement aux tâches de niveau 1. Les personnes dont le rendement est de 200, par contre, ont une probabilité d'environ 80 % de répondre correctement à ces tâches.

Sur le plan des exigences des tâches, on peut supposer que les adultes qui obtiennent une note de 200 sur l'échelle des textes suivis sont probablement capables de repérer un élément d'information dans un texte bref en l'absence d'éléments distrayeurs ou si le texte contient des informations plausibles mais incorrectes qui sont éloignées de la bonne réponse. Toutefois, ces personnes risquent d'éprouver beaucoup plus de difficulté avec des tâches des niveaux 2 à 5. Par exemple, elles n'auraient que 40 % de chances d'effectuer correctement une tâche moyenne de niveau 2, 18 % de chances de réussir des tâches de niveau 3 et seulement 7 % de chances avec des tâches des niveaux 4 et 5.

À l'opposé, les répondants qui obtiennent un résultat de 300 sur l'échelle des textes suivis ont à peu près 80 % de chances ou plus de réussir des tâches des niveaux 1, 2 et 3. Cela signifie qu'ils sont capables d'accomplir des tâches qui les amènent à faire des inférences de niveau peu élevé ou à tenir compte de renseignements conditionnels. Ils peuvent également intégrer ou comparer et distinguer des éléments d'information facilement identifiables dans le texte. Par contre, ils risquent d'éprouver de la difficulté avec les tâches qui les obligent à faire des inférences complexes fondées sur le texte ou à traiter des types d'information plus abstraits. Ces tâches plus difficiles peuvent également les obliger à aller au-delà du texte pour faire appel à des connaissances moins familières ou plus spécialisées. En moyenne, ils ont une probabilité d'environ 50 % d'effectuer correctement les tâches de niveau 4; avec les tâches de niveau 5, leurs chances de répondre correctement tombent à 40 %.

On peut interpréter de manière semblable les données concernant les échelles de compréhension de textes schématiques et de textes au contenu quantitatif. Par exemple, une personne qui se situe à 200 sur l'échelle des textes au contenu quantitatif a, en moyenne, 67 % de chances de répondre correctement aux tâches de niveau 1. Sa probabilité de réussite tombe à 47 % pour les tâches de niveau 2, à 21 % pour les tâches de niveau 3, à 6 % pour les tâches de niveau 4 et à seulement 2 % pour les tâches de niveau 5. De même, les lecteurs qui ont une note de 300 sur l'échelle des textes au contenu quantitatif auraient une probabilité de 92 % ou plus de répondre correctement aux tâches des niveaux 1 et 2. Leur probabilité moyenne tomberait à 81 % pour les tâches de niveau 3, à 57 % pour les tâches de niveau 4 et à 20 % pour les tâches de niveau 5.

Tableau 4.1 Probabilités moyennes de réussite, échelle des textes suivis

Textes suivis	Pour certains résultats				
	150	200	250	300	350
			%		
1	48	81	95	99	100
2	14	40	76	94	99
3	6	18	46	78	93
4	2	7	21	50	80
5*	2	6	18	40	68

* En fonction d'une seule tâche

Source : Enquête sur l'alphabétisation des adultes (1994).

Tableau 4.2 Probabilités moyennes de réussite, échelle des textes schématiques

Textes schématiques	Pour certains résultats				
	150	200	250	300	350
			%		
1	40	72	94	99	100
2	20	51	82	95	99
3	7	21	50	80	94
4	4	13	34	64	85
5*	<1	1	3	13	41

* En fonction d'une seule tâche

Source : Enquête sur l'alphabétisation des adultes (1994).

Les compétences dans chaque domaine sont mesurées selon une échelle continue. Chaque échelle de compétence va de zéro jusqu'à un maximum théorique de 500 points (avec quatre décimales de précision). Une personne dont les connaissances la situent à un échelon donné de l'échelle a 80 % de probabilité de réussir une tâche qui comporte ce même niveau de difficulté.

D'un point de vue analytique, des statistiques sommaires utiles peuvent être produites en vue de décrire les compétences des populations, notamment une note moyenne globale. Les populations affichant des notes moyennes similaires peuvent cependant avoir des nombres relativement différents d'adultes à rendement faible ou élevé. Par conséquent, il est également possible d'examiner la façon dont les notes sont distribuées dans les populations en utilisant les rangs-centiles. Les rangs-centiles sont les niveaux sous lesquels se situe un pourcentage

particulier d'adultes. Par exemple, le résultat au cinquième centile est celui sous lequel se situent 5 % des adultes d'une population donnée. Les écarts entre les rangs-centiles nous renseignent sur les différences qui existent dans les populations par rapport aux compétences. Les lecteurs trouveront au chapitre 5 des renseignements plus détaillés sur l'utilisation des valeurs plausibles dans leur analyse.

Les résultats de l'EIACA sont également groupés selon des niveaux de compétence représentant un ensemble de tâches à difficulté croissante. Pour les domaines de la compréhension de textes suivis et de textes schématiques, de même que pour le domaine de la numératie, les spécialistes ont défini cinq grands niveaux de difficulté, chacun correspondant à une fourchette similaire (mais non équidistante) de résultats. En ce qui concerne le domaine de la résolution de problèmes, quatre grands niveaux de difficulté ont été définis. Dans chaque domaine, le niveau 1 indique le niveau le plus faible de compétence, et les niveaux 4 (pour la résolution de problèmes) et 5 (pour les autres domaines), les niveaux les plus élevés.

Il importe, pour des motifs tant analytiques qu'opérationnels, de définir un « niveau souhaité » de compétence pour faire face aux demandes grandissantes de compétences dans une économie axée sur le savoir et l'information. Le niveau 3 de performance est généralement utilisé comme point de repère minimum parce que dans les pays développés, une performance au-delà du niveau 2 est généralement associée à un certain nombre de rendements positifs. Ceux-ci comprennent une participation accrue à la vie sociale, des chances accrues de succès économique et d'indépendance, une amélioration des chances de participer à l'apprentissage continu et d'améliorer son niveau personnel de littératie (Kirsch, I. *et coll.*, 1993; Murray, T.S. *et coll.*, 1997; Tuijnman, A., 2001), alors que les individus qui performant aux niveaux 1 et 2 ne maîtrisent généralement pas encore les compétences minimales de base permettant d'atteindre des niveaux supérieurs de performance (Strucker, J., Yamamoto, K., 2005).


Une analyse secondaire des données de l'EIAA de 1994 a permis de constater que l'écart des rendements entre les niveaux 2 et 3 des échelles de compréhension de textes suivis, de textes schématiques et de textes au contenu quantitatif est considérable et correspond à une différence importante quant au nombre d'avantages mesurables dont bénéficient les citoyens des pays de l'OCDE (OCDE et DRHC, 1997). Les résultats de l'analyse préliminaire des données de l'EIACA, y compris la nouvelle échelle mesurant la numératie, sont conformes à cette constatation. C'est pourquoi il est parfois utile de fixer les échelles à un point de séparation situé entre les niveaux 2 et 3, mettant ainsi l'accent sur la répartition des compétences au-dessus et au-dessous de ce seuil pour les domaines de la compréhension de textes suivis et de textes schématiques ainsi que pour celui de la numératie. À l'opposé, l'interprétation du domaine de la résolution de problèmes est plus complexe, et aucun seuil « souhaitable » n'a encore été établi; il conviendrait donc de situer le point de séparation au niveau 1 en attendant de trouver un seuil plus précis.

Le chapitre 8 présente d'autres outils et d'autres techniques d'utilisation des valeurs plausibles et des poids de rééchantillonnage nécessaires pour produire des estimations exactes des erreurs-types liées à chaque estimation ponctuelle.

Références

- Almond, R. G., and Mislevy, R. J. (1998). Graphical models and computerized adaptive testing. (TOEFL Tech. Rep. No. 14). Princeton, NJ: Educational Testing Service.
- Baker, D., and Street, B. (1994). Literacy and numeracy: Concepts and definitions. In T. Husen and E. A. Postlethwaite (Eds.), *Encyclopedia of education*. New York: Pergamon Press.
- Beazley, K. (1984). *Education in Western Australia: Report of the Committee of Inquiry into Education in Western Australia*. Education Department of Western Australia.
- Coben, D., O'Donoghue, J., and FitzSimons, G. E. (Eds.) (2000). *Perspectives on adults learning mathematics: Theory and practice*. London: Kluwer Academic Publishers.
- Cockcroft, W.H. (1982). *Report of the Committee of Inquiry into the Teaching of Mathematics in Schools*. London: HMSO.
- Cook-Gumperz, J., and Gumperz, J. (1981). From oral to written culture: The transition to literacy. In M. Whitman (Ed.), *Writing: The nature, development and teaching of written communication: Vol. 1*. Hillsdale, NJ: Erlbaum.
- Crandall, J. (1981, December). Functional literacy of clerical workers: Strategies for minimizing literacy demands and maximizing available information. Paper presented at the annual meeting of the American Association for Applied Linguistics, New York.
- Diehl, W. (1980). *Functional literacy as a variable construct: An examination of the attitudes, behaviours, and strategies related to occupational literacy*. Unpublished doctoral dissertation, Indiana University.
- Dossey, J.A. (1997). "Defining and measuring quantitative literacy". In L.A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America*. New York: College Entrance Examination Board.
- Fey, James T. (1990). "Quantity" In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.
- Frankenstein, M. (1989). *Relearning mathematics: A different third 'R' – Radical maths*. London: Free Association Books.
- Gal, I. (1997). Numeracy: Imperatives of a forgotten goal. In L.A. Steen (Ed.), *Why numbers count: quantitative literacy for tomorrow's America* (pp. 36-44). New York: The College Board.
- Gal, I. (2000). The numeracy challenge. In I. Gal (Ed.), *Adult numeracy development: Theory, research, practice* (pp. 1-25). Cresskill, NJ: Hampton Press.
- Gal, I. (2002). Adult Statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Jacob, E. (1982). *Literacy on the job: Final report of the ethnographic component of the industrial literacy project*. Washington, DC: Center for Applied Linguistics.
- Johnston, B. (1994, Summer). Critical numeracy? In *Fine print*, Vol. 16, No. 4.
- Heath, S.B. (1980). The functions and uses of literacy. *Journal of Communication*, 30, 123–133.
- Kirsch, I.S., and Guthrie, J.T. (1984a). Adult reading practices for work and leisure. *Adult Education Quarterly*, 34(4), 213–232.
- Kirsch, I.S., and Guthrie, J. T. (1984b). Prose comprehension and text search as a function of reading volume. *Reading Research Quarterly*, 19, 331–342.

- Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding What Was Measured* (ETS Research Report RR-01-25). Princeton, NJ: Educational Testing Service.
- Marr, B., and Tout, D. (1997). *A numeracy curriculum: Australian Association of Mathematics Teachers (AAMT) conference proceedings*. Melbourne: AAMT.
- Messick, S. (1989). *Validity*. In R. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher*, 32(2), 13-23.
- Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. *Reading Research Quarterly*, 17(3), 400–419.
- Miller, P. (1982). Reading demands in a high-technology industry. *Journal of Reading*, 26(2), 109–115.
- Mislevy, R.J. (September, 2000). Leverage points for improving educational assessment. Paper submitted to National Center for Research on Evaluation, Standards, and Student Testing (CRESST) as part of award #R305B60002 from the US Department of Education, Office of Educational Research and Improvement.
- Mosenthal, P.B., and Kirsch, I.S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent and Adult Literacy*, 41(8), 638–657.
- Murray, T.S., Clermont, Y. and Binkley, M. (Eds.) *The Adult Literacy and Life Skills Survey: Aspects of Design, Development and Validation*. Canada: Statistics Canada, in press.
- Murray, T.S., Kirsch, I.S., and Jenkins, L. (1998). *Adult Literacy in OECD Countries: Technical report on the First International Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Organization for Economic Co-operation and Development. (1992). *Adult illiteracy and economic performance*. Paris, France: Author.
- Rutherford, F.J and Ahlgren, A. (1990). *Science for all Americans*. New York: Oxford University Press.
- Rychen, D.S. and Salganik, L.H. (Eds.) *Key Competencies for a Successful Life and a Well-Functioning Society*. Cambridge, MA: Hogrefe and Huber Publishers, 2003.
- Senechal, Majorie (1990) "Shape" In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.
- Scribner, S., and Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Steen, L.A. (Ed). (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Research Council.
- Steen, L.A. (2001). *Mathematics and democracy: the case for quantitative literacy*. USA: National Council on Education and the Disciplines.
- Sticht, T.G. (Ed.). (1975). *Reading for working: A functional literacy anthology*. Alexandria, VA: Human Resources Research Organization.

- 
- Sticht, T.G. (1978). Literacy and vocational competency (Occasional Paper 39, National Center for Research in Vocational Education). Columbus, OH: Ohio State University.
- Sticht, T.G. (1982, January). Evaluation of the reading potential concept for marginally literate adults. (Final Report FR–ET50–82–2). Alexandria, VA: Human Resources Research Organization.
- Szwed, J. (1981). The ethnography of literacy. In M. Whitman (Ed.), *Writing: The nature, development, and teaching of written communication: Vol. 1*. Hillsdale, NJ: Erlbaum.
- Tobias, S. (1993). *Overcoming math anxiety*. New York: Norton.
- Venezky, R.L. (1983). The origins of the present-day chasm between adult literacy needs and school literacy instruction. *Visible Language*, 16, 113–136.

5.0 Méthodologie de l'enquête

Le Canada ayant participé au premier cycle de collecte de données de l'Enquête sur la littératie et les compétences des adultes (ELCA), le volet canadien, sous le titre d'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA), a été mis en œuvre selon les normes énoncées dans les lignes directrices relatives à l'ELCA. Ces normes établissent les exigences minimales concernant le plan d'enquête et la mise en œuvre de toutes les étapes de l'enquête, de la planification à la documentation.

5.1 Plan d'évaluation

Comme nous l'avons mentionné dans le chapitre précédent, les éléments de la littératie et des compétences fonctionnelles visés par l'EIACA sont évalués au moyen de mesures psychométriques des compétences dans les domaines de la compréhension de textes suivis, de la compréhension de textes schématiques, de la numératie et de la résolution de problèmes. Chaque question ou ensemble de questions connexes est axé sur un item. L'ensemble des items est organisé en sous-ensembles de tâches, ou blocs, de 30 minutes : quatre blocs d'items de littératie (compréhension de textes suivis et de textes schématiques), deux blocs d'items de numératie et deux blocs d'items de résolution de problèmes. Au moyen d'un plan d'évaluation par blocs incomplets équilibrés, les blocs sont combinés par paires pour constituer 28 livrets.

Les livrets ont été distribués selon le plan d'enquête à l'ensemble de l'échantillon canadien, soit bien au-delà du nombre minimal de répondants à évaluer dans chaque langue. Comme il faut parfois plus d'une heure pour administrer un livret, on a demandé à chaque répondant de n'en remplir qu'un; personne n'était tenu d'effectuer l'ensemble complet de tâches. La répartition des blocs en plusieurs livrets a réduit considérablement le fardeau des répondants. On a également suivi de près la collecte des données afin d'obtenir à peu près le même nombre d'unités complètes pour chaque livret de tâches, à l'exception de quatre livrets contenant uniquement des items de numératie ou des items de résolution de problèmes : ces livrets nécessitaient un plus grand nombre d'unités complètes.

5.2 Couverture de la population

La population cible est composée de tous les résidents canadiens qui avaient 16 ans et plus au moment de la collecte de données, à l'exclusion des résidents à long terme d'établissements, des familles de membres des forces armées vivant dans des bases militaires et des personnes vivant dans des réserves indiennes.

En raison de contraintes opérationnelles, on a également exclu de la population cible les résidents de régions peu peuplées. Même combinées à celles qui sont mentionnées plus haut, ces exclusions représentaient au plus 2 % de la population totale, ce qui respectait largement l'exigence internationale selon laquelle le sous-dénombrement ne devait pas dépasser 5 %. On estime que la couverture de l'enquête s'est établie à 98,5 % à l'échelle nationale et entre 95 %

et près de 100 % à l'échelle provinciale. Dans les territoires du Nord, on a obtenu des niveaux de couverture réduits (de 70 % à 90 %) parce que seules les collectivités couvertes par l'Enquête sur la population active étaient comprises. Le tableau C1 présente le taux de couverture estimatif par province et par territoire.

Tableau 5.1 Taux de couverture estimatif par province ou territoire, EIACA de 2003

	Taux de couverture estimatif (%)
Terre-Neuve-et-Labrador	98,1
Île-du-Prince-Édouard	99,7
Nouvelle-Écosse	99,3
Nouveau-Brunswick	98,8
Québec	98,9
Ontario	99,3
Manitoba	95,3
Saskatchewan	95,3
Alberta	98,2
Colombie-Britannique	97,1
Yukon	90,0
Territoires du Nord-Ouest	86,0
Nunavut	70,0
Canada	98,5

Source : Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Le Canada est le seul pays à avoir inclus dans sa population cible les adultes de plus de 65 ans; il en avait le loisir, puisque le plan d'échantillonnage satisfaisait déjà à l'exigence minimale internationale concernant la taille de l'échantillon de personnes de 16 à 65 ans.

5.3 Base de sondage

On a choisi comme base de sondage le plus récent Recensement de la population et du logement, en date du 15 mai 2001. Le choix de cette base existante permettait d'utiliser les caractéristiques déclarées des ménages pour repérer les logements les plus susceptibles de comprendre une personne appartenant à certaines sous-populations cibles. Ce renseignement auxiliaire a grandement favorisé l'efficacité du plan d'échantillonnage. Plus précisément, la base de sondage était composée des ménages dénombrés au moyen du questionnaire complet du recensement (échantillon de 20 %).

On aurait pu prélever l'échantillon de base national, les échantillons complémentaires provinciaux et les échantillons supplémentaires liés à l'âge à partir des ménages ayant répondu au questionnaire abrégé du recensement, mais on avait besoin des données du questionnaire complet pour repérer le reste des sous-populations particulières. Dans le cas des échantillons des langues minoritaires, la qualité des réponses au questionnaire complet est jugée supérieure à celle des réponses au questionnaire abrégé. La présence de questions sur la connaissance et l'utilisation de langues autres que la langue maternelle (langue apprise la première et encore comprise) permet aux répondants de mieux caractériser leur profil linguistique.

5.4 Plan d'échantillonnage

On a employé un plan d'échantillonnage probabiliste à plusieurs degrés pour prélever l'échantillon à partir de la base du recensement. Le plan a servi à constituer des échantillons distincts pour les deux langues officielles, l'anglais et le français. De plus, on a accru la taille de l'échantillon afin de produire des estimations concernant un certain nombre de sous-groupes de population. Des ministères provinciaux et d'autres organismes ont commandité des échantillons supplémentaires pour accroître la base ou pour cibler certaines sous-populations, dont les jeunes (de 16 à 24 ans au Québec, de 16 à 29 ans en Colombie-Britannique), les adultes de 25 à 64 ans au Québec, les minorités linguistiques (les anglophones au Québec, les francophones ailleurs), les nouveaux immigrants et les immigrants établis, les Autochtones urbains et les résidents des territoires du Nord.

Dans chacune des dix provinces, on a en outre divisé la base du recensement en deux strates, l'une urbaine, l'autre rurale. On a restreint la strate urbaine aux centres urbains d'une certaine taille, déterminée d'après le recensement antérieur. Le Système automatisé de regroupement des territoires (SARTE) de Statistique Canada a réparti le reste de la base de sondage en unités primaires d'échantillonnage (UPÉ). On a créé les UPÉ pour contenir une population suffisante en fonction du nombre de logements à l'intérieur d'un secteur limité et raisonnablement compact. De plus, on a ajouté une indication générale du niveau de scolarité de la population d'après le Recensement de 1996 pour créer des UPÉ qui reflétaient la répartition des niveaux de scolarité de leur province.

On a procédé à une deuxième stratification, implicite, en sélectionnant systématiquement des ménages dans chaque échantillon. On a retenu le niveau de scolarité le plus élevé de chaque adulte au sein du ménage, d'après la base du recensement, afin de représenter la catégorie dominante parmi quatre grands niveaux : 1) sans diplôme d'études secondaires, 2) diplôme d'études secondaires ou études postsecondaires partielles, 3) diplôme d'études collégiales et 4) diplôme d'études universitaires. Le niveau de scolarité officiel n'est pas le seul déterminant du rendement dans les évaluations de la littératie, mais il est le principal (OCDE et Statistique Canada, 2000). À l'intérieur des régions géographiques, on a ordonné les ménages selon le niveau de scolarité avant de prélever l'échantillon, ce qui a accru la capacité de représenter un éventail des antécédents scolaires.

On a réparti l'échantillon en strates selon la méthode de Neyman, en intégrant un effet prudent du plan de sondage, soit 2 pour la strate rurale et 1,5 pour la strate urbaine. Après la répartition, on a constaté que plusieurs UPÉ des strates rurales étaient suffisamment importantes pour être échantillonnées avec certitude. On a converti ces UPÉ en une nouvelle pseudo-strate urbaine, qu'on a échantillonnée de la même façon que la strate urbaine.

À la dernière étape avant le prélèvement de l'échantillon, on a gonflé la taille des échantillons primaires pour tenir compte du taux de réponse minimal ciblé à l'échelle internationale, soit 70 %, et de la mobilité (en fonction des caractéristiques d'intérêt) de chaque sous-population couverte par un échantillon supplémentaire. On a calculé un taux pondéré en utilisant comme variables substitutives les variables de mobilité sur un an et sur cinq ans d'après le recensement et on a appliqué ce taux au décalage entre le recensement et le début de la collecte de données en mars 2003. On a rajusté ces taux à la baisse dans chaque strate pour tenir compte du remplacement prévu des personnes ayant déménagé par d'autres possédant les mêmes caractéristiques cibles pour chaque échantillon supplémentaire.

À l'intérieur de la strate urbaine, on a utilisé deux degrés d'échantillonnage. Au premier degré, on a sélectionné systématiquement des ménages avec probabilité proportionnelle à la taille. On a défini la mesure de la taille en fonction du nombre d'adultes au sein d'un ménage, le plafond étant fixé à quatre pour l'échantillon de base et à trois pour les échantillons supplémentaires. Au deuxième degré, l'application IPAO a utilisé un algorithme d'échantillonnage aléatoire simple pour sélectionner une personne d'après la liste démographique d'adultes au sein des ménages admissibles. Dans la strate rurale, on a utilisé trois degrés pour prélever l'échantillon. Au premier degré, on a sélectionné des unités primaires d'échantillonnage (UPÉ) avec probabilité proportionnelle à la taille de la population d'après le nombre total d'adultes dans la population observée de chaque échantillon du Recensement de 2001. Aux deuxième et troisième degrés, on a repris la même méthode de sélection à deux degrés que dans le cas de la strate urbaine.

5.5 Précision recherchée

On a proposé trois niveaux de base pour la taille de l'échantillon de l'enquête : une stratégie minimale, une stratégie moyenne et une stratégie maximale, selon la précision recherchée pour chaque estimation. La stratégie minimale correspondait à la capacité d'estimer une proportion (comme la proportion de la population qui se situe à un niveau de compétence donné dans un domaine donné) d'au plus 15 % avec un coefficient de variation d'au plus 16,5 %, c'est-à-dire de qualité acceptable. Lors de l'EIAA de 1994, les plus faibles proportions d'adultes canadiens pour n'importe quel niveau de compétence se situaient au niveau le plus faible : 16,6 % se classaient au niveau 1 en compréhension de textes suivis et une proportion semblable, 16,9 %, se classaient au même niveau en compréhension de textes au contenu quantitatif (numératie). Comme l'EIAA ne couvrait pas le domaine de la résolution de problèmes, on n'obtiendra une bonne estimation nationale de cette proportion qu'après compilation des estimations tirées de l'EIACA. La stratégie moyenne est très semblable à la stratégie minimale, sauf qu'on misait sur la capacité d'estimer une proportion d'au plus 10 %. Enfin, la stratégie maximale correspondait à la capacité d'estimer des proportions d'au plus 25 % à l'intérieur d'un niveau de littératie, à condition que ce niveau représente au moins 15 % de la population totale.

5.6 Taille de l'échantillon

À chaque province correspondait un échantillon de base représentant la population générale. De plus, des ministères provinciaux et d'autres organismes ont commandité des échantillons supplémentaires pour accroître la base de sondage ou pour cibler certaines sous-populations. Le tableau 5.2a montre le nombre prévu de répondants dans chaque échantillon : l'échantillon de base, les jeunes (de 16 à 24 ans au Québec, de 16 à 29 ans en Colombie-Britannique), les adultes de 25 à 64 ans au Québec, les minorités linguistiques (les anglophones au Québec, les francophones ailleurs), les nouveaux immigrants et les immigrants établis, les Autochtones urbains et les résidents des territoires (les Inuits et les non-Inuits dans le cas du Nunavut).

Tableau 5.2a Composition prévue des échantillons avant la collecte, EIACA de 2003

Province ou territoire	Échantillon de base	Jeunes [*]	Adultes ^{**}	Langue	Immigrants	Populations autochtones	Populations non autochtones ^{***}	Total
Terre-Neuve-et-Labrador	1 350	1 350
Île-du-Prince-Édouard	650	650
Nouvelle-Écosse	1 350	1 350
Nouveau-Brunswick	650	760	1 410
Québec	1 110	815	1 885	570	270	4 650
Ontario	1 690	3 000	1 060	5 750
Manitoba	1 350	450	...	700	...	2 500
Saskatchewan	650	700	...	1 350
Alberta	1 350	70	1 420
Colombie-Britannique	1 350	490	280	2 120
Yukon	700	700	1 400
Territoires du Nord-Ouest	450	450	900
Nunavut	700	180	880
Canada	11 500	1 305	1 885	4 780	1 680	3 240	1 340	25 730

Tableau 5.2b Répartition finale des répondants, EIACA de 2003

Province ou territoire	Échantillon de base	Jeunes [*]	Adultes ^{**}	Langue	Immigrants	Populations autochtones	Populations non autochtones ^{***}	Total
Terre-Neuve-et-Labrador	1 299	1 350
Île-du-Prince-Édouard	645	650
Nouvelle-Écosse	1 272	1 350
Nouveau-Brunswick	636	830	1 410
Québec	1 002	693	1 737	465	269	4 650
Ontario	1 502	2 318	1 126	5 750
Manitoba	1 219	386	...	662	...	2 500
Saskatchewan	559	675	...	1 350
Alberta	1 234	73	1 420
Colombie-Britannique	1 190	345	314	2 120
Yukon	615	477	1 400
Territoires du Nord-Ouest	408	410	900
Nunavut	570	107	880
Canada	10 558	1 038	1 737	3 990	1 782	2 930	994	23 038

* On entend par jeunes les personnes de 16 à 24 ans au Québec et celles de 16 à 29 ans en Colombie-Britannique.

** On entend par adultes les personnes de 25 à 64 ans.

*** Au Nunavut, on entend par population non autochtone les personnes autres que les Inuits.

Source : Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Après correction en fonction de la non-réponse et de la mobilité prévue des sous-populations cibles, on a obtenu un échantillon global de plus de 40 000 unités. On a sélectionné les échantillons de manière séquentielle, l'un après l'autre, à partir de l'échantillon de base. Après avoir sélectionné chaque échantillon, on a retiré les ménages choisis de la base de sondage avant de passer aux sélections suivantes, rendant ainsi les échantillons dépendants. On peut considérer la sélection séquentielle de plusieurs échantillons dans une province comme un échantillonnage à phases multiples.

Pour les besoins de l'échelonnement des données de l'évaluation psychométrique, les données d'un répondant sont considérées comme complètes à condition de comprendre au moins les variables du questionnaire de référence en ce qui concerne l'âge, le sexe et le niveau de scolarité.

5.7 Collecte des données

Le plan d'enquête de l'ELCA combinait les techniques de l'évaluation pédagogique à celles d'une enquête-ménage afin d'évaluer la littératie et d'obtenir les renseignements nécessaires pour rendre ces mesures significatives. Les livrets des tâches principales constituaient le dernier d'une série d'instruments de collecte à être appliqués. On a d'abord demandé aux répondants de remplir un questionnaire de présélection qui a servi à dresser une liste pour chaque logement échantillonné. Ce questionnaire de présélection a permis de recueillir suffisamment de données démographiques pour cerner les sous-populations cibles de l'enquête et pour sélectionner au hasard un membre de chaque logement. On a alors demandé au répondant sélectionné de répondre au questionnaire de référence, qui comprenait plusieurs modules de renseignements nécessaires pour rapprocher les compétences évaluées à sa situation économique et sociale. On a posé aux répondants une série de questions sur leur scolarité, leurs activités de lecture à la maison et au travail, leur situation par rapport au marché du travail, leur utilisation des technologies de l'information et des communications, leur participation à la formation continue, ainsi qu'une auto-évaluation de leur niveau de littératie. Il fallait donc compter un temps médian d'environ 35 minutes pour administrer le questionnaire de référence.

Une fois le questionnaire de référence rempli, l'intervieweur a remis au répondant un bref livret de tâches de base comprenant six tâches de lecture relativement simples (livret des tâches de base). Puis, on a remis aux répondants qui avaient réussi les tâches de base un livret de tâches principales, plus difficiles, renfermant un nombre beaucoup plus important de tâches comprenant environ 45 items. L'évaluation n'était pas minutée et on a exhorté le répondant à essayer chaque exercice. On lui a donc accordé le maximum de chances pour faire preuve de ses compétences, même si ces dernières s'avéraient faibles. Tous les répondants devaient essayer les tâches de base puis, le cas échéant, le livret des tâches principales (temps médian pour le remplir : 58 minutes) immédiatement après avoir rempli le questionnaire de référence, ceci afin de vérifier l'incidence de la fatigue sur les outils d'évaluation.

Les livrets des tâches de base et des tâches principales constituaient des évaluations papier et crayon; toutefois, le questionnaire de présélection, le questionnaire de référence, et même l'administration des livrets des tâches de base et des tâches principales ont été menés selon la méthode de l'interview sur place assistée par ordinateur (IPAO). Grâce à un appariement à grande échelle de la base du recensement avec le registre central des adresses, on disposait de numéros de téléphone pour environ 74 % du fichier d'enquête. Dans ce cas, les intervieweurs étaient autorisés à effectuer un contact initial par téléphone pour remplir le questionnaire de présélection, puis à fixer un rendez-vous pour mener une interview sur place avec le répondant sélectionné.

La collecte de données a commencé en 2003 avec la planification des affectations des intervieweurs par les bureaux régionaux qui coordonnaient les activités de collecte. On a effectué les premiers contacts avec les répondants à l'échelle du pays en mars 2003; les dernières interviews ont été menées en août, et tous les documents relatifs à l'enquête ont été retournés au bureau central en septembre 2003.

Pour assurer un haut degré de qualité des données, on a suivi les lignes directrices relatives à l'exécution de l'enquête, tout en adhérant aux normes et méthodes internes de Statistique Canada. Les interviews ont été menées à domicile, de manière impartiale et sans aucune pression. On a assuré la formation et la supervision des intervieweurs en mettant l'accent sur l'importance des précautions à prendre contre les erreurs dues aux non-réponses. Les intervieweurs avaient pour instruction formelle de retourner à plusieurs reprises chez les ménages non répondants afin d'obtenir le plus de réponses possible. On a tout mis en œuvre pour s'assurer que les adresses fournies aux intervieweurs étaient aussi complètes que possible, afin de réduire les problèmes éventuels d'identification des ménages. Enfin, on a supervisé le travail des intervieweurs en exerçant de fréquents contrôles de la qualité au début de la collecte de données, puis moins de contrôles par la suite, et en offrant aux intervieweurs de l'aide durant la période de collecte de données. Au total, le Canada a employé 317 intervieweurs chargés d'interviewer en moyenne 62 répondants.

Comme condition de la participation à l'étude internationale, il fallait saisir et traiter les fichiers en utilisant des méthodes qui assuraient une cohérence logique et des niveaux acceptables d'erreur dans la saisie des données. Plus précisément, on a effectué une vérification complète des notes saisies (en entrant chaque enregistrement deux fois) afin de réduire les taux d'erreur au minimum. L'exactitude de la notation des tests étant essentielle pour assurer un haut degré de qualité des données, il fallait exercer un contrôle intégral par double saisie.

Il fallait coder l'industrie, la profession et le niveau de scolarité à l'aide des codes types suivants : la Classification internationale type, par industrie (CITI), la Classification internationale type des professions (CITP) et la Classification internationale type de l'éducation (CITE). On a fourni des codes pour toutes les questions ouvertes, ainsi que des directives précises relativement au codage des questions.

5.8 Notation des tests

Les personnes chargées de la notation ont reçu une formation intensive relativement à la notation des réponses aux questions ouvertes à l'aide du manuel de notation de l'ELCA. On leur a également fourni un outil pour saisir les questions fermées. Pour favoriser l'exactitude de la notation et la comparabilité entre les pays, les responsables de l'ELCA ont installé un babillard électronique permettant aux correcteurs de chaque pays de poser leurs questions aux spécialistes des domaines. Tous les pays participants avaient accès à ces renseignements et pouvaient rajuster leur notation en conséquence. Pour assurer davantage l'exactitude, on a contrôlé de deux façons la qualité de la notation.

Premièrement, au moins 20 % des tests ont dû faire l'objet d'une deuxième notation. Les lignes directrices relatives à la deuxième notation à l'intérieur d'un pays stipulaient une deuxième notation d'une grande partie des livrets au début du processus afin de cerner et de corriger le plus grand nombre possible de problèmes de notation. Dans un deuxième temps, on a sélectionné une plus petite partie d'un autre tiers des livrets à noter. La dernière étape, considérée comme une mesure de contrôle de la qualité, consistait à noter à nouveau et régulièrement une petite partie des livrets jusqu'à la fin du processus de deuxième notation. Les deux ensembles de notes devaient concorder dans une proportion d'au moins 95 % pour qu'on puisse passer à l'étape suivante du traitement. En fait, dans la plupart des cas, la fiabilité de la notation à l'intérieur d'un pays était supérieure à 95 %. Lorsqu'il y avait des erreurs, on a

demandé au pays de reprendre les livrets et de noter à nouveau toutes les questions présentant des problèmes et tous les tests notés par un correcteur problématique.

Le tableau 6.1 du chapitre suivant montre le haut niveau de concordance des notes à l'intérieur du pays.

Deuxièmement, on a effectué une deuxième notation internationale. Dans une proportion de 10 %, l'échantillon de chaque pays a fait l'objet d'une deuxième notation par les correcteurs d'un autre pays. Par exemple, un échantillon des livrets de tâches des États-Unis a été soumis à une deuxième notation par les personnes qui avaient noté les livrets en anglais du Canada, et inversement. La deuxième notation visait principalement à vérifier qu'aucun pays n'avait effectué une notation uniformément différente de celle des autres pays. Statistique Canada a calculé la fiabilité de la notation entre les pays et le Educational Testing Service de Princeton (New Jersey) a évalué les résultats. Encore une fois, on a exigé une exactitude rigoureuse, soit une concordance de 90 %, pour que les notes soient jugées acceptables. Chaque fois qu'on a décelé un problème, l'échantillon a été soumis à une deuxième notation.

Dans le chapitre suivant, nous abordons en détail la deuxième notation à l'intérieur d'un pays et entre les pays.

5.9 Réponse à l'enquête et pondération


L'échantillon canadien de l'EIACA procédait d'un plan très complexe : stratification, phases multiples, degrés multiples, échantillonnage systématique, échantillonnage avec probabilité proportionnelle à la taille et plusieurs échantillons chevauchants. En outre, il est nécessaire de compenser les cas de non-réponse survenus à divers niveaux. L'estimation des paramètres de population et des erreurs-types correspondantes dépend donc des coefficients de pondération, ou poids. On a calculé deux types de coefficients de pondération : des poids de population, qui servent à produire des estimations de population, et des poids de rééchantillonnage jackknife, qui servent à calculer les erreurs-types correspondantes.

5.9.1 Poids de population

On a calculé les poids de population en quatre étapes : 1) calcul des poids déterminés par le plan d'échantillonnage, 2) redressement par pondération de la non-réponse, 3) intégration des poids des différents échantillons et 4) calibrage.

Les poids déterminés par le plan d'échantillonnage ont été définis comme l'inverse des probabilités de sélection. La probabilité globale de sélection d'une unité d'échantillonnage était le produit de ses probabilités de sélection à chaque phase et à chaque degré de sélection. On a pris en compte la sélection séquentielle de plusieurs échantillons dans une province en intégrant la probabilité qu'une unité sélectionnée dans un échantillon donné n'ait été sélectionnée dans aucun des échantillons déjà sélectionnés.

On a calculé les redressements par pondération de la non-réponse en répartissant d'abord les unités d'échantillonnage en répondants, en ménages hors du champ de l'enquête, en ménages non répondants (ceux pour lesquels on ne disposait pas de données tirées du questionnaire de présélection) et en non-répondants (ceux qui ont répondu au questionnaire de présélection, mais sans qu'on dispose de données sur le répondant sélectionné). On a utilisé



successivement l'algorithme CHAID du logiciel Knowledge-Seeker pour former des catégories de pondération (groupes de réponses homogènes) et corriger en fonction des ménages non répondants et des non-répondants en deux degrés distincts pour chaque province et chaque type d'échantillon. Par la suite, on a corrigé les poids déterminés par le plan d'échantillonnage des répondants au moyen des facteurs calculés à chaque étape afin de représenter tous les répondants.

À cause du chevauchement des divers échantillons étudiés, il a fallu intégrer les poids pour pouvoir produire des estimations à partir de toutes les unités de tous les échantillons. La situation est comparable à celle d'une base multiple, sauf qu'ici, les échantillons sont dépendants. On a intégré les poids à l'aide de la méthode de Hartley pour les bases multiples : on a réparti l'échantillon entier en fonction des sous-populations ciblées dans les échantillons supplémentaires et on a corrigé les poids au moyen des coefficients proportionnels à la taille des divers échantillons réalisés à l'intérieur de la partition.

Enfin, on a calibré séparément les poids dans chaque province ou territoire à l'aide des variables de référence données dans le tableau 5.3. Les tentatives d'inclusion des variables « taille du ménage » et « niveau de scolarité » se sont avérées insatisfaisantes et ont été abandonnées. Les variables retenues avaient été validées au moyen d'appariements des données d'enquête recueillies avec les renseignements disponibles dans la base de sondage. On a imputé quelques données manquantes pour les besoins des variables de calibrage. Les chiffres du recensement pour toutes les variables de calibrage au niveau des secteurs de dénombrement ont été gonflés selon la croissance mesurée entre les totaux provinciaux du recensement pour l'âge et le sexe et les chiffres de population correspondants au 21 juin 2003. Cette date de référence représentait une approximation du point milieu de la collecte de données en fonction du nombre de jours civils et de réponses complètes.

Tableau 5.3 Variables de référence par province ou territoire

Province ou territoire	Variabes de calibrage
Terre-Neuve-et-Labrador	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (St. John's)
Île-du-Prince-Édouard	Groupe d'âge selon le sexe
Nouvelle-Écosse	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Halifax)
Nouveau-Brunswick	Groupe d'âge selon le sexe, strate selon le sexe, francophones selon le sexe
Québec	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Montréal, Québec), anglophones selon le sexe, immigrants selon le sexe
Ontario	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Toronto, Ottawa, groupe de six RMR), francophones selon le sexe, immigrants selon le sexe
Manitoba	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Winnipeg), francophones selon le sexe, Autochtones urbains selon le sexe
Saskatchewan	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Regina, Saskatoon), Autochtones urbains selon le sexe
Alberta	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Calgary, Edmonton), immigrants selon le sexe
Colombie-Britannique	Groupe d'âge selon le sexe, strate selon le sexe, RMR/AR (Vancouver), immigrants selon le sexe
Yukon	Groupe d'âge selon le sexe, RMR/AR (Whitehorse), Autochtones selon le sexe
Territoires du Nord-Ouest	Groupe d'âge selon le sexe, RMR/AR (Yellowknife), Autochtones selon le sexe
Nunavut	Groupe d'âge selon le sexe, Inuits selon le sexe

Note : On entend par groupes d'âge les personnes de 16 à 25 ans, de 26 à 35 ans, de 36 à 45 ans, de 46 à 55 ans, de 56 à 65 ans et de 66 ans et plus; toutefois, les jeunes groupes d'âge se définissent comme les personnes de 16 à 24 ans et de 25 à 35 ans au Québec et de 16 à 29 ans et de 30 à 45 ans en Colombie-Britannique.

Source : Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003.

Le tableau 5.4 résume la taille des échantillons et les taux de réponse de chaque province :

Tableau 5.4 Échantillon obtenu et taux de réponse par province

Région	Taille de l'échantillon et taux de réponse - Résumé				
	Population cible	Taille de l'échantillon initial	Unités hors champ ¹	Nombre de répondants ²	Taux de réponse ³
Terre-Neuve-et-Labrador	431 646	2 001	98	1 299	68,3
Île-du-Prince-Édouard	111 274	929	48	645	73,2
Nouvelle-Écosse	747 447	1 928	103	1 272	69,7
Nouveau-Brunswick	599 679	2 126	181	1 466	75,4
Québec	5 994 042	7 327	939	4 166	65,2
Ontario	9 621 290	9 600	1 613	4 946	61,9
Manitoba	852 805	4 186	767	2 267	66,3
Saskatchewan	741 829	2 542	640	1 234	64,9
Alberta	2 428 842	2 067	130	1 307	67,5
Colombie-Britannique	3 313 115	3 291	429	1 849	64,6
Yukon	20 739	2 000	249	1 092	62,4
Territoires du Nord-Ouest	26 541	1 286	110	818	69,6
Nunavut	12 592	1 257	119	677	59,5
Canada	24 901 841	40 540	5 426	23 038	65,6

1. Les unités hors du champ de l'enquête sont celles qui ont été codées comme suit : résidents non admissibles, logement introuvable, logement en construction, inoccupé ou saisonnier, ou unités en double.

2. Pour les besoins de l'échelonnement des données de l'évaluation psychométrique, les données d'un répondant sont considérées comme complètes à condition de comprendre au moins les variables du questionnaire de référence en ce qui concerne l'âge, le sexe et le niveau de scolarité.

3. On calcule le taux de réponse en divisant le nombre de répondants par la taille de l'échantillon initial, déduction faite des unités hors champ.


Source : Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003.

5.9.2 Poids jackknife

Pour simplifier l'estimation de la variance, on a supposé que les divers échantillons avaient été choisis indépendamment. On croit que cette hypothèse entraîne une légère surestimation de la variance. L'hypothèse permettait d'utiliser l'estimation jackknife de la variance proposée par Lohr et Rao (1997) pour les bases multiples. Les auteurs proposent de traiter les échantillons des différentes bases comme des échantillons de différentes strates et d'appliquer le jackknife comme pour un échantillonnage stratifié. Pour répondre aux normes internationales, on a inclus dans le fichier de microdonnées de l'enquête trente poids de rééchantillonnage jackknife qui recourent les strates.

On a mis au point ces trente poids de rééchantillonnage jackknife afin de déterminer les erreurs-types des estimations d'enquête. Les trente poids de rééchantillonnage recourent les strates; on a créé trente poids de rééchantillonnage à l'intérieur de chaque province, par type d'échantillon et par strate.

Dans les strates urbaines, on a sélectionné systématiquement des ménages au premier degré de la sélection de l'échantillon. On a formé les poids de rééchantillonnage en triant les ménages dans l'ordre qui a servi à sélectionner systématiquement l'échantillon et en les numérotant de manière séquentielle de 1 à 30 pour les ménages, en recommençant à 1 après avoir atteint 30.



Dans les strates rurales, on a sélectionné des UPÉ au premier degré de sélection. Comme on a sélectionné moins de trente UPÉ dans toutes les strates, on a subdivisé les UPÉ pour former les poids de rééchantillonnage. Dans la mesure du possible, on les a fractionnées en un nombre égal de poids. Si ce n'était pas possible, on a subdivisé les UPÉ comptant un nombre élevé de répondants en un plus grand nombre de poids. Par exemple, si on a sélectionné onze UPÉ pour une province et un échantillon donné, on a ensuite fractionné les huit UPÉ comptant le plus grand nombre de répondants en trois poids de rééchantillonnage chacune, et les trois UPÉ restantes en deux poids chacune, pour obtenir au total trente poids de rééchantillonnage.

On a formé les poids de rééchantillonnage à partir de l'échantillon initial de l'EIACA comptant plus de 40 000 unités. On a calculé des poids jackknife initiaux en fonction des poids déterminés par le plan d'échantillonnage. On a répété l'ensemble du processus de pondération pour chacun des trente poids jackknife, y compris les redressements par pondération de la non-réponse, l'intégration des poids et le calibrage.

Références

- Lohr, S. and Rao, J.N.K. (1997). Jackknife Variance Estimation in Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 552-557.
- Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 203-206.
- Organisation for Economic Co-operation and Development, OECD (2000). *Literacy in the Information Age: Final Report on the International Adult Literacy Survey*. Ottawa: Statistics Canada.
- OECD and Statistics Canada (2000), *Literacy in the information age: Final results of the International Adult Literacy Survey*. Paris and Ottawa: Authors
- Särndal, C.-E., Swenson, B., and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Special Surveys Division (1999). *International Adult Literacy and Skills Survey: National Planning Report*. Ottawa: Statistics Canada.

6.0 Méthodes d'enquête et traitement des données

6.1 Introduction

À bien des égards, les méthodes de l'EIACA sont inspirées des lignes directrices internationales concernant l'administration de l'ELCA. Le caractère normalisé des instruments, de l'échantillonnage, de la collecte et des méthodes de traitement (y compris les codes uniformisés pour la profession, l'industrie et l'éducation) contribue énormément à faire de l'EIACA une étude comparative internationale. Dans la présente section, nous décrivons ces procédés ainsi que les écarts du volet canadien de l'enquête par rapport au protocole. Nous abordons également certains aspects du traitement postcollecte des données qui a mené à la création du fichier de microdonnées à grande diffusion.

L'EIACA a permis de recueillir des renseignements descriptifs et des données sur les capacités des répondants échantillonnés au moyen d'un questionnaire de base et d'une série de livrets d'évaluation contenant des tâches de compréhension de textes suivis et de textes schématiques, de numératie et de résolution de problèmes. Les répondants ont passé environ 30 minutes à répondre à un ensemble commun de questions de base concernant leurs caractéristiques démographiques, leur scolarité, leur expérience sur le marché du travail et leurs activités en matière de littératie. Les réponses à ces questions de base ont permis de résumer les résultats de l'enquête en utilisant un éventail de variables descriptives et d'accroître l'exactitude des estimations des capacités de diverses sous-populations. Les renseignements de base ont été recueillis par des intervieweurs dûment formés.

Après avoir répondu aux questions de base, les répondants ont passé le reste du temps à remplir un livret de tâches de littératie visant à mesurer leurs capacités en compréhension de textes suivis et de textes schématiques, en numératie et en résolution de problèmes. La plupart de ces tâches comportaient des questions ouvertes, auxquelles les répondants devaient fournir une réponse écrite.

Pour obtenir une couverture suffisante de chacun des quatre aspects de la littératie, il fallait évaluer un nombre assez important de tâches. Par contre, il fallait aussi maintenir le fardeau de chaque répondant dans des limites de temps acceptables. Pour respecter ces deux exigences contradictoires, c'est-à-dire réduire le fardeau des répondants sans sacrifier une représentation suffisante des quatre aspects de la littératie, on a demandé à chaque répondant de n'exécuter qu'une partie de l'ensemble de tâches, en utilisant une variante de l'échantillonnage matriciel.

6.2 Modèles de méthodes, de manuels et d'instruments

Chaque pays participant à l'EIACA a reçu un ensemble de manuels d'administration et d'instruments d'enquête à utiliser comme modèle. Les pays étaient libres d'adapter ces modèles à leur système national de collecte de données, mais en conservant un certain nombre

de caractéristiques essentielles. Premièrement, les répondants devaient remplir les livrets de tâches de base et de tâches principales seuls, chez eux, sans l'aide d'une autre personne ou d'une calculatrice. Deuxièmement, les répondants ne devaient pas être rémunérés pour leur participation. Troisièmement, malgré l'interdiction d'offrir des incitatifs pécuniaires, les intervieweurs disposaient de méthodes visant à maximiser le nombre de questionnaires de base remplis et devaient utiliser les mêmes normes de codage pour traiter les cas de non-réponse. Cette dernière exigence est essentielle. Le fait de ne pas remplir les livrets de tâches de base et de tâche principales étant corrélé avec la compétence, on a besoin de renseignements de base sur les non-répondants afin d'imputer les données cognitives concernant ces personnes.

6.2.1 Questions de base

Le modèle de questionnaire de base remis à tous les pays participants comportait deux ensembles de questions : les questions obligatoires, que tous les pays devaient poser, et les questions facultatives, qui étaient recommandées mais non exigées. Les pays n'étaient pas tenus de traduire littéralement les questions obligatoires, mais on leur a demandé de respecter l'objectif conceptuel de chaque question en l'adaptant à leur enquête. Les pays étaient libres d'ajouter des questions à leur questionnaire de base si le fardeau supplémentaire des répondants ne risquait pas de réduire les taux de réponse.

Au Canada, on a posé toutes les questions obligatoires et facultatives, ainsi qu'une série de questions particulières concernant principalement la langue maternelle et le statut d'immigrant ou d'Autochtone.

La **section A** contient des variables liées à la scolarité des répondants.

La **section B** contient des variables liées aux caractéristiques linguistiques.

La **section C** contient des variables liées aux caractéristiques des parents.

La **section D** contient des variables liées à l'expérience de travail des répondants.

La **section E** contient des variables liées à l'utilisation de différentes compétences au travail.

La **section F** contient des variables liées à l'éducation et à la formation des adultes.

La **section G** contient des variables liées aux habitudes en matière de compétences et de santé.

La **section H** contient des variables liées à l'utilisation d'Internet.

La **section K** contient des variables liées au revenu personnel et à celui du ménage.

Lorsque les réponses à ces questions ne compromettent pas la confidentialité des répondants, le FMGD de l'EIACA comprend le plus grand nombre possible de données recueillies. Dans le chapitre 8, nous aborderons en détail la question de la confidentialité.

6.2.2 Items de tâche

À l'instar de l'ELCA qui l'a précédée, l'IALS repose sur le principe selon lequel la difficulté de diverses tâches de littératie est déterminée par certains facteurs, qui sont stables d'une langue à l'autre et d'une culture à l'autre. On a donc remis à tous les pays participants des fichiers graphiques contenant l'ensemble des items psychométriques et on les a invités à traduire chaque item de l'anglais à leur propre langue sans modifier la présentation graphique ni les caractéristiques des tâches. Dans bien des cas, on a traduit l'item initial en fonction de ce modèle anglais pour offrir à tous le même point de départ. Cette uniformité des documents de base a permis de réduire au minimum l'incidence d'erreurs de traduction et d'adaptation.

Le processus d'adaptation des items était soumis à certaines règles. Par exemple, certains items exigeaient du répondant qu'il accomplisse une tâche qui était facilitée par l'utilisation de mots-clés. Dans certains cas, les mots-clés étaient identiques dans la question et dans le corps de l'item; dans d'autres, le mot-clé était semblable, mais pas exactement le même; dans d'autres encore, le mot-clé était un synonyme du mot utilisé dans le corps de l'item. Dans un autre cas, on demandait au répondant de choisir, dans le corps de l'item, entre plusieurs mots-clés dont un seul était correct. Les pays devaient traduire les items en préservant ces associations conceptuelles.

Au besoin, certaines conventions utilisées dans les items – unités monétaires, écriture des dates, notation décimale, etc. – ont été adaptées pour chaque pays.

Pour s'assurer que le processus d'adaptation ne compromettait pas l'intégrité psychométrique des items, on a soigneusement examiné les livrets de chaque pays pour y déceler des erreurs d'adaptation.

6.2.3 Codage normalisé des non-réponses

Il était essentiel que les pays participant à l'enquête gèrent les cas de non-réponse de manière uniforme afin de limiter, dans les estimations, le biais de non-réponse.

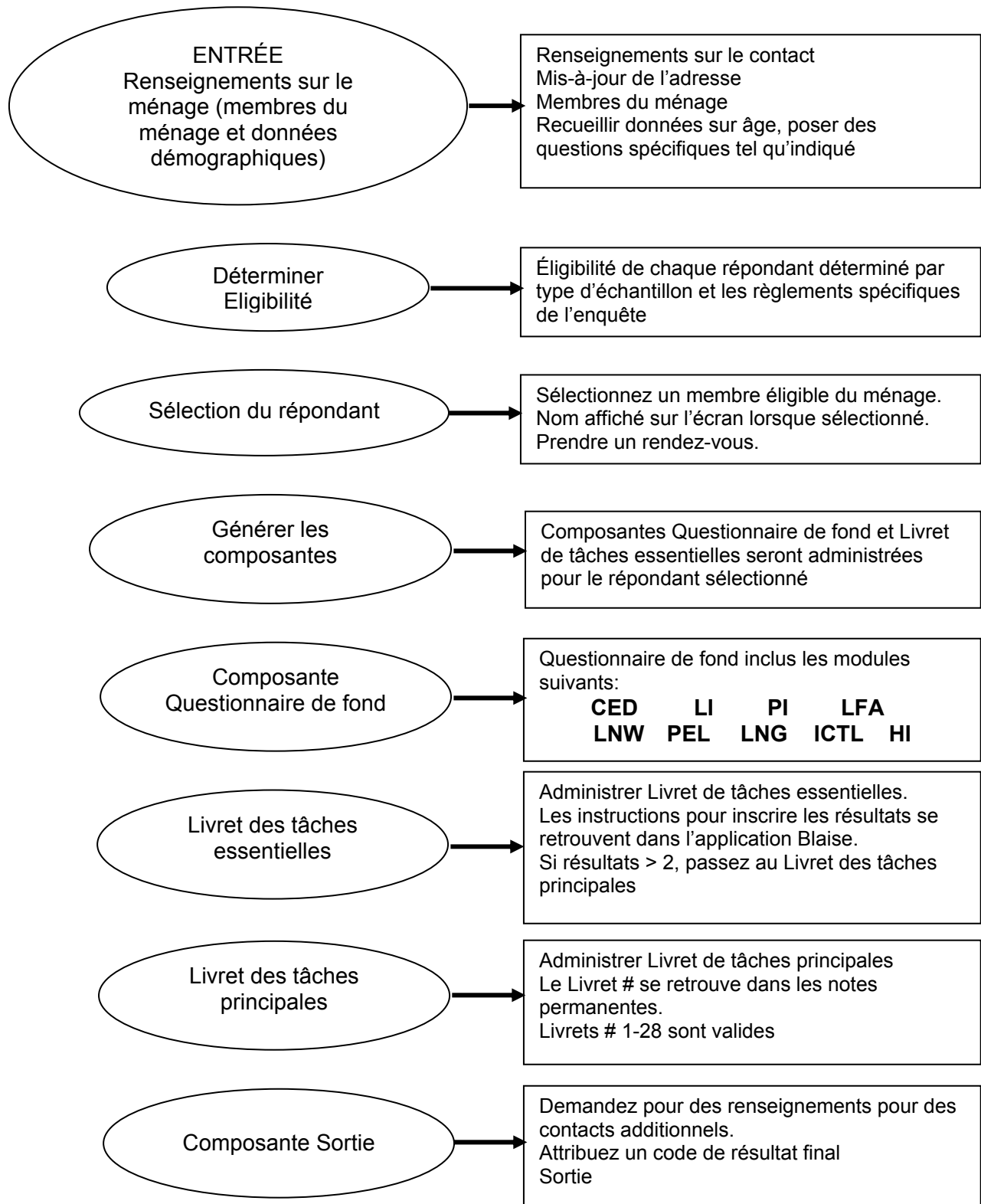
Pour les besoins de l'EIACA, le répondant devait remplir le questionnaire de base, réussir les tâches de littératie de base et tenter d'effectuer au moins cinq tâches par échelle de littératie pour que les chercheurs puissent évaluer directement ses capacités de littératie. On a imputé les données sur les capacités des personnes qui échouaient ou qui refusaient d'effectuer les tâches de base et de celles qui réussissaient les tâches de base sans tenter d'effectuer au moins cinq tâches par échelle de littératie. Comme le modèle utilisé pour imputer les estimations du niveau de littératie des non-répondants repose sur un ensemble complet de réponses aux questions de base, les pays participants ont reçu pour directive de faire remplir au moins le questionnaire de base par les personnes échantillonnées. On leur a également remis une classification détaillée des non-réponses pour les besoins de l'enquête.

Il appartenait à chaque pays de recruter ses intervieweurs. Ainsi, le nombre d'intervieweurs, leur taux de rémunération et la durée de la période d'enquête variaient d'un pays à l'autre, selon les normes et les budgets. Chaque pays a reçu un livret portant sur la formation des intervieweurs.

Au Canada, les données de l'EIACA ont été recueillies par des intervieweurs expérimentés de Statistique Canada au moyen de la technologie de l'interview personnelle assistée par ordinateur.

Le diagramme ci-dessous représente le plan de l'Enquête internationale sur l'alphabétisation et les compétences des adultes.

Enquête internationale sur l'alphabétisation et les compétences des adultes Organigramme



6.3 Notation

On a évalué les compétences en littératie des répondants d'après leur rendement à l'égard des tâches cognitives exécutées au moment de l'enquête. Contrairement aux questions à choix multiple, qu'on utilise habituellement dans les enquêtes à grande échelle et qui proposent un nombre fixe de choix de réponse, les questions ouvertes comme celles qu'on a posées dans le cadre de l'EIACA donnent lieu à un vaste éventail de réponses. Comme les données brutes sont rarement utiles à elles seules, on doit regrouper les réponses, d'une manière ou d'une autre, afin de résumer les résultats relatifs au rendement. À mesure qu'elles étaient notées, les réponses aux questions ouvertes de l'EIACA étaient classées dans trois catégories : correctes, incorrectes ou manquantes.

Les modèles employés pour évaluer les capacités des répondants et la difficulté des tâches reposent sur l'hypothèse selon laquelle les critères de notation établis pour les besoins de l'évaluation ont été appliqués uniformément à l'intérieur des pays et d'un pays à l'autre. Pour vérifier cette hypothèse, on a pris plusieurs mesures, notamment la deuxième notation à l'intérieur d'un pays et la deuxième notation entre les pays, qui sont décrites dans les sections qui suivent.

6.3.1 Deuxième notation à l'intérieur d'un pays

Pour contrôler l'exactitude de la notation, on a établi une méthode fondée sur un taux d'échantillonnage variable. Au début de la notation, presque toutes les réponses ont fait l'objet d'une deuxième notation visant à repérer les correcteurs problématiques et à déceler les réponses uniques ou difficiles qui n'étaient pas prévues dans le manuel de notation. Après avoir atteint un niveau d'exactitude satisfaisant, on a ramené le niveau de la deuxième notation à celui d'une mise à jour pour contrôler l'exactitude de tous les correcteurs. On a calculé des concordances moyennes pour tous les items. Pour s'assurer que la première et la deuxième notations étaient vraiment indépendantes, il a fallu prendre certaines précautions. Par exemple, les correcteurs devaient être des personnes différentes et le deuxième correcteur ne devait pas connaître les notes attribuées par le premier.

On s'attend à ce que les correcteurs ayant reçu une formation identique à l'intérieur d'un pays soient plus cohérents entre eux que par rapport aux correcteurs d'autres pays. C'est ce qu'on a constaté lors de l'EIAA de 1994. Dans la plupart des cas, la fiabilité de la deuxième notation était supérieure à 97 %. Il importe de signaler que les résultats respectaient largement les tolérances statistiques établies pour les besoins de l'enquête et qu'ils étaient nettement supérieurs à ceux atteints dans d'autres études à grande échelle utilisant des questions ouvertes.

Le tableau 6.1 montre la fiabilité de la notation à l'intérieur d'un pays dans le cadre de l'EIACA de 2003 au Canada.

Tableau 6.1 Fiabilité de la notation à l'intérieur d'un pays par domaine, en pourcentage

	Domaine			Total %
	Textes suivis et textes schématiques	Numératie	Résolution de problèmes	
Notation de tests en français par des anglophones	95	95	92	95
Notation de tests en anglais par des francophones	95	97	94	95

Source : Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Comme la deuxième notation à l'intérieur d'un pays visait à améliorer la qualité des données, les notes n'ont pas été mises à jour dans la base de données. En d'autres termes, les données sur la concordance présentées ici font état de la concordance minimale atteinte à l'étape de la notation. Après avoir calculé la fiabilité de la notation à l'intérieur d'un pays, on a constaté que quelques correcteurs n'étaient pas fiables. Ces correcteurs ont reçu une formation supplémentaire ou ont été remerciés. Avdenant un écart entre la première notation et la deuxième, on a remplacé la première note par la bonne note si l'inexactitude était attribuable à une erreur systématique de la part du premier correcteur. Dans certains cas, le manuel de notation s'est avéré ambigu. On a alors révisé le manuel de notation et modifié la première note en fonction de la révision, mais sans modifier la deuxième note. On n'a jamais remplacé une deuxième note, même si elle se révélait erronée par la suite.

En somme, la première note tient compte des modifications et des corrections résultant des leçons tirées de l'analyse de la deuxième notation à l'intérieur d'un pays. La première note est donc plus exacte et cohérente que la deuxième, qui peut rester erronée et qui, par conséquent, sous-estime quelque peu la fiabilité de la deuxième notation. Cette sous-estimation doit toutefois être très faible puisque, dans la plupart des cas, la fiabilité est supérieure à 97 %. Ces valeurs indiquent que tous les pays participants ont atteint un très haut degré d'uniformité dans la notation.

6.3.2 Deuxième notation entre les pays

Même après avoir vérifié que tous les correcteurs pratiquaient une notation uniforme, remédié aux ambiguïtés des manuels de notation et corrigé les erreurs de notation systématiques, il fallait encore examiner la comparabilité des notes d'un pays à l'autre. Une notation exacte et uniforme à l'intérieur d'un pays ne suppose pas nécessairement que tous les pays appliquent de la même manière les directives des manuels de notation. Un biais lié à la notation peut survenir si un pays note une certaine réponse différemment des autres pays. La deuxième notation entre les pays, décrite dans la présente section, visait à assurer la comparabilité de la notation d'un pays à l'autre.

Comme nous l'avons mentionné plus haut, chaque pays a noté séparément les réponses aux items d'évaluation de l'enquête. Pour déterminer la fiabilité de la notation entre les pays à l'égard de chaque item, deux groupes distincts ont noté les réponses d'un sous-ensemble de répondants. Habituellement, ces groupes de notation étaient de pays différents. Par exemple, un échantillon de livrets de tests a été noté par deux groupes qui avaient noté les livrets en anglais du Canada et ceux des États-Unis. Statistique Canada a calculé la fiabilité de la deuxième notation entre les pays, puis ETS a évalué les résultats. À la suite de cette

évaluation, on a demandé à chaque pays d'apporter quelques modifications mineures à la méthode de notation. Dans certains cas, on a constaté que les directives ambiguës dans le manuel de notation entraînaient des erreurs d'interprétation et, par conséquent, une baisse de la fiabilité.

En s'inspirant de la fiabilité de la notation entre les pays, les chercheurs peuvent repérer les items mal conçus, les critères de notation ambiguës, les erreurs de traduction et d'impression d'items ou de critères de notation, les inexactitudes attribuables aux correcteurs et, surtout, les cas où un pays pratiquait systématiquement une notation différente de celle d'un autre pays. Dans ce dernier cas, il arrive que les correcteurs d'un pays notent uniformément une certaine réponse comme étant correcte, alors que ceux d'un autre pays notent la même réponse comme étant incorrecte. Avant de procéder à l'échelonnage selon la méthode TRI, il faut éliminer cette asymétrie des notes. ETS et Statistique Canada ont repéré les items asymétriques, et le pays dans lequel le problème de notation était survenu a enquêté sur les causes plausibles de ce biais systématique dans les notes. Lorsqu'on repérait une erreur systématique dans un pays donné, on corrigeait les notes initiales attribuées à l'item en question pour la totalité de l'échantillon.

Le tableau 6.2 résume la fiabilité de la deuxième notation entre les pays pour le volet canadien de l'ELCA, avant corrections.

Tableau 6.2 Fiabilité de la deuxième notation entre les pays

Fiabilité de la notation par domaine, en pourcentage

Appariement des pays (deuxième notation – première notation)	Domaine psychométrique			
	Textes suivis et textes schématiques (%)	Numératie (%)	Résolution de problèmes (%)	Total (%)
Canada anglais – Canada français	95	95	92	95
Canada français – Canada anglais	95	97	94	95
Norvège – Canada	91	93	91	92
Canada – Etats-Unis	94	97	...	95
États-Unis – Canada	95	97	...	95
États-Unis – Bermudes	91	94	...	90
Bermudes – Etats-Unis	93	95	...	93
Canada français – Suisse	95	98	97	96
Suisse – Canada français	94	96	94	95
Suisse – Italie	96	98	96	96
Italie – Suisse	93	97	93	94
Canada – Bermudes	83	83
Canada – Nuevo León (Mexique)	91	95 ¹	...	92

... n'ayant pas lieu de figure

1. Textes au contenu quantitatif

6.4 Saisie, traitement et codage des données

Comme condition à sa participation à l'ELCA, on a demandé à l'équipe de l'EIACA de saisir et de traiter les fichiers en utilisant des méthodes qui assuraient une cohérence logique et des niveaux acceptables d'erreur dans la saisie des données.

Le questionnaire de base canadien a été rempli par ordinateur. Comme il fallait noter les livrets de tâches et saisir les données correspondantes, dans le cadre des lignes directrices de l'ELCA, on a demandé aux pays d'effectuer une vérification complète des notes saisies (en entrant chaque enregistrement deux fois) afin de réduire les taux d'erreur au minimum. L'exactitude de la notation des tests étant essentielle pour assurer un haut degré de qualité des données, il fallait exercer un contrôle intégral par double saisie.

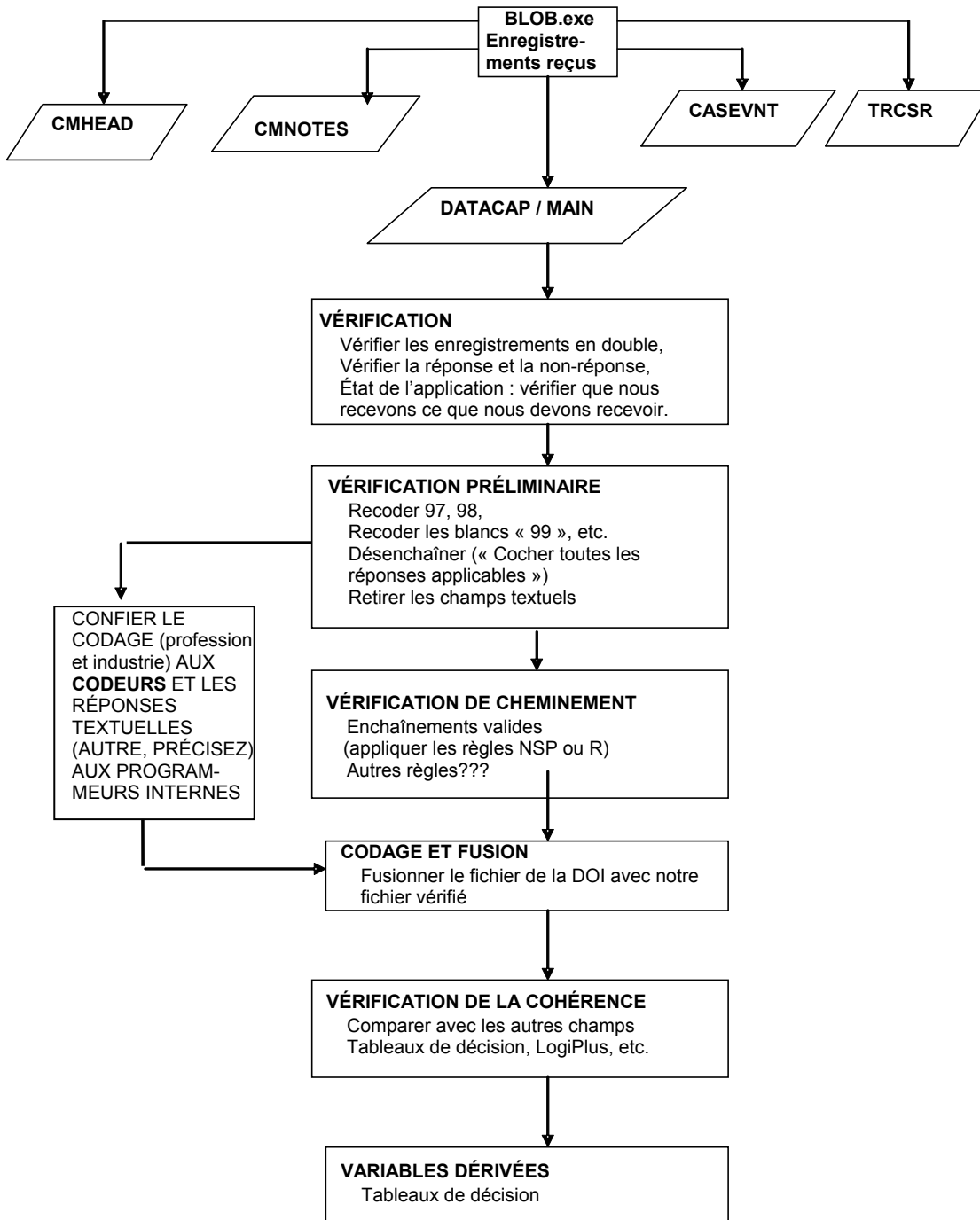
Chaque pays devait également coder l'industrie, la profession et le niveau de scolarité en utilisant les codes types internationaux établis à cette fin (Classification internationale type par industrie, CITI; Classification internationale type des professions, CITP; Classification internationale type de l'éducation, CITE). On a également fourni des codes pour les questions ouvertes, et les pays ont reçu des directives précises relativement au codage de ces questions de façon à maintenir les erreurs de codage à des niveaux acceptables.

On a aussi codé le fichier canadien selon les classifications nationales de l'éducation, des professions (CTP, CNP et CNP-S) et de l'industrie (CTI et SCIAN). À cause du chevauchement de certains codes d'industrie et de profession, il n'était pas possible d'inclure ces codes canadiens dans le fichier de microdonnées à grande diffusion, mais des totalisations spéciales utilisant ces classifications canadiennes sont disponibles auprès de Statistique Canada.

Pour améliorer la comparabilité de l'analyse des données, on a demandé à chaque pays participant à l'EIACA de schématiser son ensemble de données national sous forme de disposition d'enregistrement très structurée et uniformisée. Outre qu'elle précisait la position, le format et la longueur de chaque champ, la disposition d'enregistrement internationale comprenait une description de chaque variable et indiquait les catégories et les codes à fournir pour cette variable.

L'organigramme technique ci-dessous résume les étapes du traitement du fichier de données canadien envoyé par les intervieweurs.

ORGANIGRAMME DE TRAITEMENT – EIACA



Comme le montre l'organigramme ci-dessus, on peut répartir le processus en neuf étapes distinctes :

1. Traitement postcollecte
2. Épuration
3. Vérification préliminaire
4. Vérification de cheminement
5. Vérification de la cohérence
6. Création et imputation des variables dérivées
7. Création du fichier principal
8. Création du fichier national de microdonnées à grande diffusion
9. Création du fichier international de microdonnées à grande diffusion

6.4.1 Traitement postcollecte (étape 1)

L'EIACA a utilisé trois grandes méthodes de collecte des données : l'interview papier et crayon (IPC), l'interview assistée par ordinateur (IAO) et l'interview téléphonique assistée par ordinateur (ITAO). Les étapes postcollecte de l'ITAO, de l'IAO et de l'IPC sont décrites dans la présente section.

À l'étape de la collecte, lorsqu'on disposait du numéro de téléphone correspondant à une adresse donnée, on a recueilli des renseignements de présélection sur le répondant et les autres membres du ménage (caractéristiques démographiques : âge, sexe, statut d'immigrant, langue maternelle, etc., confirmation de l'adresse et admissibilité des membres du ménage). On a choisi le répondant au hasard parmi les membres admissibles du ménage. L'intervieweur a recueilli les renseignements nécessaires au moyen de l'ITAO, selon le principe des rendez-vous ferme et flexible, puis a demandé au répondant de lui accorder une visite sur place dans la langue de son choix. Une partie de cette interview (les renseignements de base) est effectuée au moyen de l'application IAO et le reste (les tâches de base et les tâches principales), au moyen de l'IPC. Si l'on n'avait pu communiquer avec le répondant ou si l'on ne disposait pas du numéro de téléphone correspondant à une adresse donnée, il fallait effectuer une visite sur place au ménage, selon les mêmes méthodes, mais cette fois en remplissant le questionnaire de présélection en personne plutôt que par téléphone.

Exécution du processus BLOB

Le processus BLOB consiste simplement à démonter les données en diverses structures de fichier en fonction des dispositions d'enregistrement définies par les programmeurs qui ont mis au point l'application de l'interview assistée par ordinateur (IAO). On crée ainsi une série de fichiers comme le fichier En-tête, le questionnaire de base, le fichier « casevnt » et d'autres fichiers connexes pour exécuter les opérations nécessaires au traitement.

Les dispositions d'enregistrement fournies par les programmeurs de l'IAO ont servi à créer les structures et les dispositions des fichiers de la base de données à utiliser pour poursuivre le traitement. Une unité de traitement s'est assurée que les dispositions créées par les programmeurs étaient correctes, que les noms des variables étaient significatifs et uniques pour renvoyer aux données dans les positions correctement attribuées et qu'on leur avait attribué la bonne longueur et le bon type de variable (numérique ou sous forme de caractères).

6.4.2 Épuration des données (étape 2)

L'objet de l'épuration consiste à cerner les cas qui nécessitent une certaine intervention et ceux qui ne nécessitent aucune vérification parce qu'il s'agit de cas de non-réponse. Il s'agit d'apporter des corrections au niveau de l'enregistrement, par exemple en cernant et en résolvant les enregistrements en double, en cernant et en retraçant les enregistrements manquants, en révisant l'attribution des codes d'état et en cernant les cas de non-réponse.

Enregistrements en double

Pour les besoins de l'EIACA, un fichier ne peut contenir qu'un seul enregistrement par unité d'échantillon. On a donc résolu ou éliminé tous les enregistrements en double.

On a trouvé deux types d'enregistrements en double : les enregistrements en double complets et les enregistrements présentant un numéro d'identification en double.

- ❖ On entend par enregistrements en double complets deux enregistrements identiques comportant exactement les mêmes renseignements dans tous les champs. On en trouve surtout dans le contexte de l'IAO lorsque, par exemple, un enregistrement est transmis et qu'à cause d'une défectuosité du système il est transmis de nouveau, d'où la présence de deux enregistrements pour un même cas. Dans le contexte de l'IPC, il est possible que les données d'un enregistrement soient saisies deux fois, surtout si le système de saisie des données ne comporte pas un processus intégré de validation des numéros d'identification.
- ❖ On entend par enregistrements présentant un numéro d'identification en double deux enregistrements portant le même numéro d'identification, mais contenant des données différentes. Dans le contexte de l'IAO, on en trouve habituellement lorsqu'un cas est transmis avant d'être finalisé, puis après avoir été finalisé. Dans le contexte de l'IAO, on peut avoir saisi un numéro d'identification incorrect, d'où la présence d'un double. Ces doubles doivent être examinés soigneusement par un membre de l'équipe de traitement, qui détermine s'il faut supprimer l'enregistrement ou modifier le numéro d'identification.

Comparaison du questionnaire de base avec le fichier échantillon

On a comparé le numéro d'identification saisi dans le fichier de données d'enquête à celui du fichier échantillon envoyé pour la collecte. Les enregistrements reçus quotidiennement devaient constituer un sous-ensemble du fichier échantillon. Dans le cas contraire, on a mené une enquête et pris les mesures correctives nécessaires. Les rapports ont été contrôlés quotidiennement pendant toute la période de collecte.

Vérification des codes d'état finaux (BQStatus)

On a vérifié les codes d'état du questionnaire de base (BQ) en fonction des renseignements saisis.

Le tableau ci-dessous donne une description des divers codes d'état :

Réponse

Description	Code de résultat
Rempli au complet	70
Rempli en partie	71

Non-réponse

Description	Code de résultat
Refus (Le répondant a refusé de remplir le questionnaire de base ou une autre personne a refusé à sa place.)	80
Problème de langue (L' interview ne peut pas avoir lieu parce que le répondant ne peut pas communiquer dans la langue requise de l'interview et que l'intervieweur ne peut pas trouver une personne qui parle la langue du répondant.)	22
Problème de lecture et d'écriture ou déficience physique (Il est impossible de remplir le questionnaire de base à cause de l'état mental ou physique du répondant.)	92
Décès, maladie ou circonstances exceptionnelles (Il est impossible de remplir le questionnaire de base pour cause de maladie, d'incendie, d'inondation, etc.)	90
Non admissible (Aucun membre du ménage n'est âgé de 16 ans et plus ou les membres ne répondent pas aux critères d'échantillonnage nécessaires pour cibler certaines sous-populations.)	56 /76, 40
Non-réponse liée au logement (Pas à la maison, logement en construction, etc.)	51, 52, 53, 55, 26
Incapable de communiquer avec le répondant (Pas d'interlocuteur, personne à la maison ou pas de réponse, répondant absent pendant la durée de l'enquête, période de collecte terminée.)	10, 11, 20, 98
Conditions météorologiques (Il est impossible de mener l'interview à cause des conditions météorologiques.)	18
Double (Le logement a été sélectionné deux fois pour l'enquête.)	63

Repérage des enregistrements manquants

On a repéré les enregistrements manquants dès le début du processus afin d'en faciliter la résolution.

Méthodes

Il existe bien des moyens pour prévenir la perte de documents pendant le processus d'enquête. L'EIACA utilise notamment les suivants :

- ❖ Liste de contrôle maîtresse des tâches (liste de tous les intervieweurs et de leurs affectations respectives).
- ❖ Bordereau de transmission de l'intervieweur – Les intervieweurs devaient retourner périodiquement les documents d'enquête en joignant à chaque « envoi » ou « ensemble de documents » un bordereau de transmission comprenant des renseignements semblables à ceux qu'on trouve sur la Liste de contrôle maîtresse des tâches.
- ❖ Lorsqu'on a reçu les documents à Ottawa, on a comparé le contenu de l'ensemble au bordereau de transmission de l'intervieweur. On a immédiatement signalé à l'intervieweur les documents manquants pour fins de résolution.
- ❖ Comparaison du fichier échantillon avec le fichier de données d'enquête – Après avoir reçu tous les documents ou enregistrements, saisi les données et résolu les enregistrements en double, on peut comparer le fichier échantillon principal créé par le méthodologiste chargé de l'enquête au fichier de données d'enquête (établi au moment de la saisie des données) pour repérer les enregistrements ou les composantes manquants.

Création des fichiers de réponse et de non-réponse

Après avoir saisi les données de tous les enregistrements, franchi toutes les étapes définies dans la présente section et résolu tous les problèmes, on a préparé les fichiers de réponse et de non-réponse en vue du traitement.

Fichier de données de réponse

À partir du fichier de données d'enquête, on retient tous les enregistrements qui ont certaines données et l'on crée un nouveau fichier qu'on appelle désormais le « fichier de données de réponse ». On utilise ce fichier à toutes les étapes de vérification préliminaire et de vérification décrites dans le présent document.

Fichier de données de non-réponse

À partir du fichier de données d'enquête, on retient tous les enregistrements qui n'ont pas de données. On s'assure que le fichier ne contient pas de données valides. S'il en contient, on prend les mesures pertinentes pour fusionner les enregistrements valides au fichier de réponse.

On n'a comparé le fichier échantillon principal au fichier de données d'enquête et créé les fichiers de données de réponse et de non-réponse qu'après avoir terminé la collecte des données et saisi les données de tous les enregistrements. Toutefois, on a effectué quotidiennement certaines comparaisons en localisant les questionnaires dès qu'ils étaient retournés à l'équipe de traitement. Cette activité a non seulement amélioré les chances de retracer les documents manquants, mais a également permis d'évaluer la qualité et l'exactitude des données en comparant les fichiers de réponse et de non-réponse au fichier échantillon.

6.4.3 Vérification préliminaire (étape 3)

La vérification préliminaire a lieu après l'étape de l'épuration. Elle consiste à modifier les données au niveau de la variable individuelle. On peut supprimer, recoder, redimensionner les variables ou les laisser telles quelles. Les champs « Cocher toutes les réponses qui

s'appliquent » sont « désenchaînés ». Les champs textuels sont retirés des fichiers principaux et reportés dans un fichier distinct.

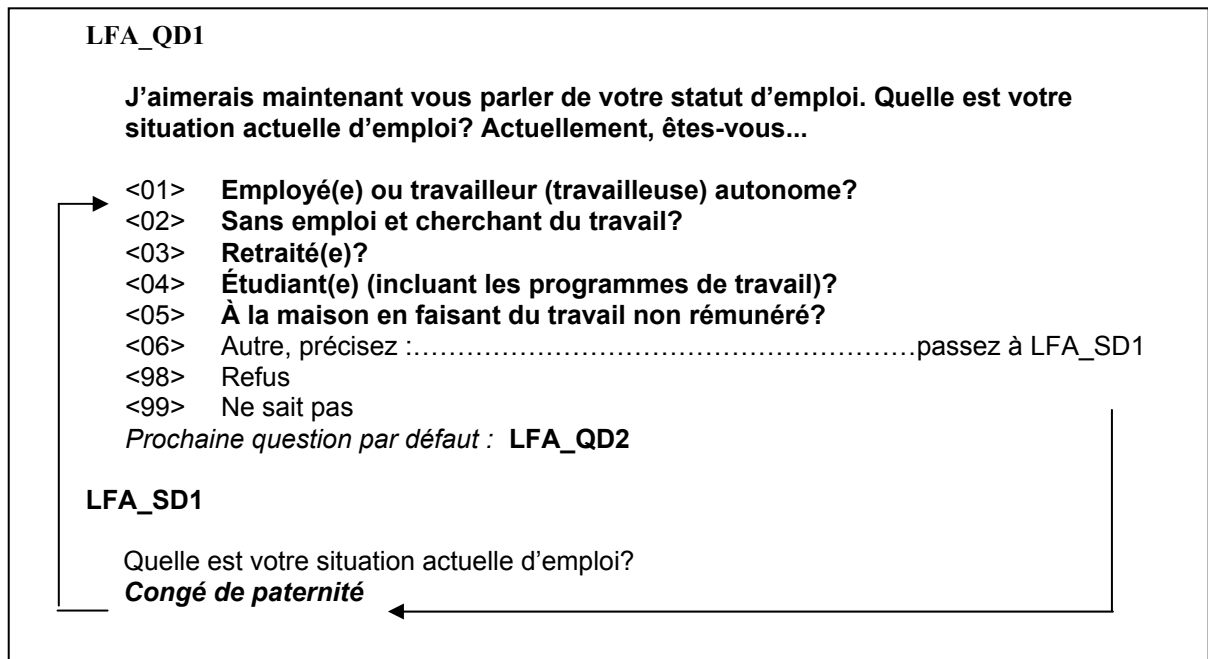
Codage des champs « Autre, précisez »

Les questions comportant une liste de catégories de réponse contiennent souvent « Autre » comme dernière catégorie. Le codage de ces champs consistait à vérifier le contenu et à déterminer si la réponse aurait dû figurer dans une autre catégorie de réponse ou s'il fallait la laisser dans la catégorie « Autre ».

À l'étape de la vérification préliminaire, on a donc saisi les réponses écrites en toutes lettres et révisé les valeurs correspondantes dans le cadre d'un processus de codage.

Si on les repère très tôt à l'étape de la collecte des données, on peut cerner les erreurs systématiques et recommander aux intervieweurs d'examiner de près les catégories de réponse pour mieux reconnaître les différentes réponses qui peuvent être codées selon des catégories de réponse données. Ce type de rétroaction au cours du processus permet de réduire au minimum le traitement postcollecte nécessaire et, naturellement, accroît la qualité des renseignements recueillis.

Dans certains cas, les renseignements contenus dans le champ « Autre, précisez » peuvent s'avérer très utiles, comme dans la question D1 du questionnaire de base. L'exemple ci-dessous illustre un cas où le renseignement entré dans le champ « Autre, précisez » serait codé selon l'une des catégories existantes, en l'occurrence « <01> », car le répondant travaille, mais il est en congé de paternité.



Pour les besoins de l'EIACA, il ne fallait pas ajouter de codes supplémentaires aux variables internationales. Les questions ci-dessous du questionnaire de base contiennent un champ « Autre, précisez ». Les réponses respectives figurent dans les questions identifiées par la flèche.

- ❖ QA0B – À quel(s) groupe(s) ethnique(s) ou culturel(s) vos ancêtres appartenaient-ils? → SA0B
- ❖ QA1D – Dans quel pays êtes-vous né? → SA1D
- ❖ QA1E – Statut d'immigrant actuel → SA1E
- ❖ QA6D – Dans quel pays avez-vous atteint le niveau de scolarité le plus élevé (à l'extérieur du Canada) → SA6D
- ❖ QB1 – Première langue apprise pendant l'enfance → SB1
- ❖ QB2 – Langue parlée le plus souvent à la maison → SB2
- ❖ QD1 – Statut/Situation d'emploi → SD1
- ❖ QF30H – Raisons de ne pas suivre de cours → SF30F
- ❖ QH3H – Autres raisons d'utiliser un ordinateur → SH3H
- ❖ QH5N – Autres raisons d'utiliser Internet → SH5N

Désenchaînement des champs « Cocher toutes les réponses qui s'appliquent »

Les questions qui permettent des réponses multiples pour chaque répondant doivent être réparties en variables binaires simples pour chaque choix admissible.

Dans l'exemple ci-dessous, la question QF14 est répartie en six variables, soit une pour chaque catégorie de réponse admissible.

Exemple de désenchaînement de la question QF14 – Qui a contribué à assumer les frais directs...

Question	Valeur initiale de la réponse	Nouvelle étiquette de valeur	Nouveau nom de la variable
QF14	<01> Vous-même ou votre famille		
		F14A	1 (Oui)
	<02> Un employeur		2 (Non)
			7 (Ne sait pas)
	<03> Un organisme gouvernemental		8 (Refus)
		F14B	1 (Oui)
	<04> Un syndicat ou une association professionnelle		2 (Non)
			7 (Ne sait pas)
	<05> Quelqu'un d'autre		8 (Refus)
		F14C	1 (Oui)
	<06> AUCUNS FRAIS DIRECTS		2 (Non)
			7 (Ne sait pas)
	<98> Refus		8 (Refus)
		F14D	1 (Oui)
	<99> Ne sait pas		2 (Non)
			7 (Ne sait pas)
			8 (Refus)
		F14E	1 (Oui)
			2 (Non)
			7 (Ne sait pas)
			8 (Refus)
		F14F	1 (Oui)
			2 (Non)
			7 (Ne sait pas)
			8 (Refus)

Si la réponse à la question F14 était 99 (Ne sait pas), 98 (Refus) ou un blanc, le programme devrait coder toutes les nouvelles variables (F14A à F14F), respectivement, comme suit : 97, 98 ou 99.

Autres mesures de vérification préliminaire

À cette étape du traitement, chaque variable soumise à la vérification préliminaire a fait l'objet d'une ou de plusieurs des mesures suivantes :

TERME	DÉFINITION
Supprimer	Retirer la variable du fichier.
Recoder	Coder les champs « Ne sait pas » et « Refus » selon les codes pertinents (7 et 8) et rendre conforme aux normes internationales tout autre code différent au moment de la collecte.
Laisser	Aucune modification de la variable n'est nécessaire.
Redimensionner	Il peut être nécessaire d'augmenter ou de réduire la longueur.
Désenchaîner	Les questions « Cocher toutes les réponses qui s'appliquent » sont retournées sous forme comprimée; il faut les convertir en valeurs « oui/non », une par réponse possible
Inscrire	Inscrire la réponse dans le fichier des réponses écrites en toutes lettres et la retirer du fichier de données

Voici un exemple de spécification de vérification préliminaire utilisant diverses sections du questionnaire de base :

InName	Insize	InType	OutName	OutSize	Indicateur de recodage	Étiquettes de valeur	Observation
ID (numéro d'identification)	9	C	ID	9	2		Laisser
...							
BT1	5	N		0	0		Supprimer
B1 (01-14, 98)	4	N	B1A	2	5		Laisser
			B1B	2	5		
B2 (01-14, 98)	2	N	B2	2	2		Laisser
BT2	5	N		0	0		Supprimer
..							
D25	80	C	D25	1	3	1=ooui 2=non 7=Ne sait pas 8=Refus	Inscrire et recoder
D26	80	C	D26	1	3	1=ooui 2=non 7=Ne sait pas 8=Refus	Inscrire et recoder
D27	80	C	D27	1	3	1=ooui 2=non 7=Ne sait pas 8=Refus	Inscrire et recoder
D28	80	C	D28	1	3	1=ooui 2=non 7=Ne sait pas 8=Refus	Inscrire et recoder
..							
..							
ET1	5	N		0	0		Supprimer
E1A	1	N		1	2		Laisser
E1B	1	N		1	2		Laisser
E1C	1	N		1	2		Laisser

6.4.4 Vérifications de cheminement (étape 4)

La vérification de cheminement reproduit l'enchaînement (ordre des questions) prévu par le plan d'enquête. Le questionnaire comprend une série de branchements conditionnels en fonction des réponses fournies antérieurement. Par exemple, si le répondant déclare n'avoir jamais suivi de cours ou de programme d'études, on saute toute la section « Éducation des adultes ». Les champs sautés en raison de l'enchaînement sont convertis en blancs (enchaînements valides), alors que les champs sautés à cause d'une non-réponse sont codés « 9 » (Non déclaré). Pour traiter les données de l'EIACA, tous les pays participants ont utilisé une méthode descendante consistant à suivre exactement l'enchaînement du questionnaire à partir de la première question. Si, par exemple, la réponse à la première question entraîne un

saut de trois questions, les trois questions sautées sont laissées en blanc, quel que soit leur contenu.

Voici un exemple de vérification de cheminement pour les questions QA4 D à QA8; on doit établir le même type de spécifications pour chaque section du questionnaire de base.

Cocher	Variable	Taille	Type	Condition	Passer à	Valeur	Observations
	QA4D – Avez-vous entrepris des études en vue de l'obtention d'un certificat ou d'un diplôme de formation professionnelle/technique, d'études collégiales (ou CÉGEP) ou universitaires que vous n'avez PAS COMPLÉTÉES?	1	N	1			
	QA4D	1	N	2	QA6	EV	Inscrire 6 en AQA4E, 96 en QA4F, 96 en QA5.
	QA4D	1	N	8 (R), 9 (NSP)	QA6	ND	Inscrire 9 en AQA4E, 99 en QA4F, 99 en QA5.
	QA4E – Le diplôme de CÉGEP faisait-il partie d'un programme d'études collégiales générales – formation pré-universitaire (en vue d'obtenir des unités, un diplôme menant à l'université ou un grade d'associé)?	1	N				Aucune mesure à prendre, puisque la question mène à la suivante (QA4F).
	QA4F – Dans le cas d'études à plein temps, quelle est la durée normale requise pour obtenir ce certificat ou diplôme?	2	N	01, 02, 03, 04, 05, 06, 07	QA6	EV	Inscrire 96 en QA5.
	QA4F	2	N	98 (R), 98 (NSP)	QA6	ND	Inscrire 99 en QA5.
	QA5 – Quel âge aviez-vous quand vous avez commencé vos études en vue de l'obtention d'un certificat ou d'un diplôme de formation professionnelle/technique, d'études collégiales (ou CÉGEP) ou universitaires?	2	N	0 - 95	CA6A	EV	Inscrire 96 en QA6.
	QA5	2	N	98 (R), 98 (NSP)	CA6A	ND	Inscrire 99 en QA6.
	QA6 – Quel âge aviez-vous quand vous avez complété votre niveau de scolarité le plus élevé?	2	N				Aucune mesure à prendre, puisque la question mène à la suivante (QA4F).
C	CA6A			Si (QA1 = 2 ou IMM_Q01 = 2) et QA3B > 0 et QA3B	QA6B	EV	
				OU	QA6E	EV	
	OU						
C	CA6A			Si ((QA1 ≠ 2 ou IMM_Q01 ≠ 2) et QA3B > 0 et QA3B <96)	QA6E	EV	Sinon, passer à la question suivante (QA6B).

Cocher	Variable	Taille	Type	Condition	Passer à	Valeur	Observations
	QA6B – Avez-vous atteint ce niveau au Canada?						Aucune mesure à prendre, puisque la question mène à la suivante (QA4C).
	QA6C – Quel est le niveau de scolarité le plus élevé que vous avez complété à l'extérieur du Canada?						Aucune mesure à prendre, puisque la question mène à la suivante (QA4D).
	QA6D – Dans quel pays avez-vous atteint ce niveau de scolarité?						Aucune mesure à prendre, puisque la question mène à la suivante (QA4E).
	QA6E – Quelle était la langue principale d'enseignement utilisée dans vos cours à l'école primaire, à l'exception des cours de langue?						Aucune mesure à prendre, puisque la question mène à la suivante (CA7).
C	CA7			Si QA4B = 1 ou QA4B = 2	QA10	EV	Sinon, passer à la question suivante (QA7A).
	QA7A – Quelle était la principale langue d'enseignement utilisée dans vos cours au premier cycle ou au deuxième cycle du secondaire, à l'exception des cours de langue?						Aucune mesure à prendre, puisque la question mène à la suivante (QA8).
	QA8 – Dans quelle province ou territoire viviez-vous lorsque vous étiez au premier cycle ou au deuxième cycle du secondaire?			10, 11, 12, 13, 35, 46, 47, 48, 59, 60, 61, 62, 76, 77	QA9A	EV	Inscrire 6 en QA8B.
	QA8			24	CA8A		
C	CA8A			QA4A ≠ 1	QA9A	EV	Inscrire 6 en QA8B.
	QA8B						Aucune mesure à prendre, puisque la question mène à la suivante (QA9A).
	QA9A						

R : Refus

NSP : Ne sait pas

EV : Enchaînement valide

ND : Non déclaré

6.4.5 Vérifications de la cohérence (étape 5)

Les vérifications de la cohérence consistent à examiner la cohérence de l'information entre différentes variables. Bon nombre des vérifications de la cohérence ont été intégrées à l'application IAQ de l'EIACA, de sorte que les valeurs qui ne concordaient pas avec l'information antérieure étaient souvent signalées par l'intervieweur et clarifiées à la source. Dans le cas des fichiers canadiens, les étapes nécessaires pour respecter les normes internationales étaient minimales. Néanmoins, on a répété toutes les vérifications nécessaires des fichiers canadiens pour assurer la qualité des données. Cette étape a suivi immédiatement le processus des vérifications de cheminement.

Pour un grand nombre de ces questions, chaque pays a défini ses propres niveaux de tolérance en fonction de la réalité du pays (limites supérieure et inférieure du nombre d'années requises pour terminer un programme, limites supérieure et inférieure des niveaux de revenu, etc.).

Pour ce qui est de l'EIACA menée au Canada, toute dérogation au niveau de tolérance incitait l'intervieweur à demander une clarification. Toutefois, dans certains cas, il n'était possible de

corriger les dérogations qu'après la collecte. Il fallait naturellement apporter des mesures correctives indépendantes du répondant, d'où la nécessité d'un nombre minimal de vérifications de la cohérence.

L'équipe de traitement de l'EIACA a procédé aux vérifications de la cohérence ci-dessous pour satisfaire aux exigences de traitement de l'ELCA.

Vérifications de la cohérence concernant l'âge du répondant

N° de vérif.	Question	Règle
E1	Âge (d'après la date de naissance)	L'âge au moment de l'interview est de 16 ans et plus. ((DOBMM > CURRMTH) et DOBY=1987) ou (E0 <16 ou Agerange=1), alors inscrire E1.
E2	QA2 – Année d'immigration	L'année d'immigration ne peut être antérieure à l'année de naissance. QA2 < DOBY et (QA1E =1 ou QA1E=2) et (QA1=2 ou IMM_Q01=2), alors inscrire E2.
E3	QA3 – Nombre d'années de scolarité	Vérifier que le nombre d'années de scolarité (A3) et le niveau de scolarité atteint (A4) sont valides. Examiner la relation entre QA3 et (QA4A-QA6C) et s'assurer que le nombre d'années de scolarité concorde avec le niveau de scolarité atteint. Dans le cas contraire, inscrire E3.
E4	QA5 – Âge au début des études...	Vérifier que l'âge inscrit n'est pas plus élevé que l'âge du répondant. QA5 >E0 et QA5 ≠ 96, alors inscrire E4.
E5	QA6 – Âge au moment de compléter le niveau de scolarité le plus élevé	Vérifier que l'âge inscrit n'est pas plus élevé que l'âge du répondant. QA6 >E0 et QA6 ≠ 96, alors inscrire E5.

Vérifications de la cohérence concernant l'emploi

N° de vérif.	Question	Règle
E6	QD3A- A travaillé la dernière fois	L'année inscrite ne peut correspondre aux douze derniers mois. QD3B QD3A > CurrYear-1 CurrMth, alors inscrire E6.
E7	QD5MA-QD5ML – Mois travaillé QD6AMA-QD6AML – A travaillé tout le mois QD6BMA-QD6BML – A travaillé une partie du mois	Le mois inscrit en QD6AMA-QD6AML et QD6BMA-QD6BML ne peut être un autre mois que ceux inscrits respectivement en QD5MA-QD5ML. (QD5MA-QD5ML) = 1 et (QD6AMA-QD6AML) ≠ 1 et (QD6BMA-QD6BML) ≠ 1 respectivement, alors inscrire E7.
E8	QD6AMA-QD6AML – A travaillé tout le mois QD6BMA-QD6BML – A travaillé une partie du mois	Les mois déclarés dans ces deux questions doivent être mutuellement exclusifs. (QD6AMA-QD6AML) = (QD6BMA-QD6BML) respectivement, alors inscrire E8.
E9	QD9 – A été sans travail ...	Le répondant doit avoir répondu « Oui » en QD4 (a travaillé chaque mois de la période de référence)

N° de vérif.	Question	Règle
		Si la réponse en QD9 et QD4 ≠ 1, alors inscrire E9 ex. : QD4 ≠ 1 et QD9 dans (1,2,8,9)
E10	QD21A –QD21B – Quand avez-vous commencé à travailler à cet emploi...	La date inscrite ne peut être postérieure à celle de l'interview QD21B QD21A > CurrYear CurrMth, alors inscrire E10.
E11	QD24A-QD24B – A travaillé la dernière fois à cet emploi...	La date inscrite doit être => QD21A-QD21B QD24B QD24A < QD21B QD21A, alors inscrire E11.
E12	QD40 – Nombre de jours par semaine	Le répondant doit avoir répondu « Par jour » (2) à la question QD39 Si la réponse à QD40 = réponse et QD39 ≠ 02, alors inscrire E12.

Vérifications de la cohérence concernant le salaire et le traitement

N° de vérif.	Question	Règle
E13	QD41 – Salaire	Le montant inscrit doit correspondre à la réponse en QD39; sinon, inscrire E13. Exemples : QD39=01 (Par heure) – tranche acceptable : 1 \$ à 1 000 \$ QD39=02 (Par jour) – tranche acceptable : 50 \$ à 600 \$ QD39=03 (Par semaine) – tranche acceptable : 25 \$ à 2 000 \$ QD39=04 (À la quinzaine) – tranche acceptable : 50 \$ à 3 000 \$ QD39=05 (Deux fois par mois) – tranche acceptable : 50 \$ à 3 000 \$ QD39=06 (Par mois) – tranche acceptable : 150 \$ à 8 000 \$ QD39=07 (Par année) – tranche acceptable : 1 500 \$ à 100 000 \$ QD39=08 (À la pièce) – tranche acceptable : 1 \$ à 5 000 \$
E15	QD42 – Salaire mensuel	Le montant inscrit doit correspondre au salaire mensuel; sinon, inscrire E15. QD39 = par mois – tranche acceptable : 150 \$ à 8 000 \$

Vérifications de la cohérence concernant le programme d'études

N° de vérif.	Question	Règle	Observations
E16	QF8 – Nombre de semaines passées à suivre le programme	Ne peut être supérieur à 52 (niveau de tolérance seulement) QF8 >52 et QF8 < 96, alors vérifier E16.	OK
E17	QF9 – Nombre moyen d'heures par semaine	Ne peut être supérieur à 30 (niveau de tolérance seulement) QF9 >30 et QF9 < 96, alors vérifier E17.	Faire passer le niveau de tolérance à >40, car il y a 98 cas ayant une valeur de 40 et 31 cas entre les valeurs de 30 et 40.
E18	QF11 – Nombre d'heures par jour	Ne peut être supérieur à 6 (niveau de tolérance seulement) et QF11 > 6 et QF11 < 96, alors vérifier E19.	Faire passer le niveau de tolérance à >8, car il y a 184 cas ayant une valeur de 8.

Vérifications de la cohérence concernant les activités hors des études

N° de vérif.	Question	Règle
E19	QG5A à G5H – Sections d'un journal (K11)	Si la réponse en QG3A est égale à 1,2 ou 3, le répondant doit avoir répondu aux questions G5A à G5H. Si QG3A dans (1,2,3), alors l'inscrire; si (QG5A ≠ 1 ou QG5B ≠ 1 ou QG5C ≠ 1 ou QG5D ≠ 1 ou QG5E ≠ 1 ou QG5F ≠ 1 ou QG5G ≠ 1 ou QG5H ≠ 1), alors inscrire E25; fin;

Vérifications de la cohérence concernant le revenu total

N° de vérif.	Question	Règle
E20	K11	Le revenu déclaré doit être égal ou supérieur à K6 (Revenu personnel) Si (. < Qk6 < 9999996,96) et (. < Qk11 < 999999,96), alors l'inscrire; Si QK11 < QK6, alors inscrire E28; fin;

Il convient de souligner que le fichier canadien n'a fait l'objet d'aucune autre vérification de la cohérence; par conséquent, il reste probablement des cas où des incohérences persistent (par exemple, entre les renseignements sur le salaire déclarés à la section D et les renseignements sur le revenu déclarés à la section K). Le traitement pertinent de ces valeurs aberrantes est laissé à la discrétion de chaque chercheur.

6.4.6 Imputations et variables dérivées (étape 6)

Selon les règles internationales, certaines variables clés devaient faire l'objet d'imputations, soit l'âge, le groupe d'âge, le niveau de scolarité, l'année d'immigration, le statut d'Autochtone et la langue maternelle. Au Canada, toutefois, on n'a effectué des imputations qu'à l'égard de cinq variables; on n'a pas imputé de valeur à la langue maternelle, le champ ayant été enregistré avec exactitude pour chaque répondant canadien. On a créé un certain nombre de variables dérivées pour favoriser la recherche sur les antécédents et les résultats des compétences. Dans la plupart des cas, on n'a pas retenu ces variables dans le fichier de microdonnées à grande diffusion. Toutefois, celles qui ont été ajoutées au fichier principal sont décrites en détail dans la section qui suit la description des étapes de l'imputation.

Imputation

De façon générale, l'imputation consiste à remplacer les valeurs manquantes par des données plausibles; dans certains cas, toutefois, les réponses au questionnaire de présélection et d'autres renseignements connexes ont permis de clarifier et de modifier certaines incohérences dans les valeurs de ces variables clés. On a habituellement effectué ces imputations juste après les vérifications de la cohérence mais, certaines incohérences ayant été traitées aux étapes précédentes de la vérification, l'imputation a été réduite au minimum.

Âge du répondant

a) Source principale : l'âge calculé d'après la date de naissance du répondant dans le questionnaire de base.

b) Si l'on ne connaissait pas l'âge calculé, on a alors utilisé l'âge déclaré dans le questionnaire de présélection s'il était cohérent avec le groupe d'âge déclaré dans le questionnaire de base. (On a compté 131 cas de ce genre.)

c) Si l'âge n'était pas disponible dans les sources a) et b), on a utilisé l'imputation par enregistrement donneur choisi au hasard parmi le groupe d'âge correspondant. (On a compté 21 cas de ce genre.)

Niveau de scolarité

Seuls quatre enregistrements de réponse ne comportaient pas de niveau de scolarité. Le modèle d'imputation a utilisé des valeurs pour une foule de questions disponibles concernant notamment le nombre d'années de scolarité ou la participation à l'éducation des adultes; dans un cas, on a eu recours à une méthode hot deck en utilisant un enregistrement donneur convenablement apparié pour produire cet élément d'information manquant mais essentiel.

Né au Canada

On n'a pas posé la question QA1 (Êtes-vous né(e) au Canada) du questionnaire de base si on l'avait déjà posée au répondant dans le questionnaire de présélection. Les valeurs manquantes comme « Refus » et « Ne sait pas » ont été imputées par la valeur inscrite dans le questionnaire de présélection. Après cette imputation, tous les cas comportaient donc une réponse à la question QA1.

Année d'immigration

La question QA2 (année d'immigration au Canada) sert à distinguer les nouveaux immigrants des immigrants établis. Les valeurs manquantes de cette variable ont été imputées.

a) On a établi l'année d'immigration comme suit : l'année en cours moins le nombre d'années vécues au Canada. $\text{YearImmig} = 2003 - \text{QA2B}$ (nombre d'années vécues au Canada). (On a compté sept cas de ce genre.)

b) On a choisi au hasard un enregistrement donneur comportant le pays de naissance et le groupe d'âge correspondants. On a assuré la cohérence en appliquant la règle de vérification suivante : année de naissance \leq année d'immigration. (On a compté deux cas de ce genre.)

Statut d'Autochtone

On a posé à tous les répondants la question sur le statut d'Autochtone. Toutefois, on n'a pas posé la question du questionnaire de base si on l'avait déjà posée au répondant dans le questionnaire de présélection. La question sur le statut d'Autochtone est une question du type « Cocher toutes les réponses qui s'appliquent ». Seuls le Manitoba, la Saskatchewan et les trois territoires ont fait l'objet d'une imputation.

a) Les valeurs manquantes comme « Refus » et « Ne sait pas » ont été imputées par la réponse au questionnaire de présélection. (On a compté trois cas de ce genre.)

b) On a compté neuf cas comportant des réponses incohérentes (par ex. : <1> Oui, Indien d'Amérique du Nord et <4> Non Autochtone). On a résolu ces cas manuellement en examinant les réponses aux questions sur l'origine ethnique, l'immigration et la langue maternelle. (On a compté dix cas de ce genre.)



Langue maternelle

La question QB1 sur la langue maternelle est une autre variable qui devait faire l'objet d'une imputation afin de respecter les normes internationales. Au Canada, comme il n'y avait pas de valeurs manquantes dans ce champ, aucune imputation n'a été nécessaire à l'égard de la langue maternelle.

Variables dérivées

Une autre étape importante concerne le calcul de variables sommaires fondées sur les renseignements détaillés fournis par les répondants. La liste ci-dessous énumère les variables dérivées (VD) comprises dans le fichier de données principal de l'EIACA. Veuillez prendre note que ces variables dérivées ne sont pas comprises dans les fichiers de microdonnées à grande diffusion national ou international de l'EIACA pour le Canada. Les descriptions sont fournies à titre informatif et comprennent la logique utilisée pour créer la variable dérivée dans le fichier de données principal.

Variables dérivées simples

Activité sur le marché du travail

Activité au moment de l'enquête (deux catégories)

Nom de la VD : LFP1
Étiquette de la VD : Activité au moment de l'enquête (deux catégories)
Valeurs de l'étiquette : 0 « Inactif »; 1 « Actif »
Question(s) posée(s) : D1 : Quelle est votre situation actuelle d'emploi?
Pseudo-logique : D1 indique si le répondant était actif au moment de l'enquête, c'est-à-dire occupé, travailleur autonome ou en chômage et cherchant du travail.

Activité au moment de l'enquête (trois catégories)

Nom de la VD : LFP2
Étiquette de la VD : Activité au moment de l'enquête (trois catégories)
Valeurs de l'étiquette : 0 « Inactif »; 1 « En chômage »; 2 « Occupé »
Question(s) posée(s) : D1 : Quelle est votre situation actuelle d'emploi?
Pseudo-logique : D1 indique si le répondant était actif au moment de l'enquête et, parmi ceux qui étaient actifs, distingue entre ceux qui étaient occupés ou en chômage.

Régime de travail au cours des 12 derniers mois

Nom de la VD : LFSTAT12
Étiquette de la VD : Régime de travail au cours des 12 derniers mois
Valeurs de l'étiquette : 1 « Occupé toute l'année »; 2 « En chômage toute l'année »; 3 « Inactif toute l'année »; 4 « Occupé une partie de l'année, en chômage une partie de l'année »; 5 « Occupé une partie de l'année, inactif une partie de l'année », 6 « En chômage une partie de l'année, inactif une partie de l'année », 7 « Occupé, en chômage et inactif une partie de l'année »
Question(s) posée(s) : D2 : Avez-vous travaillé à un emploi ou dans une entreprise au cours des 12 derniers mois?
D4 : Au cours des 12 derniers mois, avez-vous travaillé chaque mois (au moins 8 heures par mois)?
D5 : Au cours de quels mois avez-vous travaillé au moins 8 heures par mois?
D6A : Au cours de quels mois avez-vous travaillé tout le mois, c'est-à-dire au moins 8 heures à chaque semaine?
D6B : Au cours de quels mois avez-vous travaillé une partie du mois?
D9 : Au cours des 12 derniers mois, avez-vous déjà été sans travail pendant une semaine ou plus parce que vous aviez changé d'employeur?
D10 : Durant quel(s) mois avez-vous été sans travail pendant une semaine ou plus parce que vous changiez d'employeur?
D11 : Au cours des 12 derniers mois, avez-vous été absent du travail pendant une semaine ou plus en raison d'une mise à pied temporaire?
D12 : Durant quel(s) mois avez-vous été mis à pied temporairement?

D13 : Avez-vous cherché du travail au cours des 12 derniers mois, c'est-à-dire cherché un nouvel emploi ou un emploi différent?
D14 : Au cours de quel(s) mois avez-vous cherché du travail?

Pseudo-logique : Les questions D2, D4, D5 (*mo1 à mo12*), D6A (*mo1 à mo12*), D6B (*mo1 à mo12*), D9, D10 (*mo1 à mo12*), D11, D12 (*mo1 à mo12*), D13 et D14 (*mo1 à mo12*) servent à créer douze variables provisoires EMPMO1 à EMPMO12 correspondant à la situation d'emploi des répondants pour chacun des 12 derniers mois. Les catégories de ces variables provisoires sont 1 « Occupé tout le mois »; 2 « En chômage tout le mois »; 3 « Inactif tout le mois »; 4 « Occupé une partie du mois, en chômage une partie du mois »; 5 « Occupé une partie du mois, inactif une partie du mois ». Les renseignements fournis par ces 12 variables provisoires servent à calculer la variable LFSTAT12 (Régime de travail sur 12 mois).

Nombre de périodes de chômage au cours des 12 derniers mois

Nom de la VD : UNEMP1
Étiquette de la VD : Nombre de périodes de chômage au cours des 12 derniers mois
Valeurs de l'étiquette : 1 « 0 période »; 2 « 1 période »; 3 « 2 périodes »; 4 « 3 périodes »; 5 « 4 périodes »; 6 « 5 périodes »

Question(s) posée(s) : LFSTAT12 (vd) : Régime de travail au cours des 12 derniers mois
EMPMO1 (vd) : Situation d'emploi au cours du mois 1
EMPMO2 (vd) : Situation d'emploi au cours du mois 2
EMPMO3 (vd) : Situation d'emploi au cours du mois 3
EMPMO4 (vd) : Situation d'emploi au cours du mois 4
EMPMO5 (vd) : Situation d'emploi au cours du mois 5
EMPMO6 (vd) : Situation d'emploi au cours du mois 6
EMPMO7 (vd) : Situation d'emploi au cours du mois 7
EMPMO8 (vd) : Situation d'emploi au cours du mois 8
EMPMO9 (vd) : Situation d'emploi au cours du mois 9
EMPMO10 (vd) : Situation d'emploi au cours du mois 10
EMPMO11 (vd) : Situation d'emploi au cours du mois 11
EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées LFSTAT12 et EMPMO1 à EMPMO12 servent à calculer le nombre de périodes de chômage vécues par les répondants au cours des 12 derniers mois. Pour les mois où il y a deux situations d'emploi, on fait une supposition. Dans ce cas, la proportion attribuée à chaque situation est un demi-mois. On établit que l'année consiste en 24 parties de mois. En outre, pour les mois où il y a deux situations d'emploi, on fait une supposition quant à l'ordre des activités. On établit l'ordre de façon à réduire au minimum le nombre de fois qu'une personne passe d'un état à l'autre et à maximiser la durée d'une situation donnée. On applique d'abord ce principe aux situations d'emploi, puis aux situations de chômage.

Nombre de semaines consécutives de chômage au moment de l'enquête (semaines)

Nom de la VD : UNEMP2
Étiquette de la VD : Nombre de semaines consécutives de chômage au moment de l'enquête (semaines)
Valeurs de l'étiquette : Nombre de semaines

Question(s) posée(s) : LFP2 (vd) : Activité au moment de l'enquête (trois catégories)
 LFSTAT12 (vd) : Régime de travail au cours des 12 derniers mois
 EMPMO1 (vd) : Situation d'emploi au cours du mois 1
 EMPMO2 (vd) : Situation d'emploi au cours du mois 2
 EMPMO3 (vd) : Situation d'emploi au cours du mois 3
 EMPMO4 (vd) : Situation d'emploi au cours du mois 4
 EMPMO5 (vd) : Situation d'emploi au cours du mois 5
 EMPMO6 (vd) : Situation d'emploi au cours du mois 6
 EMPMO7 (vd) : Situation d'emploi au cours du mois 7
 EMPMO8 (vd) : Situation d'emploi au cours du mois 8
 EMPMO9 (vd) : Situation d'emploi au cours du mois 9
 EMPMO10 (vd) : Situation d'emploi au cours du mois 10
 EMPMO11 (vd) : Situation d'emploi au cours du mois 11
 EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées LFP2, LFSTAT12 et EMPMO1 à EMPMO12 servent à calculer le nombre de semaines consécutives de chômage au moment de l'enquête.

Nombre de semaines consécutives de chômage au moment de l'enquête (deux catégories)

Nom de la VD : UNEMP3
 Étiquette de la VD : Nombre de semaines consécutives de chômage au moment de l'enquête (deux catégories)
 Valeurs de l'étiquette : 0 « Chômage de courte durée – moins de 6 mois »; 1 « Chômage de longue durée – 6 mois ou plus »
 Question(s) posée(s) : UNEMP2 (vd) : Nombre de semaines consécutives de chômage au moment de l'enquête (semaines)
 Pseudo-logique : UEMP2 est recodée en deux catégories.

Nombre de semaines de chômage au cours des 12 derniers mois

Nom de la VD : UNEMP4
 Étiquette de la VD : Nombre de semaines de chômage au cours des 12 derniers mois
 Valeurs de l'étiquette : Nombre de semaines
 Question(s) posée(s) : EMPMO1 (vd) : Situation d'emploi au cours du mois 1
 EMPMO2 (vd) : Situation d'emploi au cours du mois 2
 EMPMO3 (vd) : Situation d'emploi au cours du mois 3
 EMPMO4 (vd) : Situation d'emploi au cours du mois 4
 EMPMO5 (vd) : Situation d'emploi au cours du mois 5
 EMPMO6 (vd) : Situation d'emploi au cours du mois 6
 EMPMO7 (vd) : Situation d'emploi au cours du mois 7
 EMPMO8 (vd) : Situation d'emploi au cours du mois 8
 EMPMO9 (vd) : Situation d'emploi au cours du mois 9
 EMPMO10 (vd) : Situation d'emploi au cours du mois 10
 EMPMO11 (vd) : Situation d'emploi au cours du mois 11
 EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées EMPMO1 à EMPMO12 servent à calculer le nombre de semaines de chômage au cours des 12 derniers mois. On convertit les situations d'emploi mensuelles en semaines comme suit : d'abord, on établit que l'année consiste en 24 parties de mois; puis, on

convertit une partie du mois en semaines selon le calcul suivant :
365,25/ (7*24).

Nombre de mois sans travail pendant une semaine ou plus en raison d'une mise à pied temporaire

Nom de la VD : UNEMP5
Étiquette de la VD : Nombre de mois sans travail pendant une semaine ou plus en raison d'une mise à pied temporaire
Valeurs de l'étiquette : Nombre de mois
Question(s) posée(s) : D4 : Au cours des 12 derniers mois, avez-vous travaillé chaque mois (au moins 8 heures par mois)?
D11 : Au cours des 12 derniers mois, avez-vous été absent du travail pendant une semaine ou plus en raison d'une mise à pied temporaire?
D12 : Durant quel(s) mois avez-vous été mis à pied temporairement?
Pseudo-logique : D4, D11 et D12 (*mo1 à mo12*) déterminent le nombre de mois sans travail pendant une semaine ou plus pour ceux qui étaient occupés pendant les 12 mois. Les répondants qui ont travaillé au cours des cinq dernières années doivent répondre à la question D29, qui est recodée de façon à exclure ceux qui n'ont pas travaillé au cours des cinq dernières années.

Nombre de mois passés à chercher du travail au cours des 12 derniers mois

Nom de la VD : LK4WORK
Étiquette de la VD : Nombre de mois passés à chercher du travail au cours des 12 derniers mois
Valeurs de l'étiquette : Nombre de mois
Question(s) posée(s) : D13 : Avez-vous cherché du travail au cours des 12 derniers mois, c'est-à-dire cherché un nouvel emploi ou un emploi différent?
D14 : Au cours de quel(s) mois avez-vous cherché du travail?
Pseudo-logique : D14 détermine le nombre de mois passés à chercher du travail au cours des 12 derniers mois.

Nombre de semaines d'inactivité au cours des 12 derniers mois

Nom de la VD : NLFWKS1
Étiquette de la VD : Nombre de semaines d'inactivité au cours des 12 derniers mois
Valeurs de l'étiquette : Nombre de semaines
Question(s) posée(s) : EMPMO1 (vd) : Situation d'emploi au cours du mois 1
EMPMO2 (vd) : Situation d'emploi au cours du mois 2
EMPMO3 (vd) : Situation d'emploi au cours du mois 3
EMPMO4 (vd) : Situation d'emploi au cours du mois 4
EMPMO5 (vd) : Situation d'emploi au cours du mois 5
EMPMO6 (vd) : Situation d'emploi au cours du mois 6
EMPMO7 (vd) : Situation d'emploi au cours du mois 7
EMPMO8 (vd) : Situation d'emploi au cours du mois 8
EMPMO9 (vd) : Situation d'emploi au cours du mois 9
EMPMO10 (vd) : Situation d'emploi au cours du mois 10
EMPMO11 (vd) : Situation d'emploi au cours du mois 11
EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées EMPMO1 à EMPMO12 servent à calculer le nombre de semaines d'inactivité au cours des 12 derniers mois. On convertit les situations d'emploi mensuelles en semaines comme suit : d'abord, on établit que l'année consiste en 24 parties de mois; puis, on convertit une partie du mois en semaines selon le calcul suivant : $365,25 / (7 * 24)$.

Nombre de semaines consécutives d'inactivité au moment de l'enquête

Nom de la VD : NLFWKS2
Étiquette de la VD : Nombre de semaines consécutives d'inactivité au cours des 12 derniers mois

Valeurs de l'étiquette : Nombre de semaines

Question(s) posée(s) : LFP2 (vd) : Activité au moment de l'enquête (trois catégories)
LFSTAT12 (vd) : Régime de travail au cours des 12 derniers mois
EMPMO1 (vd) : Situation d'emploi au cours du mois 1
EMPMO2 (vd) : Situation d'emploi au cours du mois 2
EMPMO3 (vd) : Situation d'emploi au cours du mois 3
EMPMO4 (vd) : Situation d'emploi au cours du mois 4
EMPMO5 (vd) : Situation d'emploi au cours du mois 5
EMPMO6 (vd) : Situation d'emploi au cours du mois 6
EMPMO7 (vd) : Situation d'emploi au cours du mois 7
EMPMO8 (vd) : Situation d'emploi au cours du mois 8
EMPMO9 (vd) : Situation d'emploi au cours du mois 9
EMPMO10 (vd) : Situation d'emploi au cours du mois 10
EMPMO11 (vd) : Situation d'emploi au cours du mois 11
EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées LFP2, LFSTAT12 et EMPMO1 à EMPMO12 servent à calculer le nombre de semaines consécutives d'inactivité au moment de l'enquête.

Nombre de semaines de travail au cours des 12 derniers mois

Nom de la VD : EMPWKS1
Étiquette de la VD : Nombre de semaines de travail au cours des 12 derniers mois

Valeurs de l'étiquette : Nombre de semaines

Question(s) posée(s) : EMPMO1 (vd) : Situation d'emploi au cours du mois 1
EMPMO2 (vd) : Situation d'emploi au cours du mois 2
EMPMO3 (vd) : Situation d'emploi au cours du mois 3
EMPMO4 (vd) : Situation d'emploi au cours du mois 4
EMPMO5 (vd) : Situation d'emploi au cours du mois 5
EMPMO6 (vd) : Situation d'emploi au cours du mois 6
EMPMO7 (vd) : Situation d'emploi au cours du mois 7
EMPMO8 (vd) : Situation d'emploi au cours du mois 8
EMPMO9 (vd) : Situation d'emploi au cours du mois 9
EMPMO10 (vd) : Situation d'emploi au cours du mois 10
EMPMO11 (vd) : Situation d'emploi au cours du mois 11
EMPMO12 (vd) : Situation d'emploi au cours du mois 12

Pseudo-logique : Les variables dérivées EMPMO1 à EMPMO12 servent à calculer le nombre de semaines de travail au cours des 12 derniers mois. On convertit les situations d'emploi mensuelles en semaines comme suit : d'abord, on établit que l'année consiste en 24 parties de mois; puis, on

convertit une partie du mois en semaines selon le calcul suivant :
365,25/ (7*24).

Nombre de semaines consécutives de travail au moment de l'enquête

Nom de la VD : EMPWKS2
Étiquette de la VD : Nombre de semaines consécutives de travail au cours des 12 derniers mois
Valeurs de l'étiquette : Nombre de semaines
Question(s) posée(s) : LFP2 (vd) : Activité au moment de l'enquête (trois catégories)
LFSTAT12 (vd) : Régime de travail au cours des 12 derniers mois
EMPMO1 (vd) : Situation d'emploi au cours du mois 1
EMPMO2 (vd) : Situation d'emploi au cours du mois 2
EMPMO3 (vd) : Situation d'emploi au cours du mois 3
EMPMO4 (vd) : Situation d'emploi au cours du mois 4
EMPMO5 (vd) : Situation d'emploi au cours du mois 5
EMPMO6 (vd) : Situation d'emploi au cours du mois 6
EMPMO7 (vd) : Situation d'emploi au cours du mois 7
EMPMO8 (vd) : Situation d'emploi au cours du mois 8
EMPMO9 (vd) : Situation d'emploi au cours du mois 9
EMPMO10 (vd) : Situation d'emploi au cours du mois 10
EMPMO11 (vd) : Situation d'emploi au cours du mois 11
EMPMO12 (vd) : Situation d'emploi au cours du mois 12
Pseudo-logique : Les variables dérivées LFP2, LFSTAT12 et EMPMO1 à EMPMO12 servent à calculer le nombre de semaines consécutives de travail au moment de l'enquête.

Nombre de semaines de travail à l'emploi principal (actuel ou dernier)

Nom de la VD : EMPWKS3
Étiquette de la VD : Nombre de semaines de travail à l'emploi principal (actuel ou dernier)
Valeurs de l'étiquette : Nombre de semaines
Question(s) posée(s) : D34 : Pendant combien de mois par année travaillez/travailliez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?
D35 : Travaillez/travailliez-vous habituellement toutes les semaines du mois à cet emploi ou dans cette entreprise?
D36 : En moyenne, combien de semaines par mois travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
Pseudo-logique : D34, D35 et D36 déterminent le nombre de semaines de travail à l'emploi principal (actuel ou dernier). Si, à la question D35, le répondant a déclaré avoir travaillé toutes les semaines du mois, le nombre estimatif de semaines de travail au cours d'un mois est 4,33.

Nombre d'heures de travail à l'emploi principal (actuel ou dernier)

Nom de la VD : EMPHRS
Étiquette de la VD : Nombre d'heures de travail à l'emploi principal (actuel ou dernier)
Valeurs de l'étiquette : Nombre d'heures
Question(s) posée(s) : D34 : Pendant combien de mois par année travaillez/travailliez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?
D35 : Travaillez/travailliez-vous habituellement toutes les semaines du mois à cet emploi ou dans cette entreprise?

D36 : En moyenne, combien de semaines par mois travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
D37 : En moyenne, combien d'heures par semaine travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
Pseudo-logique : D34, D35, D36 et D37 déterminent le nombre d'heures de travail à l'emploi principal (actuel ou dernier). Si, à la question D35, le répondant a déclaré avoir travaillé toutes les semaines du mois, le nombre estimatif de semaines de travail au cours d'un mois est 4,33.

Intensité du travail au cours des 52 dernières semaines (six catégories)

Nom de la VD : EMPFULL1
Étiquette de la VD : Intensité du travail au cours des 52 dernières semaines (six catégories)
Valeurs de l'étiquette : 1 « Occupé à temps plein au cours des 52 dernières semaines »; 2 « Occupé à temps plein, mais < 52 semaines car inactif une partie de l'année »; 3 « Occupé à temps plein, mais < 52 semaines car en chômage une partie de l'année »; 4 « Occupé à temps partiel ou à parts égales à temps plein et à temps partiel au cours des 52 dernières semaines »; 5 « Occupé à temps partiel ou à parts égales à temps plein et à temps partiel, mais < 52 semaines car inactif une partie de l'année »; 6 « Occupé à temps partiel ou à parts égales à temps plein et à temps partiel, mais < 52 semaines car en chômage une partie de l'année »
Question(s) posée(s) : D7 : Au cours des 12 derniers mois, avez-vous travaillé principalement à temps plein, c'est-à-dire 30 heures par semaine ou plus, ou à temps partiel, c'est-à-dire moins de 30 heures par semaine?
LFP2 (vd) : Activité au moment de l'enquête (trois catégories)
LFSTAT12 (vd) : Régime de travail au cours des 12 derniers mois
EMPWKS1 : Nombre de semaines de travail au cours des 12 derniers mois
Pseudo-logique : D7, LFSTAT12, LFP2, EMPWKS1 déterminent l'intensité du travail au cours des 52 dernières semaines précédant l'enquête.

Statut de travailleur autonome

Nom de la VD : EMPSELF
Étiquette de la VD : Statut de travailleur autonome
Valeurs de l'étiquette : 0 « Sans statut de travailleur autonome dans l'emploi principal »; 1 « Avec statut de travailleur autonome dans l'emploi principal »
Question(s) posée(s) : D29 : Quelle est/était votre situation professionnelle dans cet emploi ou dans cette entreprise?
Pseudo-logique : D29 détermine si le répondant est un travailleur autonome ou non à son emploi principal (actuel ou dernier).

Responsabilités de supervision

Nom de la VD : SUPERVIS
Étiquette de la VD : Responsabilités de supervision
Valeurs de l'étiquette : 0 « Sans responsabilités de supervision dans l'emploi principal »; 1 « Avec responsabilités de supervision dans l'emploi principal »
Question(s) posée(s) : D29 : Quelle est/était votre situation professionnelle dans cet emploi ou dans cette entreprise?

Pseudo-logique : D29 détermine si le répondant a des responsabilités de supervision ou non à son emploi principal (actuel ou dernier). On suppose qu'un travailleur autonome ayant des employés assume ce genre de responsabilité.

Taille de l'entreprise

Nom de la VD : FIRMSIZE
Étiquette de la VD : Taille de l'entreprise
Valeurs de l'étiquette : 1 « Moins de 20 »; 2 « entre 20 et 99 »; 3 « entre 100 et 499 »; 4 « entre 500 et 999 »; 5 « 1 000 ou plus »
Question(s) posée(s) : D30 : Environ combien de personnes sont ou ont été employées à l'endroit où vous travaillez/iez? Est-ce moins de 20, entre 20 et 99, entre 100 et 499, entre 500 et 999 ou 1 000 ou plus?
D31 : Votre employeur a/avait-il/elle des opérations à plusieurs endroits au Canada?
D33 : Au total, environ combien de personnes travaillent/aient pour cet employeur à ces différents endroits? Est-ce moins de 20, entre 20 et 99, entre 100 et 499, entre 500 et 999 ou 1 000 et plus?
Pseudo-logique : D30, D31 et D33 déterminent la taille de l'entreprise dans laquelle le répondant travaille ou a travaillé, tous les endroits compris.

Ancienneté (mois)

Nom de la VD : TENURE1
Étiquette de la VD : Nombre de mois de travail au même emploi
Valeurs de l'étiquette : Nombre de mois
Question(s) posée(s) : D1 : Quelle est votre situation actuelle d'emploi?
D2 : Avez-vous travaillé à un emploi ou dans une entreprise au cours des 12 derniers mois?
D3 : Quand avez-vous travaillé la dernière fois à un emploi ou dans une entreprise?
D21 : Quand avez-vous commencé à travailler à cet emploi ou pour cette entreprise?
D23 : Travaillez-vous encore à cet emploi ou à cette entreprise?
D24 : Quand avez-vous travaillé la dernière fois à cet emploi ou pour cette entreprise?
Pseudo-logique : D1, D2, D3, D21, D23 et D24 déterminent le nombre de mois pendant lesquels les répondants ont travaillé au même emploi (actuel ou principal).

Ancienneté (quatre catégories)

Nom de la VD : TENURE2
Étiquette de la VD : Ancienneté (quatre catégories)
Valeurs de l'étiquette : 1 « Ancienneté établie – poste occupé pendant les 12 derniers mois »; 2 « Ancienneté récente – poste occupé pendant moins des 12 derniers mois »; 3 « Durée d'occupation de l'emploi précédent – occupé dans un autre emploi »; 4 « Durée d'occupation de l'emploi précédent – inoccupé au cours des 12 derniers mois »
Question(s) posée(s) : TENURE1 : Nombre de mois de travail dans le même emploi
D23 : Travaillez-vous encore à cet emploi ou à cette entreprise?

Pseudo-logique : D23 et TENURE1 déterminent quand la personne a acquis son ancienneté et si c'était au cours des 12 derniers mois.

Participation à l'éducation et à l'apprentissage

Destination du programme d'études ou de formation suivi au cours des 12 derniers mois (étudiants)

Nom de la VD : PROGDES1

Étiquette de la VD : Destination du programme d'études ou de formation suivi au cours des 12 derniers mois (étudiants)

Valeurs de l'étiquette : 1 « Étudiant à un programme d'études ou de formation, niveaux 0 à 3 de la CITE »; 2 « Étudiant à un programme d'études ou de formation, niveaux 4 et 5B de la CITE »; 3 « Étudiant à un programme d'études ou de formation, niveaux 5A et 6 de la CITE »; 4 « Étudiant à un programme d'études ou de formation dans un établissement régulier, ne sait pas à quel niveau »; 5 « Étudiant à un programme d'études ou de formation, mais pas dans un établissement régulier »; 6 « Étudiant à un programme d'études ou de formation, ne sait pas si c'est dans un établissement régulier »; 7 « Étudiant ayant suivi des études ou une formation, mais pas dans le cadre d'un programme »; 8 « Étudiant ayant suivi des études ou une formation, mais ne sait pas si c'était dans le cadre d'un programme »; 9 « Étudiant, mais n'a participé à aucun programme d'études ou de formation »; 10 « Pas aux études au moment de l'enquête »

Question(s) posée(s) : D1 : Quelle est votre situation actuelle d'emploi? (L'un des choix est « Étudiant(e) ».)

F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?

F2 : Au cours des 12 derniers mois, avez-vous suivi des cours dans le cadre d'un programme d'études menant à un certificat ou à un diplôme?

F3 : Ce programme d'études était-il offert par une école, un collège ou une université?

F4 : Quel type de certificat, de grade ou de diplôme désiriez-vous obtenir grâce à ces études ou à cette formation? (Cf. les niveaux de la CITE97.)

Pseudo-logique : D1, F1, F2, F3 et F4 déterminent la destination du programme d'études ou de formation suivi au cours des 12 derniers mois. Cette variable concerne uniquement les participants à un programme qui étaient aux études au moment de l'enquête.

Destination du programme d'études ou de formation suivi au cours des 12 derniers mois (tous les répondants)

Nom de la VD : PROGDES2

Étiquette de la VD : Destination du programme d'études ou de formation suivi au cours des 12 derniers mois (tous les répondants)

Valeurs de l'étiquette : 1 « Suit un programme d'études ou de formation, niveaux 0 à 3 de la CITE »; 2 « Suit un programme d'études ou de formation, niveaux 4 et 5B de la CITE »; 3 « Suit un programme d'études ou de formation, niveaux 5A et 6 de la CITE »; 4 « Suit un programme d'études ou de

formation dans un établissement régulier, ne sait pas à quel niveau »; 5 « Suit un programme d'études ou de formation, mais pas dans un établissement régulier »; 6 « Suit un programme d'études ou de formation, ne sait pas si c'est dans un établissement régulier »; 7 « A suivi des études ou une formation, mais pas dans le cadre d'un programme »; 8 « A suivi des études ou une formation, mais ne sait pas si c'était dans le cadre d'un programme »; 9 « N'a participé à aucun programme d'études ou de formation »

Question(s) posée(s) : F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?

F2 : Au cours des 12 derniers mois, avez-vous suivi des cours dans le cadre d'un programme d'études menant à un certificat ou à un diplôme?

F3 : Ce programme d'études était-il offert par une école, un collège ou une université?

F4 : Quel type de certificat, de grade ou de diplôme désiriez-vous obtenir grâce à ces études ou à cette formation? (Cf. les niveaux de la CITE97.)

Pseudo-logique : F1, F2, F3 et F4 déterminent la destination du programme d'études ou de formation suivi au cours des 12 derniers mois. Cette variable concerne tous les répondants qui ont déclaré participer à un programme.

Participation à un programme d'études ou de formation au cours des 12 derniers mois

Nom de la VD : PROGPARG

Étiquette de la VD : Participation à un programme d'études ou de formation au cours des 12 derniers mois

Valeurs de l'étiquette : 1 « A participé à un programme d'études ou de formation »; 2 « A suivi des études ou une formation, mais pas dans le cadre d'un programme »; 3 « A suivi des études ou une formation, mais ne sait pas si c'était dans le cadre d'un programme »; 4 « N'a participé à aucun programme d'études ou de formation »

Question(s) posée(s) : PROGDES2 : Destination du programme d'études ou de formation suivi au cours des 12 derniers mois (tous les répondants)

Pseudo-logique : PROGDES2 est recodée en quatre catégories pour spécifier si les répondants ont suivi des études ou une formation et si c'était dans le cadre d'un programme ou non.

Intensité du programme d'études ou de formation suivi au cours des 12 derniers mois (étudiants)

Nom de la VD : PROGIN1

Étiquette de la VD : Intensité du programme d'études ou de formation suivi au cours des 12 derniers mois (étudiants)

Valeurs de l'étiquette : 1 « Aux études à temps plein dans le cadre d'un programme officiel »; 2 « Aux études à temps partiel dans le cadre d'un programme officiel »; 3 « Aux études à parts égales à temps plein et à temps partiel dans le cadre d'un programme officiel »; 4 « Aux études dans le cadre d'un programme officiel, mais ne connaît pas l'intensité »; 5 « Aux études, mais pas dans le cadre d'un programme officiel, ne connaît pas l'intensité »; 6 « Aux études dans le cadre d'un programme, ne sait pas ou refuse de dire s'il s'agit d'un programme officiel, ne connaît pas

l'intensité »; 7 « Aux études, mais n'a pas étudié dans le cadre d'un programme »; 8 « Aux études, mais n'a participé à aucun programme d'études ou de formation »; 9 « Pas aux études au moment de l'enquête »

Question(s) posée(s) : D1 : Quelle est votre situation actuelle d'emploi? (L'un des choix est « Étudiant(e) ».)

F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?

F2 : Au cours des 12 derniers mois, avez-vous suivi des cours dans le cadre d'un programme d'études menant à un certificat ou à un diplôme?

F3 : Ce programme d'études était-il offert par une école, un collège ou une université?

F6 : Étiez-vous aux études principalement à temps plein ou à temps partiel?

Pseudo-logique : D1, F1, F2, F3 et F6 déterminent l'intensité du programme d'études ou de formation suivi au cours des 12 derniers mois. Cette variable concerne uniquement les participants au programme qui étaient aux études au moment de l'enquête.

Intensité du programme d'études ou de formation suivi au cours des 12 derniers mois (tous les répondants)

Nom de la VD : PROGINT2

Étiquette de la VD : Intensité du programme d'études ou de formation suivi au cours des 12 derniers mois (tous les répondants)

Valeurs de l'étiquette : 1 « Aux études à temps plein dans le cadre d'un programme officiel »; 2 « Aux études à temps partiel dans le cadre d'un programme officiel »; 3 « Aux études à parts égales à temps plein et à temps partiel dans le cadre d'un programme officiel »; 4 « Aux études dans le cadre d'un programme officiel, mais ne connaît pas l'intensité »; 5 « Aux études, mais pas dans le cadre d'un programme officiel, ne connaît pas l'intensité »; 6 « Aux études dans le cadre d'un programme, ne sait pas ou refuse de dire s'il s'agit d'un programme officiel, ne connaît pas l'intensité »; 7 « N'a pas étudié dans le cadre d'un programme »; 8 « N'a participé à aucun programme d'études ou de formation »

Question(s) posée(s) : F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?

F2 : Au cours des 12 derniers mois, avez-vous suivi des cours dans le cadre d'un programme d'études menant à un certificat ou à un diplôme?

F3 : Ce programme d'études était-il offert par une école, un collège ou une université?

F6 : Étiez-vous aux études principalement à temps plein ou à temps partiel?

Pseudo-logique : F1, F2, F3 et F6 déterminent l'intensité du programme d'études ou de formation suivi au cours des 12 derniers mois. Cette variable concerne tous les répondants qui ont déclaré participer à un programme.

Durée du programme d'études ou de formation suivi au cours des 12 derniers mois (heures)

Nom de la VD : PROGDUR1

Étiquette de la VD : Durée du programme d'études ou de formation suivi au cours des 12 derniers mois (heures)

Valeurs de l'étiquette : Nombre d'heures

Question(s) posée(s) : F7 : Au cours des 12 derniers mois, la durée totale de ce programme s'exprime-t-elle le plus facilement en nombre de semaines, de jours ou d'heures?
 F8 : Du <insérez le mois et l'année>, pendant combien de semaines avez-vous suivi des cours dans le cadre de ce programme d'études?
 F9 : En moyenne, combien d'heures par semaine cela représentait-il?
 F10 : Du <insérez le mois et l'année>, pendant combien de jours au total avez-vous suivi des cours dans le cadre de ce programme?
 F11 : En moyenne, combien d'heures par jour cela représentait-il?
 F12 : Du <insérez le mois et l'année>, combien d'heures ce programme d'études a-t-il duré au total?

Pseudo-logique : F7, F8, F9, F10, F11, F12 et PROGINT2 déterminent le nombre d'heures consacrées au programme d'études par tous les répondants. Pour ceux qui ont répondu 'Principalement à temps plein' en F6 et pour lesquels on est passé à la question F13, on impute une durée de 40 heures par semaine pendant 52 semaines.

Durée du programme d'études ou de formation suivi au cours des 12 derniers mois (trois catégories)

Nom de la VD : PROGDUR2

Étiquette de la VD : Durée du programme d'études ou de formation suivi au cours des 12 derniers mois (trois catégories)

Valeurs de l'étiquette : 1 « Courte durée – 40 heures ou moins »; 2 « Durée moyenne – entre 41 et 100 heures »; 3 « Longue durée – 101 heures et plus »

Question(s) posée(s) : PROGDUR1 : Durée du programme d'études ou de formation suivi au cours des 12 derniers mois (heures)

Pseudo-logique : PROGDUR1 est recodée en trois catégories.

Population participant à l'éducation des adultes

Nom de la VD : ADEDPOP

Étiquette de la VD : Population participant à l'éducation des adultes

Valeurs de l'étiquette : 1 « Population participant à l'éducation des adultes »; 2 « Population ne participant pas à l'éducation des adultes »

Question(s) posée(s) : AGE_RESP : Âge des répondants.
 PROGDES2 : Destination du programme pour tous ceux qui étudiaient au cours des 12 derniers mois.
 PROGINT2 : Intensité des études au cours des 12 derniers mois pour tous ceux qui étudiaient dans le cadre d'un programme.
 F14B : Qui a contribué à assumer les frais directs de ce programme d'études, c'est-à-dire les dépenses engagées pour les droits de scolarité, le matériel de cours d'apprentissage, les déplacements, l'hébergement, etc.? Un employeur?
 F14D : Qui a contribué à assumer les frais directs de ce programme d'études, c'est-à-dire les dépenses engagées pour les droits de scolarité, le matériel de cours d'apprentissage, les déplacements, l'hébergement, etc.? Un syndicat ou une association professionnelle?

Pseudo-logique : F14B, F14D, AGE_RESP, PROGDES1 et PROGIN2 déterminent la population qui participait à l'éducation et à la formation des adultes. Les étudiants à temps plein âgés de 16 à 25 ans inscrits à des programmes des catégories 0 à 6 de la CITE et dont les dépenses n'étaient pas assumées par un employeur, un syndicat ou une association professionnelle sont exclus de la population adulte.

Participation à l'éducation ou à la formation au cours des 12 derniers mois

Nom de la VD : PELPART1
Étiquette de la VD : Participation à l'éducation ou à la formation au cours des 12 derniers mois
Valeurs de l'étiquette : 0 « N'a participé à aucun programme d'études ou de formation »; 1 « A participé à un programme d'études ou de formation »
Question(s) posée(s) : F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?
Pseudo-logique : F1 est recodée en deux catégories, selon que les répondants ont suivi ou non des études ou une formation au cours des 12 derniers mois.

Participation à un programme d'études ou de formation au cours des 12 derniers mois

Nom de la VD : PELPART2
Étiquette de la VD : Participation à un programme d'études ou de formation au cours des 12 derniers mois
Valeurs de l'étiquette : 0 'N'a pas participé à un programme d'études ou de formation'; 1 'A participé à un programme d'études ou de formation'
Question(s) posée(s) : F2 : Au cours des 12 derniers mois, avez-vous suivi des cours dans le cadre d'un programme d'études menant à un certificat ou à un diplôme?
Pseudo-logique : F2 est recodée en deux catégories, selon que les répondants ont participé ou non à un programme d'études ou de formation au cours des 12 derniers mois.

Participation à un cours d'études ou de formation ou plus au cours des 12 derniers mois

Nom de la VD : PELPART3
Étiquette de la VD : Participation à un cours d'études ou de formation ou plus au cours des 12 derniers mois
Valeurs de l'étiquette : 0 « N'a pas participé à un ou des cours d'études ou de formation »; 1 « A participé à un ou des cours d'études ou de formation »
Question(s) posée(s) : F1 : Au cours des 12 derniers mois, avez-vous reçu une formation ou suivi des études?
F15 : Au cours des 12 derniers mois, avez-vous suivi des cours qui ne faisaient pas partie de votre programme d'études?
F16 : Au cours des 12 derniers mois, avez-vous suivi des cours qui ne faisaient pas partie d'un programme d'études?
Pseudo-logique : F1, F15 et F16 déterminent si les répondants ont participé à un ou des cours d'études ou de formation qui ne faisaient pas partie d'un programme d'études au cours des 12 derniers mois.

Participation à des types d'apprentissage non structuré au cours des 12 derniers mois

Nom de la VD :	PELPART4
Étiquette de la VD :	Participation à des types d'apprentissage non structuré au cours des 12 derniers mois
Valeurs de l'étiquette :	0 'N'a participé à aucun type d'apprentissage non structuré'; 1 'A participé à des types d'apprentissage non structuré'
Question(s) posée(s) :	F26A : Visiter des salons professionnels, aller à des conférences ou à des congrès. F26B : Assister à des exposés, des séminaires, des ateliers ou des causeries spéciales ne faisant pas partie d'un cours. F26C : Lire des manuels, des ouvrages de référence, des revues et d'autres documents hors du cadre d'un cours. F26D : Faire des visites guidées dans des musées, des galeries d'art ou ailleurs. F26E : Apprendre au moyen de l'ordinateur ou d'Internet hors du cadre d'un cours. F26F : Apprendre au moyen de la télévision, de la vidéo, de cassettes hors du cadre d'un cours. F26G : Apprendre en observant, en obtenant de l'aide ou des conseils d'autres personnes, mais pas d'un professeur. F26H : Apprendre par soi-même en faisant des essais, en s'exerçant, en essayant différentes méthodes de faire les choses. F26I : Être envoyé(e) dans une grande organisation pour connaître différents aspects de cette organisation.
Pseudo-logique :	F26A à F26I déterminent si les répondants ont participé à des types d'apprentissage non structuré au cours des 12 derniers mois. S'ils ont répondu oui à au moins une de ces questions, on considère qu'ils ont participé à des types d'apprentissage non structuré.

Participation à l'éducation ou à la formation au cours des 12 derniers mois (huit catégories)

Nom de la VD :	PELPART5
Étiquette de la VD :	Participation à l'éducation ou à la formation au cours des 12 derniers mois (huit catégories)
Valeurs de l'étiquette :	1 « A participé à un programme d'études ou de formation seulement »; 2 « A participé à un ou des cours d'études ou de formation seulement »; 3 « A participé à des types d'apprentissage non structuré seulement »; 4 « A participé à un programme d'études ou de formation, à un ou des cours et à l'apprentissage non structuré »; 5 « A participé à un programme et à un ou des cours d'études ou de formation »; 6 « A participé à un programme d'études ou de formation et à des types d'apprentissage non structuré »; 7 « A participé à un ou des cours d'études ou de formation et à des types d'apprentissage non structuré »; 8 « N'a participé à aucun programme d'études ou de formation ni à des types d'apprentissage non structuré »
Question(s) posée(s) :	PELPART1 : Participation à l'éducation ou à la formation au cours des 12 derniers mois PELPART2 : Participation à un programme d'études ou de formation au cours des 12 derniers mois

PELPART3 : Participation à un ou des cours d'études ou de formation au cours des 12 derniers mois
PELPART4 : Participation à des types d'apprentissage non structuré au cours des 12 derniers mois
Pseudo-logique : PELPART1, PELPART2, PELPART3 et PELPART4 déterminent les types de participation à l'éducation ou à la formation ou les types d'apprentissage non structuré au cours des 12 derniers mois.

Participation à l'éducation ou à la formation au cours des 12 derniers mois (quatre catégories)

Nom de la VD : PELPART6
Étiquette de la VD : Participation à l'éducation ou à la formation au cours des 12 derniers mois (quatre catégories)
Valeurs de l'étiquette : 1 « A participé à un programme d'études ou de formation seulement »; 2 « A participé à un ou des cours d'études ou de formation seulement »; 3 « A participé à un programme d'études ou de formation, à un ou des cours et à l'apprentissage non structuré »; 4 « N'a participé à aucun programme d'études ou de formation ni à des types d'apprentissage non structuré »
Question(s) posée(s) : Même(s) que pour PELPART5
Pseudo-logique : PELPART1, PELPART2, PELPART3 et PELPART4 déterminent les types de participation à l'éducation ou à la formation ou les types d'apprentissage non structuré au cours des 12 derniers mois.

Nombre de cours suivis au cours des 12 derniers mois qui ne font pas partie d'un programme d'études ou de formation

Nom de la VD : ETCNUM
Étiquette de la VD : Nombre de cours suivis au cours des 12 derniers mois qui ne font pas partie d'un programme d'études ou de formation
Valeurs de l'étiquette : 1 « 1 cours »; 2 « 2 cours »; 3 « 3 cours »; 4 « 4 cours ou plus »
Question(s) posée(s) : *Pour ceux qui avaient suivi des cours dans le cadre d'un programme au cours des 12 derniers mois :*
F15 : Avez-vous suivi des cours qui ne faisaient pas partie de votre programme d'études?
F15A : Si oui, combien?
Pour ceux qui n'avaient pas suivi de cours dans le cadre d'un programme au cours des 12 derniers mois :
F16 : Avez-vous suivi des cours qui ne faisaient pas partie d'un programme d'études?
F16A : Si oui, combien?
Pseudo-logique : F15, F15A, F16, F16A déterminent le nombre de cours suivis au cours des 12 derniers mois qui ne font pas partie d'un programme d'études ou de formation.

Durée du cours 1 (heure)

Nom de la VD : ETCBUR1
Étiquette de la VD : Durée du cours 1 (heures)
Valeurs de l'étiquette : Nombre d'heures

Question(s) posée(s) : F18A : La durée totale de ce cours s'exprime-t-elle plus facilement en nombre de semaines, de jours ou d'heures?
F19A : Au cours des 12 derniers mois, pendant combien de semaines avez-vous suivi ce cours?
F20A : En moyenne, combien d'heures par semaine cela représentait-il?
F21A : Au cours des 12 derniers mois, pendant combien de jours au total avez-vous suivi ce cours?
F22A : En moyenne, combien d'heures par jour cela représentait-il?
F23A : Au cours des 12 derniers mois, combien d'heures ce cours a-t-il duré au total?

Pseudo-logique : F18A, F19A, F20A, F21A, F22A et F23A déterminent la durée du cours 1 en heures.

Durée du cours 2 (heures)

Nom de la VD : ETCDUR2
Étiquette de la VD : Durée du cours 2 (heures)
Valeurs de l'étiquette : Nombre d'heures

Question(s) posée(s) : F18B : La durée totale de ce cours s'exprime-t-elle plus facilement en nombre de semaines, de jours ou d'heures?
F19B : Au cours des 12 derniers mois, pendant combien de semaines avez-vous suivi ce cours?
F20B : En moyenne, combien d'heures par semaine cela représentait-il?
F21B : Au cours des 12 derniers mois, pendant combien de jours au total avez-vous suivi ce cours?
F22B : En moyenne, combien d'heures par jour cela représentait-il?
F23B : Au cours des 12 derniers mois, combien d'heures ce cours a-t-il duré au total?

Pseudo-logique : F18B, F19B, F20B, F21B, F22B et F23B déterminent la durée du cours 2 en heures.

Durée du cours 3 (heures)

Nom de la VD : ETCDUR3
Étiquette de la VD : Durée du cours 3 (heures)
Valeurs de l'étiquette : Nombre d'heures

Question(s) posée(s) : F18C : La durée totale de ce cours s'exprime-t-elle plus facilement en nombre de semaines, de jours ou d'heures?
F19C : Au cours des 12 derniers mois, pendant combien de semaines avez-vous suivi ce cours?
F20C : En moyenne, combien d'heures par semaine cela représentait-il?
F21C : Au cours des 12 derniers mois, pendant combien de jours au total avez-vous suivi ce cours?
F22C : En moyenne, combien d'heures par jour cela représentait-il?
F23C : Au cours des 12 derniers mois, combien d'heures ce cours a-t-il duré au total?

Pseudo-logique : F18C, F19C, F20C, F21C, F22C et F23C déterminent la durée du cours 3 en heures.

Durée de tous les cours (heures)

Nom de la VD : ETCDUR1

Étiquette de la VD : Durée de tous les cours (heures)
Valeurs de l'étiquette : Nombre d'heures
Question(s) posée(s) : ETC DUR1 : Durée du cours 1 (heures)
ETC DUR2 : Durée du cours 2 (heures)
ETC DUR3 : Durée du cours 3 (heures)
ETC NUM : Nombre de cours suivis au cours des 12 derniers mois qui ne font pas partie d'un programme d'études ou de formation
Pseudo-logique : ETC DUR1, ETC DUR2, ETC DUR3 et ETC NUM déterminent le nombre total d'heures consacrées à tous les cours au cours des 12 derniers mois.

Renseignements sur le ménage et le revenu

Enfants à charge de moins de 16 ans vivant dans le ménage

Nom de la VD : KIDSHOME
Étiquette de la VD : Enfants à charge de moins de 16 ans vivant dans le ménage
Valeurs de l'étiquette : 0 « Pas d'enfants de moins de 16 ans vivant dans le ménage »;
1 « Enfants de moins de 16 ans vivant dans le ménage »
Question(s) posée(s) : K1 : Y compris vous-même, combien de personnes vivent dans votre ménage?
K2 : Avez-vous des enfants à charge qui vivent avec vous dans votre ménage (enfants dont vous êtes responsable sur le plan financier et/ou dont vous avez la garde exclusive ou partagée)?
K3 : Quel est l'âge du plus jeune enfant de votre ménage?
Pseudo-logique : K1, K2 et K3 déterminent s'il y a un ou des enfants à charge de moins de 16 ans vivant dans le ménage.

Gains annuels tirés de l'emploi principal

Nom de la VD : EARNJOB1
Étiquette de la VD : Gains annuels tirés de l'emploi principal
Valeurs de l'étiquette : Valeur des gains annuels dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US), Norvège (couronne norvégienne) et Suisse (franc suisse).*
Question(s) posée(s) : D29 : Quelle est/était votre situation professionnelle dans cet emploi ou dans cette entreprise?
D34 : Pendant combien de mois par année travaillez/travailliez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?
D35 : Travaillez/travailliez-vous habituellement toutes les semaines du mois à cet emploi ou dans cette entreprise?
D36 : En moyenne, combien de semaines par mois travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
D37 : En moyenne, combien d'heures par semaine travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
D39 : Quelle est la façon la plus facile pour vous de nous dire votre salaire ou traitement habituel pour cet emploi? Serait-il par heure, par jour, par semaine, par année, ou autrement?
D40 : Combien de jours par semaine travaillez/travailliez-vous habituellement à cet emploi ou pour cette entreprise?

D41 : Pour cet emploi, quel est/était votre salaire ou traitement avant impôt et autres retenues à la source dans cet emploi, incluant les pourboires et les commissions?

D41A : Dans ce cas, pourriez-vous nous indiquer votre salaire ou traitement après taxes et autres déductions pour ce travail?

D42 : Quel est/était approximativement votre salaire mensuel à cet emploi, avant impôt et autres retenues à la source, en incluant les pourboires et les commissions?

D42A : Dans ce cas, pourriez-vous nous indiquer votre salaire ou traitement après taxes et autres déductions pour ce travail?

D43 : Quel est/était votre revenu personnel annuel net avant impôt et autres retenues à la source provenant de cette entreprise (c'est-à-dire après avoir enlevé toutes les dépenses d'entreprise?)

Pseudo-logique :

D29, D34, D35, D36, D37, D39, D40, D41, D41A, D42, D42A et D43 servent à calculer les gains annuels tirés de l'emploi principal. Si le répondant a déclaré travailler ou avoir travaillé toutes les semaines du mois, on impute le nombre de semaines travaillées dans cet emploi comme suit : $2*(365,25/(24*7))$.

Gains mensuels tirés de l'emploi principal

Nom de la VD : EARNJOB2

Étiquette de la VD : Gains mensuels tirés de l'emploi principal

Valeurs de l'étiquette : Valeur des gains mensuels dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US), Norvège (couronne norvégienne) et Suisse (franc suisse).*

Question(s) posée(s) : EARNJOB1 : Gains annuels tirés de l'emploi principal

D34 : Pendant combien de mois par année travaillez/travaillez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?

Pseudo-logique :

On divise EARNJOB1 par la réponse obtenue en D34, soit le nombre de mois travaillés dans l'emploi principal.

Gains hebdomadaires tirés de l'emploi principal

Nom de la VD : EARNJOB3

Étiquette de la VD : Gains hebdomadaires tirés de l'emploi principal

Valeurs de l'étiquette : Valeur des gains hebdomadaires dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US), Norvège (couronne norvégienne) et Suisse (franc suisse).*

Question(s) posée(s) : EARNJOB1 : Gains annuels tirés de l'emploi principal

D34 : Combien de mois par année travaillez/travaillez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?

D36 : En moyenne, combien de semaines par mois travaillez/travaillez-vous habituellement à cet emploi ou dans cette entreprise?

Pseudo-logique :

On divise EARNJOB1 par les réponses obtenues en D34 et D36, soit le nombre de mois et de semaines travaillés dans l'emploi principal. Si le répondant a déclaré travailler ou avoir travaillé toutes les semaines du mois, on impute le nombre de semaines travaillées dans l'emploi principal comme suit : $2*(365,25/(24*7))$.

Gains horaires tirés de l'emploi principal

Nom de la VD : EARNJOB3
Étiquette de la VD : Gains horaires tirés de l'emploi principal
Valeurs de l'étiquette : Valeur des gains horaires dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US), Norvège (couronne norvégienne) et Suisse (franc suisse).*

Question(s) posée(s) : EARNJOB1 : Gains annuels tirés de l'emploi principal
D34 : Pendant combien de mois par année travaillez/travailliez-vous habituellement dans le cadre de cet emploi ou de cette entreprise?
D36 : En moyenne, combien de semaines par mois travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?
D37 : En moyenne, combien d'heures par semaine travaillez/travailliez-vous habituellement à cet emploi ou dans cette entreprise?

Pseudo-logique : On divise EARNJOB1 par les réponses obtenues en D34, D36 et D37, soit le nombre de mois, de semaines et d'heures travaillés dans l'emploi principal. Si le répondant a déclaré travailler ou avoir travaillé toutes les semaines du mois, on impute le nombre de semaines travaillées dans l'emploi principal comme suit : $2*(365,25/(24*7))$.

Revenu personnel total du répondant

Nom de la VD : INCTOT1
Étiquette de la VD : Revenu personnel total du répondant
Valeurs de l'étiquette : Valeur du revenu personnel dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US), Norvège (couronne norvégienne) et Suisse (franc suisse).*

Question(s) posée(s) : K6 : Pour <l'année précédant l'entrevue> quelle est votre meilleure estimation de votre revenu personnel de toutes provenances, y compris des sources susmentionnées, avant impôt et retenues?
K7 : Pourriez-vous alors estimer si votre revenu personnel total était de moins de 20 000 \$ ou de 20 000 \$ et plus?
K8 : Était-il de moins de 5 000 \$, de 5 000 \$ à moins de 10 000 \$, de 10 000 \$ à moins de 15 000 \$, de 15 000 \$ ou plus?
K9 : Était-il de moins de 30 000 \$, de 30 000 \$ à moins de 40 000 \$, de 40 000 \$ à moins de 50 000 \$, de 50 000 \$ à moins de 60 000 \$, de 60 000 \$ à moins de 80 000 \$, de 80 000 \$ ou plus?

Pseudo-logique : On inscrit directement le revenu personnel total selon la réponse à la question K6 ou on l'estime d'après la réponse à la question K7, K8 ou K9. Si l'on impute le revenu d'après les intervalles énoncés en K7, K8 ou K9, on choisit le point milieu de l'intervalle. Les utilisateurs de cette VD doivent faire preuve de prudence à l'égard du dernier intervalle (par exemple, 80 000 \$ ou plus dans le cas du Canada), qu'on a plafonné afin d'arriver à un point milieu. Dans le cas du Canada, on a fixé un maximum de 99 999 \$.

Revenu total du ménage du répondant

Nom de la VD : INCTOT2
Étiquette de la VD : Revenu total du ménage du répondant
Valeurs de l'étiquette : Valeur du revenu personnel dans la devise nationale. *Canada (dollar CA), Italie (euro), Bermudes (dollar US), États-Unis (dollar US),*

Norvège (couronne norvégienne) et Suisse (franc suisse).

Question(s) posée(s) : K1 : Y compris vous-même, combien de personnes vivent dans votre ménage?
K11 : Pour <l'année précédant l'entrevue> quelle est votre meilleure estimation du revenu total de toutes provenances de tous les membres du ménage (y compris vous-même) avant impôt et retenues?
K12 : Pourriez-vous alors estimer si le revenu total de toutes les personnes de votre ménage était de moins de 40 000 \$ ou de 40 000 \$ et plus?
K13 : Était-il de moins de 10 000 \$, de 10 000 \$ à moins de 20 000 \$, de 20 000 \$ à moins de 30 000 \$, de 30 000 \$ ou plus?
K14 : Était-il de moins de 50 000 \$, de 50 000 \$ à moins de 60 000 \$, de 60 000 \$ à moins de 80 000 \$, de 80 000 \$ à moins de 100 000 \$, de 100 000 \$ ou plus?

Pseudo-logique : Si le répondant est la seule personne du ménage, le revenu personnel total INCTOT1 est aussi son revenu total du ménage. On inscrit directement le revenu total du ménage selon la réponse à la question K11 ou on l'estime d'après la réponse à la question K12, K13 ou K14. Si l'on impute le revenu d'après les tranches énoncées en K12, K13 ou K14, on choisit le point milieu de la tranche. Les utilisateurs de cette VD doivent faire preuve de prudence à l'égard de la dernière tranche (par exemple, 100 000 \$ ou plus dans le cas du Canada), qu'on a plafonnée afin d'arriver à un point milieu. Dans le cas du Canada, on a fixé un maximum de 119 999 \$.

Revenu personnel total du répondant selon la parité de pouvoir d'achat

Nom de la VD : INCPPP1
Étiquette de la VD : Revenu personnel total du répondant selon la parité de pouvoir d'achat
Valeurs de l'étiquette : Valeur du revenu personnel selon la PPA en \$US (2001)
Question(s) posée(s) : INCTOT1 : Revenu personnel total du répondant
Pseudo-logique : On divise INCTOT1 par la parité de pouvoir d'achat, facteur de conversion correspondant au nombre d'unités d'une devise nécessaires pour acheter sur le marché intérieur la même quantité de biens et de services que le dollar US permettrait d'acheter aux États-Unis. Les facteurs de conversion PPA sont établis par la Banque mondiale (les chiffres les plus récents datent de 2001). Dans le cas des Bermudes, comme on ne connaît pas cette donnée, on suppose que le facteur de conversion PPA est de 1,0.

Revenu total du ménage du répondant selon la parité de pouvoir d'achat

Nom de la VD : INCPPP2
Étiquette de la VD : Revenu total du ménage du répondant selon la parité de pouvoir d'achat
Valeurs de l'étiquette : Valeur du revenu du ménage selon la PPA en dollars US (2001)
Question(s) posée(s) : INCTOT2 : Revenu total du ménage du répondant
Pseudo-logique : On divise INCTOT2 par la parité de pouvoir d'achat, facteur de conversion correspondant au nombre d'unités d'une devise nécessaires pour acheter sur le marché intérieur la même quantité de biens et de services que le dollar US permettrait d'acheter aux États-Unis. Les facteurs de conversion PPA sont établis par la Banque mondiale (les chiffres les plus récents datent de 2001). Dans le cas des Bermudes,

comme on ne connaît pas cette donnée, on suppose que le facteur de conversion PPA est de 1,0.

Soutien du revenu

Nom de la VD : INCSUPP
Étiquette de la VD : Soutien du revenu
Valeurs de l'étiquette : 0 « N'a reçu aucun soutien social du revenu »; 1 « A reçu un soutien social du revenu »
Question(s) posée(s) : K5E : Au cours de <l'année précédant l'interview>, avez-vous touché un revenu provenant de prestations de l'assurance-emploi?
K5F : Au cours de <l'année précédant l'interview>, avez-vous touché un revenu provenant de prestations d'aide sociale et d'un programme provincial de supplément de revenu?
Pseudo-logique : K5E et K5F déterminent si le répondant a reçu un soutien social du revenu au cours de l'année précédant l'enquête.

Variables dérivées complexes

Profession et industrie

Classes d'Erikson et Goldthorpe

Nom de la VD : EGP
Étiquette de la VD : Classes d'Erikson et Goldthorpe
Échelle de la VD : 1 « Travailleur hautement qualifié du secteur des services »; 2 « Travailleur peu qualifié du secteur des services »; 3 « Travailleur manuel »; 4 « Travailleur autonome avec employés »; 5 « Travailleur autonome sans employés »; 7 « Contremaître »; 8 « Travailleur manuel qualifié »; 9 « Travailleur manuel peu qualifié »; 10 « Travailleur agricole »; 11 « Travailleur agricole autonome »
Question(s) posée(s) : ISCOR : Classification internationale type des professions (CITP-88) pour le répondant
Pseudo-logique : ISCO88 est recodée selon les classes d'Erikson et Goldthorpe.

Indice socioéconomique international – Répondant (variable continue)

Nom de la VD : ISEI_R
Étiquette de la VD : Indice socioéconomique international – Répondant (variable continue)
Échelle de la VD : Les notes s'échelonnent de 16 à 90.
Question(s) posée(s) : ISCOR : Profession du répondant (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Indice socioéconomique international.

Indice socioéconomique international – Répondant (quartiles)

Nom de la VD : ISEI_R_Q
Étiquette de la VD : Indice socioéconomique international – Répondant (quartiles)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : ISEI_R : Indice socioéconomique international – Répondant (variable continue)
Pseudo-logique : ISEI_R est recodée en quatre catégories.

Échelle internationale type de prestige des professions – Répondant (variable continue)

Nom de la VD : SIOPS_R
Étiquette de la VD : Échelle internationale type de prestige des professions – Répondant (variable continue)
Échelle de la VD : Les notes s'échelonnent de 6 à 78.
Question(s) posée(s) : ISCOR : Profession du répondant (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Échelle internationale type de prestige des professions.

Échelle internationale type de prestige des professions – Répondant (quartiles)

Nom de la VD : SIOPS_R_Q
Étiquette de la VD : Échelle internationale type de prestige des professions – Répondant (quartile)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : SIOPS_R : Échelle internationale type de prestige des professions – Répondant (variable continue)
Pseudo-logique : SIOPS_R est recodée en quatre catégories.

Indice socioéconomique international – Mère (variable continue)

Nom de la VD : ISEI_M
Étiquette de la VD : Indice socioéconomique international – Mère (variable continue)
Échelle de la VD : Les notes s'échelonnent de 16 à 90.
Question(s) posée(s) : C3_C4 : Emploi de la mère quand le répondant avait 16 ans (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Indice socioéconomique international.

Indice socioéconomique international – Mère (quartiles)

Nom de la VD : ISEI_M_Q
Étiquette de la VD : Indice socioéconomique international – Mère (quartiles)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : ISEI_M : Indice socioéconomique international – Mère (variable continue)
Pseudo-logique : ISEI_M est recodée en quatre catégories.

Échelle internationale type de prestige des professions – Mère (variable continue)

Nom de la VD : SIOPS_M
Étiquette de la VD : Échelle internationale type de prestige des professions – Mère (variable continue)
Échelle de la VD : Les notes s'échelonnent de 6 à 78.
Question(s) posée(s) : C3_C4 : Emploi de la mère quand le répondant avait 16 ans (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Échelle internationale type de prestige des professions.

Échelle internationale type de prestige des professions – Mère (quartiles)

Nom de la VD : SIOPS_M_Q

Étiquette de la VD : Échelle internationale type de prestige des professions – Mère (quartiles)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : SIOPS_M : Échelle internationale type de prestige des professions – Mère (variable continue)
Pseudo-logique : SIOPS_M est recodée en quatre catégories.

Indice socioéconomique international – Père (variable continue)

Nom de la VD : ISEI_F
Étiquette de la VD : Indice socioéconomique international – Père (variable continue)
Échelle de la VD : Les notes s'échelonnent de 16 à 90.
Question(s) posée(s) : C7_C8 : Emploi du père quand le répondant avait 16 ans (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Indice socioéconomique international.

Indice socioéconomique international – Père (quartiles)

Nom de la VD : ISEI_F_Q
Étiquette de la VD : Indice socioéconomique international – Père (quartiles)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : ISEI_F : Indice socioéconomique international – Répondant (variable continue)
Pseudo-logique : ISEI_F est recodée en quatre catégories.

Échelle internationale type de prestige des professions – Père (variable continue)

Nom de la VD : SIOPS_F
Étiquette de la VD : Échelle internationale type de prestige des professions – Père (variable continue)
Échelle de la VD : Les notes s'échelonnent de 6 à 78.
Question(s) posée(s) : C7_C8 : Emploi du père quand le répondant avait 16 ans (Classification internationale type des professions – CITP-88)
Pseudo-logique : ISCO88 est recodée selon l'Échelle internationale type de prestige des professions.

Échelle internationale type de prestige des professions – Père (quartiles)

Nom de la VD : SIOPS_F_Q
Étiquette de la VD : Échelle internationale type de prestige des professions – Père (quartiles)
Échelle de la VD : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : SIOPS_F : Échelle internationale type de prestige des professions – Père (variable continue)
Pseudo-logique : SIOPS_F est recodée en quatre catégories.

Classes de savoir et d'information

Nom de la VD : OCCTYPE
Étiquette de la VD : Classes de savoir et d'information
Échelle de la VD : 1 « Spécialistes »; 2 « Gestionnaires »; 3 « Information, travailleurs hautement qualifiés »; 4 « Information, travailleurs peu qualifiés »; 5 « Services, travailleurs peu qualifiés »; 6 « Biens »

Question(s) posée(s) : ISCOR : Profession du répondant (Classification internationale type des professions – CITP-88)

Pseudo-logique : ISCO88 est recodée selon les classes de savoir et d'information.

Classification axée sur le savoir des industries manufacturières et des services

Nom de la VD : INDTYPE

Étiquette de la VD : Classification axée sur le savoir des industries manufacturières et des services

Échelle de la VD : 1 « Industries manufacturières de haute technologie »; 2 « Industries manufacturières de haute ou moyenne technologie »; 3 « Industries manufacturières de basse ou moyenne technologie »; 4 « Industries manufacturières de basse technologie »; 5 « Services marchands axés sur le savoir »; 6 « Administration publique, défense, santé »; 7 « Autres services communautaires, sociaux et personnels »; 8 « Services publics et construction »; 9 « Commerce de gros et de détail, hôtellerie et restauration »; 10 « Transport et entreposage »; 11 « Industries primaires »

Question(s) posée(s) : ISICR : Industrie du répondant (Classification internationale type par industrie)

Pseudo-logique : ISIC est recodée selon la classification axée sur le savoir des industries manufacturières et des services.

Habitudes de littératie et de numératie au travail

Indice de l'engagement professionnel en matière de lecture (TRI – variable continue)

Nom de la VD : WLIT1

Étiquette de la VD : Indice de l'engagement professionnel en matière de lecture (TRI – variable continue)

Échelle de la VD : Échelle type, avec moyenne internationale et écart-type fixé à (0, 1), adultes âgés de 16 à 65 ans dans les pays suivants : Bermudes, Canada, États-Unis, Italie, Norvège et Suisse.

Question(s) posée(s) : E1 : Dans le cadre de votre emploi principal, à quelle fréquence lisez/lisiez-vous ou utilisez/utilisiez-vous l'information contenue dans les documents ci-dessous? Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?

E1A : Lettres, notes ou messages électroniques

E1B : Rapports, articles, magazines ou revues

E1C : Manuels ou ouvrages de référence, incluant des catalogues

E1D : Diagrammes ou schémas

E1E : Directives ou instructions

E1F : Comptes, factures, feuilles de calcul ou budgets

Pseudo-logique : On calcule cette variable au moyen de techniques d'analyse factorielle confirmatoire (CFA). On spécifie un modèle international qu'on met à l'essai à l'aide du progiciel LISREL (Relations structurelles linéaires). On estime les covariances et les covariances asymptotiques connexes des variables sources à l'aide du progiciel PRELIS et on les importe dans le LISREL. On estime le modèle au moyen de la méthode des moindres carrés pondérés (MCP). On évalue la fiabilité et la validité du modèle en

utilisant les statistiques d'ajustement suivantes : RMSEA, RMR, GFI, AGFI, NFI, NNFI, CFI et RFI. Le modèle international répond aux critères suivants : erreur quadratique moyenne d'approximation (RMSEA) inférieure à 0,08, résidu quadratique moyen (RMR) inférieur à 0,05 et indice corrigé de la qualité de l'ajustement (AGFI) supérieur à 0,90. Dans l'évaluation de l'ajustement, on tolère des covariances parmi les erreurs lorsque la théorie soutient la relation. Puis, on procède à l'échelonnement final des VD selon la méthode de la théorie de la réponse d'item (TRI) à l'aide du progiciel PARSCALE.

Indice de l'engagement professionnel en matière de lecture (TRI – quartiles)

Nom de la VD : WLIT1_Q
Étiquette de la VD : Indice de l'engagement professionnel en matière de lecture (quartiles)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour WLIT1.
Pseudo-logique : WLIT1 est recodée en quatre catégories.

Indice de l'engagement professionnel en matière de lecture (échelle cumulative)

Nom de la VD : E1SS
Étiquette de la VD : Indice de l'engagement professionnel en matière de lecture (échelle cumulative)
Valeurs de l'étiquette : 1 « Jamais »; 2 « Rarement »; 3 « Moins d'une fois par semaine »; « Au moins une fois par semaine »
Question(s) posée(s) : Voir WLIT1.
Pseudo-logique : Les variables sources sont recodées en échelles de catégories ordonnées ascendantes. On divise la somme des valeurs par le nombre de variables sources. Seuls les cas comportant des réponses valides pour toutes les variables sources sont inclus.

Indice de l'engagement professionnel en matière d'écriture (TRI – variable continue)

Nom de la VD : WLIT2
Étiquette de la VD : Indice de l'engagement professionnel en matière d'écriture (TRI – variable continue)
Échelle de la VD : Même que pour WLIT1.
Question(s) posée(s) : E2 : Dans le cadre de votre emploi principal, à quelle fréquence avez/aviez-vous à écrire ou à remplir chacun des documents ci-dessous? Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?
E2A : Lettres, notes ou courriels
E2B : Rapports, articles, magazines ou revues
E2C : Manuels ou ouvrages de référence, incluant des catalogues
E2D : Directives ou instructions
E2E : Comptes, factures, feuilles de calcul ou budgets
Pseudo-logique : Même que pour WLIT1.

Indice de l'engagement professionnel en matière d'écriture (TRI – quartiles)

Nom de la VD : WLIT2_Q
Étiquette de la VD : Indice de l'engagement professionnel en matière d'écriture (quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour WLIT2.
Pseudo-logique : WLIT2 est recodée en quatre catégories.

Indice de l'engagement professionnel en matière d'écriture (échelle cumulative)

Nom de la VD : E2SS
Étiquette de la VD : Indice de l'engagement professionnel en matière d'écriture (échelle cumulative)
Valeurs de l'étiquette : 1 « Jamais »; 2 « Rarement »; 3 « Moins d'une fois par semaine »; 4 « Au moins une fois par semaine »
Question(s) posée(s) : Même(s) que pour WLIT2.
Pseudo-logique : Même que pour E1SS.

Indice de l'engagement professionnel en matière de numératie (TRI – variable continue)

Nom de la VD : WLIT3
Étiquette de la VD : Indice de l'engagement professionnel en matière de numératie (TRI – variable continue)
Échelle de la VD : Même que pour WLIT1.
Question(s) posée(s) : E3 : À quelle fréquence faites/faisiez-vous chacune des activités suivantes dans le cadre de votre emploi principal? Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?
E3B : Calculer des prix, des coûts ou des budgets
E3C : Compter ou lire des chiffres pour vous garder à jour
E3D : Gérer le temps ou établir des horaires
E3E : Donner ou suivre des directives ou utiliser des cartes ou des répertoires de rues
E3F : Utiliser des données statistiques pour parvenir à des conclusions
Pseudo-logique : Même que pour WLIT1.

Indice de l'engagement professionnel en matière de numératie (TRI – quartiles)

Nom de la VD : WLIT3_Q
Étiquette de la VD : Indice de l'engagement professionnel en matière de numératie (quartiles)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour WLIT3.
Pseudo-logique : WLIT3 est recodée en quatre catégories.

Indice de l'engagement professionnel en matière de numératie (échelle cumulative)

Nom de la VD : E3SS
Étiquette de la VD : Indice de l'engagement professionnel en matière de numératie (échelle cumulative)
Valeurs de l'étiquette : 1 « Jamais »; 2 « Rarement »; 3 « Moins d'une fois par semaine »; 4 « Au moins une fois par semaine »
Question(s) posée(s) : E3 : À quelle fréquence faites/faisiez-vous chacune des activités suivantes dans le cadre de votre emploi principal? Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?

- E3A : Mesurer ou estimer la taille ou le poids d'objets
- E3B : Calculer des prix, des coûts ou des budgets
- E3C : Compter ou lire des chiffres pour vous garder à jour
- E3D : Gérer le temps ou établir des horaires
- E3E : Donner ou suivre des directives ou utiliser des cartes ou des répertoires de rues
- E3F : Utiliser des données statistiques pour parvenir à des conclusions

Pseudo-logique : Même que pour E1SS.

Apprentissage non structuré

Indice de la participation à l'apprentissage non structuré (TRI – variable continue)

- Nom de la VD : INFL1
 Étiquette de la VD : Indice de la participation à l'apprentissage non structuré (TRI – variable continue)
 Échelle de la VD : Même que pour WLIT1.
 Question(s) posée(s) : F26 : Au cours des 12 derniers mois, c'est-à-dire depuis <insérez le mois et l'année>, avez-vous participé à l'une ou l'autre des activités d'apprentissage suivantes?
 F26A : Visiter des salons professionnels, aller à des conférences ou à des congrès
 F26B : Assister à des exposés, des séminaires, des ateliers ou des causeries spéciales ne faisant pas partie d'un cours
 F26D : Faire des visites guidées dans des musées, des galeries d'art ou ailleurs
 F26I : Être envoyé(e) dans une organisation pour connaître différents aspects de cette organisation

Pseudo-logique : Même que pour WLIT1.

Indice de la participation à l'apprentissage non structuré (TRI – quartiles)

- Nom de la VD : INFL1_Q
 Étiquette de la VD : Indice de la participation à l'apprentissage non structuré (TRI – quartiles)
 Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
 Question(s) posée(s) : Même(s) que pour INFL1.
 Pseudo-logique : INFL1 est recodée en quatre catégories.

Indice de l'apprentissage par la pratique (TRI – variable continue)

- Nom de la VD : INFL2
 Étiquette de la VD : Indice de l'apprentissage par la pratique (TRI – variable continue)
 Échelle de la VD : Même que pour WLIT1.
 Question(s) posée(s) : F26 : Au cours des 12 derniers mois, c'est-à-dire depuis <insérez le mois et l'année>, avez-vous participé à l'une ou l'autre des activités d'apprentissage suivantes?
 F26E : Apprendre au moyen de l'ordinateur ou d'Internet hors du cadre d'un cours
 F26F : Apprendre au moyen de la télévision, de la vidéo, de cassettes hors du cadre d'un cours

F26G : Apprendre en observant, en obtenant de l'aide ou des conseils d'autres personnes, mais pas d'un professeur

F26H : Apprendre par soi-même en faisant des essais, en s'exerçant, en essayant différentes méthodes de faire les choses

Pseudo-logique : Même que pour WLIT1.

Indice de l'apprentissage par la pratique (TRI – quartiles)

Nom de la VD : INFL2_Q

Étiquette de la VD : Indice de l'apprentissage par la pratique (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour INFL2.

Pseudo-logique : INFL2 est recodée en quatre catégories.

Indice de l'apprentissage non structuré (échelle cumulative)

Nom de la VD : F26SS

Étiquette de la VD : Indice de l'apprentissage non structuré (échelle cumulative)

Valeurs de l'étiquette : 1 « Oui »; 0 « Non »

Question(s) posée(s) : F26 : Au cours des 12 derniers mois, c'est-à-dire depuis <insérez le mois et l'année>, avez-vous participé à l'une ou l'autre des activités d'apprentissage suivantes?

F26A : Visiter des salons professionnels, aller à des conférences ou à des congrès

F26B : Assister à des exposés, des séminaires, des ateliers ou des causeries spéciales ne faisant pas partie d'un cours

F26C : Lire des manuels, des ouvrages de référence, des revues et d'autres documents hors du cadre d'un cours

F26D : Faire des visites guidées dans des musées, des galeries d'art ou ailleurs

F26E : Apprendre au moyen de l'ordinateur ou d'Internet hors du cadre d'un cours

F26F : Apprendre au moyen de la télévision, de la vidéo, de cassettes hors du cadre d'un cours

F26G : Apprendre en observant, en obtenant de l'aide ou des conseils d'autres personnes, mais pas d'un professeur

F26H : Apprendre par soi-même en faisant des essais, en essayant différentes méthodes de faire les choses

F26I : Être envoyé(e) dans une organisation pour connaître différents aspects de cette organisation

Pseudo-logique : Même que pour E1SS.

Habitudes de littératie et de numératie en général

Indice de la fréquence et de la variété de la lecture (TRI – variable continue)

Nom de la VD : HLIT1

Étiquette de la VD : Indice de la fréquence et de la variété de la lecture (TRI – variable continue)

Échelle de la VD : Même que pour WLIT1.

Question(s) posée(s) : G3 : Dans votre vie quotidienne, à quelle fréquence lisez-vous les documents ci-dessous ou utilisez-vous l'information qui y est contenue? Ne tenez pas compte de vos lectures dans le cadre de votre travail ou de vos études. Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?

G3A : À quelle fréquence lisez-vous les journaux ou utilisez-vous l'information qui y est contenue?

G3B : À quelle fréquence lisez-vous des revues ou des articles ou utilisez-vous l'information qui y est contenue?

G3C : À quelle fréquence lisez-vous des livres (de fiction ou autres) ou utilisez-vous l'information qui y est contenue?

G3D : À quelle fréquence lisez-vous des lettres, des notes ou des messages électroniques/courriels ou utilisez-vous l'information qui y est contenue?

F26 : Au cours des 12 derniers mois, c'est-à-dire depuis <insérez le mois et l'année>, avez-vous participé à l'une ou l'autre des activités d'apprentissage suivantes?

F26C : Lire des manuels, des ouvrages de référence, des revues et d'autres documents hors du cadre d'un cours

Pseudo-logique : Même que pour WLIT1.

Indice de la fréquence et de la variété de la lecture (TRI – quartiles)

Nom de la VD : HLIT1_Q

Étiquette de la VD : Indice de la fréquence et de la variété de la lecture (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour HLIT1.

Pseudo-logique : HLIT1 est recodée en quatre catégories.

Indice de la fréquence et de la variété de la lecture (échelle cumulative)

Nom de la VD : G3SS

Étiquette de la VD : Indice de la fréquence et de la variété de la lecture (échelle cumulative)

Valeurs de l'étiquette : 1 « Jamais »; 2 « Rarement »; 3 « Moins d'une fois par semaine »; 4 « Au moins une fois par semaine »

Question(s) posée(s) : G3 : Dans votre vie quotidienne, à quelle fréquence lisez-vous les documents ci-dessous ou utilisez-vous l'information qui y est contenue? Ne tenez pas compte de vos lectures dans le cadre de votre travail ou de vos études. Diriez-vous au moins une fois par semaine, moins d'une fois par semaine, rarement ou jamais?

G3A : À quelle fréquence lisez-vous les journaux ou utilisez-vous l'information qui y est contenue?

G3B : À quelle fréquence lisez-vous des revues ou des articles ou utilisez-vous l'information qui y est contenue?

G3C : À quelle fréquence lisez-vous des livres (de fiction ou autres) ou utilisez-vous l'information qui y est contenue?

G3D : À quelle fréquence lisez-vous des lettres, des notes ou des messages électroniques/courriels ou utilisez-vous l'information qui y est contenue?

Pseudo-logique : Même que pour E1SS.

Indice de l'attitude à l'égard de la lecture (TRI continue)

Nom de la VD : HLIT4
Étiquette de la VD : Indice de l'attitude à l'égard de la lecture (TRI – variable continue)
Échelle de la VD : Même que pour WLIT1.
Question(s) posée(s) : G7 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :
G7C : Je lis seulement quand je suis obligé.
G7D : La lecture est l'une de mes activités préférées.
G7E : J'aime parler de mes lectures avec d'autres personnes.
Pseudo-logique : Même que pour WLIT1.

Indice de l'attitude à l'égard de la lecture (TRI quartiles)

Nom de la VD : HLIT4_Q
Étiquette de la VD : Indice de l'attitude à l'égard de la lecture (TRI – quartiles)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour HLIT4.
Pseudo-logique : HLIT4 est recodée en quatre catégories.

Indice de l'attitude à l'égard de la lecture et des chiffres (échelle cumulative)

Nom de la VD : G7SS
Étiquette de la VD : Indice de l'attitude à l'égard de la lecture (échelle cumulative)
Valeurs de l'étiquette : 1 « Tout à fait en désaccord »; 2 « En désaccord »; 3 « D'accord »; 4 « Tout à fait d'accord »
Question(s) posée(s) : G7 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :
G7A : Je me sens à l'aise avec les chiffres et les calculs.
G7B : Je me sens nerveux (nerveuse) quand je dois calculer des rabais, les taxes de vente ou les pourboires.
G7C : Je lis seulement quand je suis obligé.
G7D : La lecture est l'une de mes activités préférées.
G7E : J'aime parler de mes lectures avec d'autres personnes.
Pseudo-logique : Même que pour E1SS. Veuillez noter qu'au besoin, la matière associée aux variables sources est recodée en catégories ordonnées positivement (plus vaut mieux que moins).

Indice de l'aisance et de la capacité perçue en mathématiques (TRI continue)

Nom de la VD : HLIT5
Étiquette de la VD : Indice de l'aisance et de la capacité perçue en mathématiques (TRI – variable continue)
Échelle de la VD : Même que pour WLIT1.
Question(s) posée(s) : G7 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :
G7A : Je me sens à l'aise avec les chiffres et les calculs.
A9 : Pensez à l'apprentissage des mathématiques et à la façon dont elles vous étaient enseignées quand vous avez fait vos études secondaires de premier ou de deuxième cycle. Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :

- A9A : J'aimais les mathématiques à l'école.
 A9B : J'avais de bonnes notes en mathématiques.
 A9C : Les professeurs allaient trop vite et j'étais souvent perdu(e).
 A9D : Je comprenais habituellement les explications dans les cours de mathématiques.

Pseudo-logique : Même que pour WLIT1.

Indice de l'aisance et de la capacité perçue en mathématiques (TRI quartiles)

- Nom de la VD : HLIT5_Q
 Étiquette de la VD : Indice de l'aisance et de la capacité perçue en mathématiques (TRI – quartiles)
 Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
 Question(s) posée(s) : Même(s) que pour HLIT5.
 Pseudo-logique : HLIT5 est recodée en quatre catégories.

Indice de l'aisance et de la capacité perçue en mathématiques (échelle cumulative)

- Nom de la VD : A9SS
 Étiquette de la VD : Indice de l'aisance et de la capacité perçue en mathématiques (échelle cumulative)
 Valeurs de l'étiquette : 1 « Tout à fait en désaccord »; 2 « En désaccord »; 3 « D'accord »; 4 « Tout à fait d'accord »
 Question(s) posée(s) : A9 : Pensez à l'apprentissage des mathématiques et à la façon dont elles vous étaient enseignées quand vous avez fait vos études secondaires de premier ou de deuxième cycle. Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :
 A9A : J'aimais les mathématiques à l'école.
 A9B : J'avais de bonnes notes en mathématiques.
 A9C : Les professeurs allaient trop vite et j'étais souvent perdu(e).
 A9D : Je comprenais habituellement les explications dans les cours de mathématiques.
 Pseudo-logique : Même que pour E1SS. Veuillez noter qu'au besoin, la matière associée aux variables sources est recodée en catégories ordonnées positivement (plus vaut mieux que moins).

Indice de fréquentation d'une bibliothèque ou d'une librairie (échelle cumulative)

- Nom de la VD : G1SS
 Étiquette de la VD : Indice de fréquentation d'une bibliothèque ou d'une librairie (échelle cumulative)
 Valeurs de l'étiquette : 1 « Jamais »; 2 « Une ou deux fois par année »; 3 « Plusieurs fois par année »; 4 « Chaque mois »; 5 « Chaque semaine »
 Question(s) posée(s) : G1 : À quelle fréquence faites-vous les activités suivantes, que ces activités soient faites en personne ou par ordinateur? Est-ce chaque semaine, chaque mois, plusieurs fois par année, une fois ou deux par année ou jamais?
 G1A : Aller dans une bibliothèque publique
 G1B : Aller dans une librairie

Pseudo-logique : Même que pour E1SS.

Indice de la variété de la lecture d'un journal (échelle cumulative)

Nom de la VD : G5SS

Étiquette de la VD : Indice de la variété de la lecture d'un journal (échelle cumulative)

Valeurs de l'étiquette : 1 « Oui »; 0 « Non »

Question(s) posée(s) : G5 : Je vais maintenant vous lire une liste de certaines rubriques ou sections d'un journal. Veuillez me dire lesquelles vous lisez généralement.

G5A : Nouvelles nationales ou internationales

G5B : Nouvelles régionales ou locales

G5C : Sports

G5D : Chroniques sur la maison, la mode, l'alimentation ou la santé

G5E : Page éditoriale

G5F : Nouvelles financières ou cotes boursières

G5G : Section des livres, du cinéma ou des arts

G5H : Courrier du cœur

Pseudo-logique : Même que pour E1SS.

Technologies de l'information et des communications

Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (TRI – variable continue)

Nom de la VD : ICT1

Étiquette de la VD : Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (TRI – variable continue)

Échelle de la VD : Même que pour WLIT1.

Question(s) posée(s) : H15 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, ni d'accord ni en désaccord, en désaccord ou en total désaccord avec chacun des énoncés suivants :

H15C : Les ordinateurs me permettent d'en faire plus en moins de temps.

H15D : Les ordinateurs me permettent d'obtenir plus facilement de l'information utile.

H15E : Les ordinateurs me permettent d'acquérir de nouvelles connaissances dans des domaines autres que l'informatique.

H15F : Les ordinateurs m'aident à communiquer avec les gens.

H15G : Les ordinateurs m'aident à atteindre mes objectifs de carrière.

Pseudo-logique : Même que pour WLIT1.

Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (TRI – quartiles)

Nom de la VD : ICT1_Q

Étiquette de la VD : Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour ICT1.

Pseudo-logique : ICT1 est recodée en quatre catégories.

Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (échelle cumulative)

- Nom de la VD : H15SS
- Étiquette de la VD : Indice de l'utilité perçue et de l'attitude à l'égard de l'ordinateur (échelle cumulative)
- Valeurs de l'étiquette : 1 « Tout à fait en désaccord »; 2 « En désaccord »; 3 « D'accord »; 4 « Tout à fait d'accord »
- Question(s) posée(s) : H15 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, ni d'accord ni en désaccord, en désaccord ou en total désaccord avec chacun des énoncés suivants :
- H15A : Mon niveau de connaissances en informatique répond à mes besoins actuels.
 - H15B : Je me sens à l'aise lorsque j'installe ou que je mets à jour un logiciel.
 - H15C : Les ordinateurs me permettent d'en faire plus en moins de temps.
 - H15D : Les ordinateurs me permettent d'obtenir plus facilement de l'information utile.
 - H15E : Les ordinateurs me permettent d'acquérir de nouvelles connaissances dans des domaines autres que l'informatique.
 - H15F : Les ordinateurs m'aident à communiquer avec les gens.
 - H15G : Les ordinateurs m'aident à atteindre mes objectifs de carrière.
- Pseudo-logique : Même que pour E1SS.

Indice de la diversité et de l'intensité de l'utilisation d'Internet (TRI – variable continue)

- Nom de la VD : ICT3
- Étiquette de la VD : Indice de la diversité et de l'intensité de l'utilisation d'Internet (TRI – variable continue)
- Échelle de la VD : Même que pour WLIT1.
- Question(s) posée(s) : H5 : Au cours d'un mois typique, à quelle fréquence avez-vous utilisé Internet pour les raisons suivantes? (Tous les jours ou presque, quelques fois par semaine, quelques fois par mois, jamais.)
- H5A : Courrier électronique (courriel)
 - H5B : Participer à des groupes de bavardage ou à d'autres discussions en ligne
 - H5C : Magasinage (y compris regarder les produits et services sans nécessairement acheter)
 - H5D : Transactions financières
 - H5E : Études ou formation officielles (partie d'un cadre structuré de formation comme un cours ou un programme d'études)
 - H5F : Obtenir ou sauvegarder de la musique
 - H5G : Lire les nouvelles ou lire à propos d'événements d'actualité
 - H5H : Chercher des possibilités d'emploi
 - H5I : Chercher de l'information liée à la santé
 - H5J : Chercher de l'information météorologique
 - H5K : Chercher de l'information gouvernementale
 - H5L : Jouer à des jeux avec d'autres personnes
 - H5M : Exploration générale

H5N : Autres raisons, précisez :
H13 : Au cours d'un mois typique, pendant combien d'heures avez-vous utilisé un ordinateur à la maison?
Pseudo-logique : Même que pour WLIT1.

Indice de la diversité et de l'intensité de l'utilisation d'Internet (TRI – quartiles)

Nom de la VD : ICT3_Q
Étiquette de la VD : Indice de la diversité et de l'intensité de l'utilisation d'Internet (TRI – quartiles)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour ICT3.
Pseudo-logique : ICT3 est recodée en quatre catégories.

Indice de la diversité et de l'intensité de l'utilisation d'Internet (échelle cumulative)

Nom de la VD : H5SS
Étiquette de la VD : Indice de la diversité et de l'intensité de l'utilisation d'Internet (échelle cumulative)
Valeurs de l'étiquette : 1 « Jamais »; 2 « Quelques fois par mois »; 3 « Quelques fois par semaine »; 4 « Tous les jours ou presque »
Question(s) posée(s) : H5 : Au cours d'un mois typique, à quelle fréquence avez-vous utilisé Internet pour les raisons suivantes? (Tous les jours ou presque, quelques fois par semaine, quelques fois par mois, jamais.)
H5A : Courrier électronique (courriel)
H5B : Participer à des groupes de bavardage ou à d'autres discussions en ligne
H5C : Magasinage (y compris regarder les produits et services sans nécessairement acheter)
H5D : Transactions financières
H5E : Études ou formation officielles (partie d'un cadre structuré de formation comme un cours ou un programme d'études)
H5F : Obtenir ou sauvegarder de la musique
H5G : Lire les nouvelles ou lire à propos d'événements d'actualité
H5H : Chercher des possibilités d'emploi
H5I : Chercher de l'information liée à la santé
H5J : Chercher de l'information météorologique
H5K : Chercher de l'information gouvernementale
H5L : Jouer à des jeux avec d'autres personnes
H5M : Exploration générale
H5N : Autres raisons, précisez :
Pseudo-logique : Même que pour E1SS.

Indice de l'utilisation de l'ordinateur pour certaines tâches (TRI – variable continue)

Nom de la VD : ICT6
Étiquette de la VD : Indice de l'utilisation de l'ordinateur pour certaines tâches (TRI – variable continue)
Échelle de la VD : Même que pour WLIT1.
Question(s) posée(s) : H3 : Au cours d'un mois typique, à quelle fréquence avez-vous utilisé un ordinateur pour les raisons suivantes? (Tous les jours ou presque, quelques fois par semaine, quelques fois par mois, jamais.)

H3A : Rédiger ou éditer du texte
 H3B : Comptes, feuilles de calcul ou analyses statistiques
 H3C : Créer des éléments graphiques, des dessins, des illustrations ou des présentations
 H3D : Faire de la programmation ou écrire du code machine
 H3E : Tenir un calendrier ou un échéancier
 H3F : Lire de l'information sur CD-ROM ou DVD
 H13 : Au cours d'un mois typique, pendant combien d'heures avez-vous utilisé un ordinateur à la maison?
 Pseudo-logique : Même que pour WLIT1.

Indice de l'utilisation de l'ordinateur pour certaines tâches (TRI – quartiles)

Nom de la VD : ICT6_Q
 Étiquette de la VD : Indice de l'utilisation de l'ordinateur pour certaines tâches (TRI – quartiles)
 Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
 Question(s) posée(s) : Même(s) que pour ICT6.
 Pseudo-logique : ICT6 est recodée en quatre catégories.

Indice de l'utilisation de l'ordinateur pour certaines tâches (échelle cumulative)

Nom de la VD : H3SS
 Étiquette de la VD : Indice de l'utilisation de l'ordinateur pour certaines tâches (échelle cumulative)
 Valeurs de l'étiquette : 1 « Jamais »; 2 « Quelques fois par mois »; 3 « Quelques fois par semaine »; 4 « Tous les jours ou presque »
 Question(s) posée(s) : H3 : Au cours d'un mois typique, à quelle fréquence avez-vous utilisé un ordinateur pour les raisons suivantes? (Tous les jours ou presque, quelques fois par semaine, quelques fois par mois, jamais.)
 H3A : Rédiger ou éditer du texte
 H3B : Comptes, feuilles de calcul ou analyses statistiques
 H3C : Créer des éléments graphiques, des dessins, des illustrations ou des présentations
 H3D : Faire de la programmation ou écrire du code machine
 H3E : Tenir un calendrier ou un échéancier
 H3F : Lire de l'information sur CD-ROM ou DVD
 H3G : Jouer à des jeux
 H3H : Autres raisons, précisez :
 Pseudo-logique : Même que pour E1SS.

Indice de l'intensité et de la variété de l'utilisation de la technologie en général (échelle cumulative)

Nom de la VD : H1SS
 Étiquette de la VD : Indice de l'intensité et de la variété de l'utilisation de la technologie en général (échelle cumulative)
 Valeurs de l'étiquette : 1 « Jamais »; 2 « Quelques fois par mois »; 3 « Quelques fois par semaine »; 4 « Tous les jours ou presque »
 Question(s) posée(s) : H1 : Au cours d'un mois typique, à quelle fréquence avez-vous utilisé chacun des objets suivants?
 H1A : Un téléphone cellulaire

- H1B : Une calculatrice
- H1C : Un télécopieur (un appareil autonome et non un logiciel de télécopie dans un ordinateur)
- H1D : Un téléphone à clavier pour faire des achats, payer une facture ou faire des transactions financières
- H1E : Un guichet automatique bancaire

Pseudo-logique : Même que pour E1SS.

Indice de l'intensité et de la variété de l'utilisation de la technologie au travail (échelle cumulative)

- Nom de la VD : H10SS
 Étiquette de la VD : Indice de l'intensité et de la variété de l'utilisation de la technologie au travail (échelle cumulative)
 Valeurs de l'étiquette : 1 « Oui »; 0 « Non »
 Question(s) posée(s) : H10 : Au cours d'un mois typique, avez-vous utilisé une des technologies suivantes à votre travail?
 H10A : Calculatrice
 H10B : Téléphone cellulaire
 H10C : Télécopieur
 H10D : Agenda électronique
- Pseudo-logique : Même que pour E1SS.

Indice de l'utilisation de l'ordinateur à plusieurs endroits (échelle cumulative)

- Nom de la VD : H14SS
 Étiquette de la VD : Indice de l'utilisation de l'ordinateur à plusieurs endroits (échelle cumulative)
 Valeurs de l'étiquette : 1 « Oui »; 0 « Non »
 Question(s) posée(s) : H14 : Au cours d'un mois typique, avez-vous utilisé un ordinateur...
 H14A : chez un(e) ami(e)
 H14B : chez un membre de la famille
 H14C : dans une bibliothèque municipale (ne pas inclure la consultation du catalogue de la bibliothèque)
 H14D : dans un café Internet
 H14E : dans un centre de ressources communautaires (ex. : un centre d'emploi)
 H14F : dans un établissement d'enseignement ou de formation
 H14G : dans un autre endroit, précisez
- Pseudo-logique : Même que pour E1SS.

Participation à la vie sociale

Indice de la participation aux activités de groupes ou d'organisations communautaires (TRI – variable continue)

- Nom de la VD : CIVI1
 Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires (TRI – variable continue)
 Échelle de la VD : Même que pour WLIT1.
 Question(s) posée(s) : G8 : Les questions qui suivent portent sur vos activités bénévoles et les organisations dont vous faites partie. Au cours des 12 derniers mois,

avez-vous fait partie de l'un ou l'autre des groupes ou des organisations énumérés ci-dessous?

- G8A : Une organisation politique
- G8B : Une organisation sportive ou récréative (ex. : une ligue de baseball, un club de tennis, etc.)
- G8C : Un groupe à caractère culturel, éducatif ou récréatif (ex. : une troupe de théâtre, un club de lecture, un club de cartes, etc.)
- G8D : Une organisation à but non lucratif (ex. : club Kiwanis, Chevaliers de Colomb, Shriners, etc.)
- G8E : Une association de quartier, de citoyens ou communautaire ou une association scolaire (ex. : association de parents/de professeurs, votre association communautaire de quartier)

Pseudo-logique : Même que pour WLIT1.

Indice de la participation aux activités de groupes ou d'organisations communautaires (TRI – quartiles)

Nom de la VD : CIVI1_Q
Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires (TRI – quartiles)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : Même(s) que pour CIVI1.
Pseudo-logique : CIVI1 est recodée en quatre catégories.

Indice de la participation aux activités de groupes ou d'organisations communautaires (échelle cumulative)

Nom de la VD : G8SS
Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires (échelle cumulative)
Valeurs de l'étiquette : 1 « Oui »; 0 « Non »
Question(s) posée(s) : G8 : Les questions qui suivent portent sur vos activités bénévoles et les organisations dont vous faites partie. Au cours des 12 derniers mois, avez-vous fait partie de l'un ou l'autre des groupes ou des organisations énumérés ci-dessous?
G8A : Une organisation politique
G8B : Une organisation sportive ou récréative (ex. : une ligue de baseball, un club de tennis, etc.)
G8C : Un groupe à caractère culturel, éducatif ou récréatif (ex. : une troupe de théâtre, un club de lecture, un club de cartes, etc.)
G8D : Une organisation à but non lucratif (ex. : club Kiwanis, Chevaliers de Colomb, Shriners, etc.)
G8E : Une association de quartier, de citoyens ou communautaire ou une association scolaire (ex. : association de parents/de professeurs, votre association communautaire de quartier)
G8F : Un groupe affilié à une organisation à caractère spirituel (ex. : un groupe de jeunes affilié à une église)
G8G : Tout autre groupe ou organisation

Pseudo-logique : Même que pour E1SS.

Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (TRI – variable continue)

Nom de la VD : CIVI2

Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (TRI – variable continue)

Échelle de la VD : Même que pour WLIT1.

Question(s) posée(s) : G9 : Au cours des 12 derniers mois, avez-vous participé à l'une ou l'autre des activités suivantes en tant que BÉNÉVOLE au sein d'un groupe ou d'une organisation?

G9A : Campagne de financement

G9B : Être membre d'un conseil d'administration sans rémunération

G9C : Enseigner, donner un entraînement ou donner des conseils gratuitement

G9D : Recueillir de la nourriture et d'autres produits pour des œuvres de charité

G9E : D'autres activités telles qu'organiser ou superviser des activités, faire du travail de bureau ou fournir des renseignements pour le compte d'une organisation

Pseudo-logique : Même que pour WLIT1.

Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (TRI – quartiles)

Nom de la VD : CIVI2_Q

Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour CIVI2.

Pseudo-logique : CIVI2 est recodée en quatre catégories.

Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (échelle cumulative)

Nom de la VD : G9SS

Étiquette de la VD : Indice de la participation aux activités de groupes ou d'organisations communautaires en tant que bénévole (échelle cumulative)

Valeurs de l'étiquette : 1 « Oui »; 0 « Non »

Question(s) posée(s) : Même(s) que pour CIVI2.

Pseudo-logique : Même que pour E1SS.

Santé et bien-être

Indice de la satisfaction et du bien-être en général (TRI – variable continue)

Nom de la VD : WELL1

Étiquette de la VD : Indice de la satisfaction et du bien-être en général (TRI – variable continue)

Valeurs de l'étiquette : Mêmes que pour WLIT1.

Question(s) posée(s) : G10 : Dans l'ensemble, quel bilan faites-vous de votre vie au cours des 12 derniers mois? Diriez-vous que vous êtes extrêmement satisfait(e), satisfait(e), insatisfait(e) ou extrêmement insatisfait(e)?

G16 : Les questions suivantes portent sur la façon dont vous vous sentiez et dont les choses se sont passées pour vous au cours des 4 dernières semaines. Pour chaque question, veuillez donner la réponse qui se rapproche le plus de ce que vous ressentiez. (Tout le temps, la plupart du temps, une bonne partie du temps, parfois, rarement, jamais.)
Au cours des 4 dernières semaines, dans quelle proportion du temps...

G16A : Vous sentiez-vous calme et paisible?

G16B : Aviez-vous beaucoup d'énergie?

G16C : Vous sentiez-vous abattu(e) et triste?

Pseudo-logique : Même que pour WLIT1.

Indice de la satisfaction et du bien-être en général (TRI – quartiles)

Nom de la VD : WELL1_Q

Étiquette de la VD : Indice de la satisfaction et du bien-être en général (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour WELL1.

Pseudo-logique : WELL1 est recodée en quatre catégories.

Indice de l'incidence de la santé physique sur le travail (TRI – variable continue)

Nom de la VD : WELL2

Étiquette de la VD : Indice de l'incidence de la santé physique sur le travail (TRI – variable continue)

Valeurs de l'étiquette : Mêmes que pour WLIT1.

Question(s) posée(s) : G13 : Au cours des 4 dernières semaines, avez-vous éprouvé les problèmes suivants au travail ou dans d'autres activités quotidiennes en raison de votre santé physique?

G13A : Vous en avez accompli moins que vous auriez voulu.

G13B : Vous étiez limité(e) dans la sorte de travail que vous pouviez faire ou dans d'autres activités.

G15 : Au cours des 4 dernières semaines, dans quelle mesure la douleur a-t-elle nui à votre travail (tant à la maison qu'à l'extérieur)? (Pas du tout, un peu, moyennement, passablement, extrêmement)

Pseudo-logique : Même que pour WLIT1.

Indice de l'incidence de la santé physique sur le travail (TRI – quartiles)

Nom de la VD : WELL2_Q

Étiquette de la VD : Indice de l'incidence de la santé physique sur le travail (TRI – quartiles)

Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »

Question(s) posée(s) : Même(s) que pour WELL2.

Pseudo-logique : WELL2 est recodée en quatre catégories.

Résumé de la composante physique – indice de santé SF-12 (variable continue)

Nom de la VD : PCS

Étiquette de la VD : Résumé de la composante physique – indice de santé SF-12 (variable continue)

Valeurs de l'étiquette : L'échelle PCS va de 9,94738 à 70,13284.

Question(s) posée(s) : G11 : De façon générale, diriez-vous que votre santé est excellente, très bonne, bonne, passable, mauvaise?
 G12 : Les questions qui suivent portent sur les activités que vous pouvez accomplir au cours d'une journée typique. Est-ce que votre santé vous limite actuellement dans les activités suivantes? Si oui, dans quelle mesure – beaucoup ou un peu?
 G12A : Des activités modérées comme déplacer une table, passer l'aspirateur, jouer aux quilles ou au golf
 G12B : Monter plusieurs escaliers
 G13 : Au cours des 4 dernières semaines, avez-vous éprouvé les problèmes suivants au travail ou dans d'autres activités quotidiennes en raison de votre santé physique?
 G13A : Vous en avez accompli moins que vous auriez voulu.
 G13B : Vous étiez limité(e) dans la sorte de travail que vous pouviez faire ou dans d'autres activités.
 G14 : Au cours des 4 dernières semaines, avez-vous éprouvé les problèmes suivants au travail ou dans vos activités quotidiennes en raison de problèmes émotionnels (dépression, anxiété)?
 G14A : Vous en avez accompli moins que vous auriez voulu.
 G14B : Vous n'avez pas accompli votre travail ou vos autres activités aussi soigneusement que d'habitude.
 G15 : Au cours des 4 dernières semaines, dans quelle mesure la douleur a-t-elle nui à votre travail (tant à la maison qu'à l'extérieur)? (Pas du tout, un peu, moyennement, passablement, extrêmement).
 G16 : Les questions suivantes portent sur la façon dont vous vous sentiez et dont les choses se sont passées pour vous au cours des 4 dernières semaines. Pour chaque question, veuillez donner la réponse qui se rapproche le plus de ce que vous ressentiez. (Tout le temps, la plupart du temps, une bonne partie du temps, parfois, rarement, jamais.)
 Au cours des 4 dernières semaines, dans quelle proportion du temps...
 G16A : Vous sentiez-vous calme et paisible?
 G16B : Aviez-vous beaucoup d'énergie?
 G16C : Vous sentiez-vous abattu(e) et triste?
 G17 : Au cours des 4 dernières semaines, dans quelle mesure votre santé physique ou vos problèmes émotionnels ont-ils nui à vos activités sociales (comme visiter des amis, des parents, etc.)? Tout le temps, la plupart du temps, parfois, rarement, jamais.

Pseudo-logique : Dans la première version de l'enquête sur la santé SF-12, la constante (1990) pour le poids physique est 56,57706. Chaque variable comprend une ou plusieurs catégories; on attribue à chaque catégorie un poids physique dont la valeur peut être positive ou négative. Les poids physiques associés à la catégorie choisie dans chaque variable sont ajoutés à la constante (1990) pour calculer l'échelle PCS finale.

Résumé de la composante physique – indice de santé SF-12 (quartiles)

Nom de la VD : PCS_Q
 Étiquette de la VD : Résumé de la composante physique – indice de santé SF-12 (quartile)
 Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
 Question(s) posée(s) : PCS : Résumé de la composante physique – indice de santé SF-12 (variable continue)

Pseudo-logique : PCS est recodée en quatre catégories.

Résumé de la composante mentale – indice de santé SF-12 (variable continue)

Nom de la VD : MCS
Étiquette de la VD : Résumé de la composante mentale – indice de santé SF-12 (variable continue)
Valeurs de l'étiquette : L'échelle MCS va de 5,89058 à 72,28379.
Question(s) posée(s) : Même(s) que pour PCS.
Pseudo-logique : Dans la première version de l'enquête sur la santé SF-12, la constante (1990) pour le poids physique est 60,75781. Chaque variable comprend une ou plusieurs catégories; on attribue à chaque catégorie un poids physique dont la valeur peut être positive ou négative. Les poids physiques associés à la catégorie choisie dans chaque variable sont ajoutés à la constante (1990) pour calculer l'échelle MCS finale.

Résumé de la composante mentale – indice de santé SF-12 (quartiles)

Nom de la VD : MCS_Q
Étiquette de la VD : Résumé de la composante mentale – indice de santé SF-12 (quartile)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : MCS : Résumé de la composante mentale – indice de santé SF-12 (variable continue)
Pseudo-logique : MCS est recodée en quatre catégories.

Auto-évaluation

Indice de satisfaction des compétences de travail (TRI – variable continue)

Nom de la VD : SELF1
Étiquette de la VD : Indice de satisfaction des compétences de travail (TRI – variable continue)
Valeurs de l'étiquette : Mêmes que pour WLIT1.
Question(s) posée(s) : E4 : Veuillez me dire si vous êtes tout à fait d'accord, d'accord, en désaccord ou tout à fait en désaccord avec les énoncés suivants :
E4A : Je possède les capacités de lecture en <insérez la langue> dont j'ai besoin pour bien faire mon travail dans le cadre de mon emploi principal.
E4B : Je possède les capacités d'écriture en <insérez la langue> dont j'ai besoin pour bien faire mon travail dans le cadre de mon emploi principal.
E4C : Je possède les compétences en mathématiques dont j'ai besoin pour bien faire mon travail dans le cadre de mon emploi principal.
Pseudo-logique : Même que pour WLIT1.

Indice de satisfaction des compétences de travail (TRI – quartile)

Nom de la VD : SELF1_Q
Étiquette de la VD : Indice de satisfaction des compétences de travail (TRI – quartile)
Valeurs de l'étiquette : 1 « Quartile 1 »; 2 « Quartile 2 »; 3 « Quartile 3 »; 4 « Quartile 4 »
Question(s) posée(s) : SELF1 : Indice de satisfaction des compétences de travail (TRI – variable continue)

Pseudo-logique : SELF1 est recodée en quatre catégories.

Indice de satisfaction des compétences de travail (échelle cumulative)

Nom de la VD : E4SS

Étiquette de la VD : Indice de satisfaction des compétences de travail (échelle cumulative)

Valeurs de l'étiquette : 1 « Tout à fait en désaccord »; 2 « En désaccord »; 3 « D'accord »; 4 « Tout à fait d'accord »

Question(s) posée(s) : Même(s) que pour SELF1.

Pseudo-logique : Même que pour E1SS.

6.4.7 Création du fichier principal (étape 7)

Le fichier principal est un fichier final épuré contenant toutes les variables nécessaires, à partir duquel on peut créer le FMGD et le fichier international de microdonnées. Pour créer ce fichier, il a fallu soumettre les données à une série de fonctions : recodage, renommage, couplage, etc. Bon nombre de questions non résolues aux étapes antérieures à cause de vérifications des comptes de fréquence, des valeurs manquantes ou des valeurs aberrantes ont été réglées à cette étape, qui a suivi directement celle de l'imputation.

Renommage des variables

Afin d'assurer la compatibilité avec les normes internationales, on a renommé les variables utilisées dans le questionnaire complet. Dans chaque variable, on a retiré la lettre 'Q' au début (exemple : QA1 est devenue A1). On n'a pas modifié les variables créées en réponse aux exigences internationales ni celles qui n'avaient aucune incidence sur le fichier international. Certaines variables provenant de sources différentes et ajoutées au fichier ont été renommées parce que le nom de la variable était trop long; par exemple, on a remplacé « imp_tqlang2 » par « itlang2 », « fforestat » par « fforest », « fmainstat » par « fmainst », « finalstat » par « finalst », « taskstatus » par « Taskstat », « MotherTongue » par « MotherTo », « Immigrant » par « Immigrnt », « Aboriginal » par « Aborig », etc. (la longueur normale étant de huit octets). S'il fallait créer des variables indicatrices ou des variables d'imputation, on a alors créé la convention d'appellation en conséquence. Par exemple, on a remplacé les variables indicatrices comme « BQStatus » par « BQStFlag » et la variable d'imputation « imp_age » par « iageflag ». On a aussi renommé d'autres variables en fonction des normes de collecte de données; par exemple, « BQLP_Q01 » est devenue « LP_Q01 » et « Primary » est devenue « SampleID ».

Nouvelles variables

On a créé plusieurs nouvelles variables pour ajouter des renseignements au fichier existant ainsi que pour inclure des renseignements nécessaires au fichier international. On en a créé certaines à partir des variables existantes, d'autres en fusionnant différents fichiers. Outre l'état final, la Méthodologie a créé de nombreuses nouvelles variables nécessaires aux fichiers national et international. Elles sont décrites ci-dessous, mais il convient de signaler que la majorité des variables ajoutées à cette étape étaient de nature administrative et qu'elles ne font pas partie du fichier de microdonnées à grande diffusion.

Fusion des ensembles de données

L'équipe d'enquête a créé les variables suivantes : « Rural/Urban » (indicateur de région urbaine ou rurale), « Stratum » (indicateur géographique), « DA » (aire de diffusion), qui étaient

également nécessaires aux deux fichiers, national et international. Elle a également imputé les valeurs manquantes pour les variables suivantes destinées au fichier national : province, langue maternelle, statut d'immigrant (né au Canada), statut d'Autochtone. On a fusionné ces nouvelles variables à notre fichier principal existant, les ajoutant ainsi à la liste de variables.

Fusion de l'ensemble de données IRLFLAGS

Pour répondre à des besoins méthodologiques, on a créé plusieurs variables qu'on a conservées dans un fichier d'indicateurs appelé IRLFLAGS (abréviation de *International Record Layout Flags*). Par exemple, comme on disposait de renseignements géographiques détaillés, on a défini les régions rurales et urbaines et créé une variable fondée sur les données du recensement concernant la taille de la localité de chaque ménage répondant.

La définition de la variable UARatype (région urbaine ou rurale) provient directement du recensement, alors qu'on a recréé la variable RUFLAG (indicateur de région rurale) pour condenser la subdivision à l'égard de la densité de la population. On a également créé des états finaux pour le questionnaire de base, les tâches de base et les tâches principales à partir des renseignements provenant d'une combinaison de variables. Pour créer les états finaux des tâches de base et des tâches principales, la Méthodologie a utilisé une combinaison de deux fichiers, soit core1.txt (variable=CS_Q02) et out_core.txt (variable=CSTAT) pour les tâches de base et main.txt (variable=MT_Q02) et out_main.txt (variable=MSTAT) pour les tâches principales.

La Méthodologie a déterminé le code d'état du questionnaire de base selon le nombre de réponses complètes, partielles et incomplètes à chaque module.

Données complètes (23 098 enregistrements) : Au moins 75 % des modules devaient être complets dans une proportion d'au moins 75 % plus le sexe, l'âge et le niveau de scolarité ont été déclarés

Données partielles (179 enregistrements) : Au moins 50 % des modules devaient être complets dans une proportion de 75 % plus le sexe, l'âge et le niveau de scolarité ont été déclarés.

Données incomplètes (472 enregistrements) : Si moins de 50 % des modules étaient complets dans une proportion de 75 % ou s'il manquait l'âge, le sexe ou la scolarité (niveau ou nombre d'années), la réponse était considérée incomplète.

On a établi le code d'état en le codant comme suit : 2 => données complètes, 1 => données partielles, 0 => données incomplètes), selon les données du questionnaire de base.

Fusion de l'ensemble de données DV (variables dérivées)

On a posé certaines questions dans le cadre de la composante Entrée de l'enquête et d'autres dans le cadre du questionnaire de base. Lorsque le répondant et la première personne jointe (Questionnaire de présélection) étaient la même personne, on n'a pas répété la question; sinon, on l'a posée de nouveau. L'équipe d'enquête a combiné les deux réponses provenant des fichiers distincts en une question commune qui contenait une seule réponse par enregistrement. L'ensemble de données DV (variables dérivées) a donc permis de fusionner les variables langue maternelle, statut d'immigrant et statut d'Autochtone. On a mis au point ce genre de questions pour réduire le fardeau du répondant.

Fusion de l'ensemble de données TQLANG2

On a créé TQLANG2 (indicateur de pays) selon le profil linguistique de chaque répondant, en vue de l'utiliser plus tard dans le fichier international sous le nom de CNTRID. Les renseignements nécessaires à TQLANG2 ont été tirés de TQLANG (langue du livret de tâches). L'équipe d'enquête a imputé les valeurs manquantes en utilisant un enregistrement donneur présentant un profil linguistique semblable. On a également créé un indicateur appelé ITQLANG2 pour indiquer si l'enregistrement avait été imputé ou non.

Fusion du fichier DA (aire de diffusion)

Afin de faciliter la création d'estimations régionales de la littératie, l'équipe d'enquête a ajouté l'aire de diffusion (Dissemination Area ou DA) du ménage selon les structures définies par le recensement.

Fusion de l'ensemble de poids

L'équipe d'enquête a également créé une série de poids nécessaires pour évaluer la qualité de l'enquête et pour permettre aux utilisateurs de calculer exactement l'erreur-type des estimations des compétences. Les variables annexées à cette étape sont les suivantes :

THEORWT : Poids théorique. Il correspond à l'inverse de la probabilité de sélection. Il est fourni par le plan d'échantillonnage. On a utilisé ce poids pour l'analyse de la qualité.

SUBWT : Sous-poids. C'est le produit du poids théorique (THEORWT) par le facteur de correction de la non-réponse.

POPWT : Poids de population. On utilise cette variable pour donner à l'échantillon la possibilité de représenter la population canadienne âgée de 16 ans et plus. Elle est le produit du sous-poids (SUBWT) par le facteur de correction en fonction de totaux connus. Il s'agit du poids d'échantillonnage global de la personne. La somme de ces poids à l'intérieur d'un échantillon donne la taille de la population.

REPLIC01 à REPLIC30 : Série de 30 poids de rééchantillonnage qu'on applique aux cinq valeurs plausibles de compétence (textes suivis, textes schématiques, numérotation, santé ou résolution de problèmes) pour produire des estimations ponctuelles exactes de la population avec des erreurs-types calculables. La méthode d'utilisation de ces variables est décrite au chapitre 8 (Qualité des données).

Le poids de population et tous les poids de rééchantillonnage sont compris dans le fichier de microdonnées à grande diffusion.

Fusion du fichier Casevnt

L'unité de traitement a également créé de nouvelles variables à partir d'une variable existante ou en combinant certaines variables. On a fusionné le fichier Casevnt afin de créer plusieurs variables comme C01 (nombre total de visites au ménage par l'intervieweur) et C07_1 à C07_10 (état de ces visites : pas de réponse, refus, répondant déménagé, répondant interviewé, etc.). On a calculé les valeurs de C07_1 à C07_10 en utilisant certaines valeurs de APLCSTAT, alors qu'on a créé BrnInCan en utilisant deux variables, soit A1 et Imm_Q01.

Fusion du fichier Times

On a également produit l'heure de fin et l'heure de début de différentes composantes de l'enquête à partir du fichier Times produit par l'application IAO. Il s'agit des dates et des heures de début et de fin de chaque composante de l'enquête (questionnaire de base, livret des tâches de base et livret des tâches principales). À l'instar du fichier CASEVNT, le fichier « Times » comporte plusieurs enregistrements pour le même ménage; dans le cas d'un appariement identique, on a donc fusionné les renseignements nécessaires tirés de l'enregistrement au fichier principal.

Fusion du fichier COREFIN

On a saisi les variables CS_Q02 (détail de l'état physique et émotif), SCORE (note totale calculée à la suite des tâches de base), C_QA1 (langue des tâches de base) et OUTCORE (état final du livret des tâches de base). On a fusionné certains renseignements tirés de ce fichier pour créer de nouvelles variables, alors qu'on a importé d'autres variables brutes telles quelles afin d'ajouter plus de renseignements au fichier existant.

Fusion du fichier MAINFIN

On a fusionné au fichier principal les variables MT_R29 (utilisation d'une calculatrice), MS_Q02 (détail de l'état physique et émotif au moment de remplir le livret de tâches principales) et OUTMAIN (état final du livret des tâches principales). Certaines variables de ce fichier ont servi à créer de nouvelles variables comme CBLOCK5, CBLOCK6, CBLOCK56 à partir de MT_R29, alors que d'autres ont simplement servi à ajouter plus de renseignements au fichier existant.

Fusion de l'ensemble de données Bookflag

En vérifiant les réponses valides et invalides dans les livrets de tâches principales, on a remarqué dans certains livrets des marques inscrites par les répondants, peut-être à la suggestion erronée de l'intervieweur, pour montrer qu'ils avaient lu la question et tenté d'y répondre. On a créé un fichier pour identifier chacun de ces cas en créant un indicateur de 1 (marques dans les livrets) ou de 2 (directives fournies aux intervieweurs). On a fusionné ce fichier au fichier principal pour ajouter la variable « BookFlag ».

Fusion des fichiers CanadaScore et B8BQ3D et de l'ensemble de données corescore

On a créé des fichiers distincts pour les notes des tâches de base (COREQ1S1 à COREQ6S2) en attribuant à chaque question une note de 1 si la réponse était bonne et de 0 si elle était mauvaise. Comme on a vérifié chaque question deux fois, on a attribué deux notes par question : la note 1 pour la première note (S1) et la note 2 (S2) pour la deuxième vérification. Il y avait en tout six questions. Lorsque le total des notes était égal ou supérieur à 3, on produisait la composante livret de tâches principales pour permettre aux répondants de répondre à des questions plus détaillées en matière de littérature et de numératie. On a créé dans ce fichier une variable appelée « SCORE » pour additionner les notes des six questions. On a créé deux variables, B8BQ3DS1 (note 1) et B8BQ3DS2 (note 2) pour tenir compte d'une erreur trouvée dans l'un des items en français (à l'égard d'aspects précis concernant l'ensemble de données canadiennes, veuillez consulter la section sur la qualité).

Fusion du fichier BckGrn

Les valeurs initiales de certaines variables ont été recodées à l'étape antérieure. Afin de saisir de nouveau les valeurs initiales, on a fusionné le fichier BckGrn en conservant uniquement les variables nécessaires. Ces variables étaient QA1, BQIQ01 et BQLP_Q01. La variable QA1 a permis d'attribuer de nouveau les valeurs initiales à la question A1 (« Êtes-vous né au Canada? »). On a également utilisé cette question pour créer la variable BrnInCan (né au Canada) à partir des questions BQIQ01 (dans le questionnaire de base) et IMM_Q01 (dans le module Entrée). La variable BQLP_Q01 (langue de l'interview) a permis de rétablir la valeur de LP_Q01. On a également choisi BQIQ01 pour rétablir les valeurs initiales. Toutefois, lorsque les valeurs de cette variable étaient manquantes, on a inscrit « 6 » (enchaînement valide).

Fusion du fichier Entry (entrée)

On a fusionné le fichier Entry (entrée) parce que, pour une variable essentielle (BOOKID) servant à indiquer lequel des 28 livrets avait effectivement été administré au répondant, il manquait une valeur dans cinq enregistrements. Pour cette raison, on a retiré du fichier Entry la variable « BookLtID » du fichier échantillon, indiquant lequel des 28 livrets avait été attribué à chaque répondant, et l'on a utilisé la valeur de cette variable pour imputer une valeur à BOOKID.

Fusion des notes sur les connaissances en matière de santé

À la suite de travaux menés par Educational Testing Service, on a découvert que les domaines de la compréhension de textes suivis et de textes schématiques et de la numération présentaient suffisamment de renseignements liés à la santé pour produire une quatrième mesure psychométrique appelée Connaissances en matière de santé. En pratique, ces tâches mesurent la capacité d'une personne d'utiliser et de comprendre des textes et des documents contenant de renseignements liés à la santé. À cette étape, afin d'ajouter de la valeur à l'ensemble de données, on a calculé les cinq valeurs plausibles liées à ce domaine supplémentaire pour les données canadiennes de l'EIACA et on les a fusionnées à l'ensemble de données.

Création des variables Education (scolarité)

On a également créé les variables Education (scolarité) pour le fichier international en utilisant une combinaison de variables.

On a ainsi créé A4_ISCED, soit le niveau de scolarité le plus élevé atteint par le répondant, en utilisant les questions A4A, A4B, A4C et A4F. Pour faciliter les comparaisons internationales, on a attribué à ces questions les codes de la Classification internationale type de l'éducation (CITE). Après avoir vérifié ces codes, on les a utilisés pour recréer la classification canadienne de l'éducation qu'on trouve dans la variable A4. On a ainsi assuré la concordance entre les deux systèmes de classification.

De même, on a utilisé les questions C2 et C6 pour créer respectivement C2_ISCED (scolarité de la mère du répondant) et C6_ISCED (scolarité du père du répondant). Enfin, on a utilisé la question F4 pour créer F4_ISCED, variable liée à l'éducation des adultes, qui mesurait si le répondant suivait ou avait suivi des cours en vue d'obtenir un grade ou un diplôme supplémentaire. Comme le nom l'indique, on l'a codée selon le système CITE et, à l'instar de A4, on l'a codée de nouveau selon les étiquettes canadiennes.

Attribution des valeurs manquantes

Les valeurs manquantes dans les variables ont été recodées « 9 »/« 99 » ou « 6 »/« 96 », selon la logique de la question. On a calculé certaines valeurs manquantes en utilisant les valeurs d'une autre variable. Par exemple, lorsque les valeurs de la variable « BookID » étaient manquantes, on a utilisé celles de la variable « BookLTID ».

Recodage

Certaines variables n'étaient pas compatibles avec le fichier international; il a donc fallu les rajuster. La réponse « Ne sait pas » aux questions A9A à A9D en constitue un bon exemple. Dans le fichier international, on ne pouvait pas utiliser cette réponse, car « Pas d'opinion » était essentiellement assimilé à « Ne sait pas ». Aux questions Q9A à Q9D, les valeurs de « 7 » (Ne sait pas) ont donc été recodées « 5 » (Pas d'opinion). À la question A10, la valeur de « 7 » est devenue « 8 » afin de respecter les normes internationales.

Recodage des questions B1A, B1B et B1C

La question B1 servait à déterminer la « première langue apprise et encore comprise » (langue maternelle). Toutefois, afin de refléter la réalité multilingue canadienne, l'EIACA avait prévu jusqu'à trois réponses au choix, à condition que toutes les langues aient été apprises et comprises en même temps (selon l'interprétation subjective des répondants). Il fallait combiner ces trois choix en une seule variable utilisable. On a créé cette variable en retenant les trois premiers choix de chaque répondant et en créant pour chaque répondant une matrice selon ses premier, deuxième et troisième choix. On a compté neuf cas où le répondant avait choisi trois langues apprises en même temps. Près de 300 répondants (1,3 %) ont mentionné deux langues apprises également en même temps.

Recodage des variables ISCOM (mère) et ISCOF (père)

On a utilisé la Classification internationale type des professions (CITP) pour coder la profession du père (F pour Father) et celle de la mère (M pour Mother) du répondant. Conformément aux exigences internationales, on a étendu les codes de ces variables en les faisant passer de deux à quatre chiffres. On a simplement ajouté deux zéros au code de deux chiffres existant. Dans le fichier international, le codage de la profession du père et de la mère prévoyait une valeur pour « Sans objet » qu'il fallait déterminer et, comme il y avait une question distincte servant à établir si un répondant âgé de 16 ans était en mesure de répondre aux questions sur chaque parent, on a attribué l'état « Sans objet » pour le code ISCO pertinent. On a ainsi créé un code pour « Sans objet » en utilisant les renseignements des questions C3A pour la mère du répondant et C7A pour son père.

Recodage des autres codes d'industrie et de profession

En ce qui concerne la profession et l'industrie des répondants, on a saisi les valeurs de « Ne sait pas » et de « Refus » d'après les questions initiales posées sur place avant d'envoyer le fichier au codage. Pour la profession, on a créé les codes CITP (Classification internationale type des professions), CTP-91 (Classification type des professions, 1991) et CNP-S (Classification nationale des professions pour statistiques, 2001) d'après les questions D27 et D28; pour l'industrie, on a créé les codes SCIAN (Système de classification des industries de l'Amérique du Nord), CTI-80 (Classification type des industries, 1980) et CITI (Classification internationale type par industrie) d'après les questions D25 et D26.

Recodage des questions sur le revenu

Aux questions D41 (salaire et traitement avant impôt) et D42 (salaire mensuel brut), le fichier international comportait un code supplémentaire indiquant que le répondant ne pouvait déclarer que ses gains nets (après impôt). Afin d'inclure ce code dans le fichier national, on a créé un code « 9999999995 », lorsque des valeurs valides existaient respectivement dans les réponses aux questions D41A et D42A, c'est-à-dire des valeurs supérieures à 0 et inférieures à 9999999995.

Changement de longueur variable

On a créé certaines variables avec deux octets au lieu d'un. Il y avait un grand nombre de variables de ce genre, notamment les variables « mois » saisies sur place, car il s'agissait de questions de type « Cocher toutes les réponses qui s'appliquent » et qu'il fallait tenir compte des mois 1 à 12. Au cours du processus, on en a fait des variables à deux chiffres au moment de les désenchaîner, vu la nature de la question. Comme elles n'avaient que des valeurs de 1, 2, 6, 7, 8 et 9, une longueur d'un octet était suffisante. On en a donc fait des questions à un chiffre. Les questions D5MA à D5MI, D6AMA à D6AML, D6BMA à D6BML, D10MA à D10ML, D12MA à D12ML, D14MA à D14ML et F5MA à F5ML en constituent quelques exemples.

En outre, dans le module « Activités sur le marché du travail », on a fait passer les questions D42 et D42A de 9,2 à 10,2 chiffres pour les rendre cohérentes avec les questions sur le revenu.

On a retiré du fichier national toutes les décimales, à l'exception des variables poids.

Création d'un numéro de répondant unique

Le système a produit un numéro d'identification à cinq chiffres (IDNO) qui rend impossible d'identifier un enregistrement par l'emplacement géographique ou par la personne elle-même. Dans le fichier international, on a retiré le numéro d'identification (« sampleid ») et seul le numéro à cinq chiffres sert à lier le fichier international au fichier national.

6.4.8 Création du fichier national de microdonnées à grande diffusion (étape 8)

Une fois le fichier principal finalisé, il est possible de construire le fichier de microdonnées à grande diffusion canadien de l'EIACA de 2003. De ce deuxième fichier découle le fichier international de microdonnées à grande diffusion de l'ELCA.

Le FMGD national est donc un sous-ensemble du fichier principal compte tenu des variables diffusées dans le fichier et du niveau de détail présenté.

Statistique Canada met tout en œuvre pour protéger la confidentialité de chaque répondant à chacune de ses enquêtes. Naturellement, la mesure la plus restrictive serait l'absence complète d'un fichier de microdonnées à grande diffusion. Dans le cas de l'EIACA, il serait difficile d'identifier un répondant à partir d'une note de « 283,7839 » en compréhension de textes suivis. Par contre, plus on ajoute de variables, plus un répondant se distingue des autres. Il ne se fond plus dans l'anonymat des nombres et redevient une personne à part entière. Prenons, par exemple, un homme de 24 à 35 ans d'origine ukrainienne, diplômé universitaire, vivant en région rurale en Saskatchewan, travaillant comme vétérinaire dans un établissement agricole

qui emploie plus de 75 personnes et touchant un salaire annuel supérieur à 47 000 \$. Il est marié et père d'un enfant de moins de cinq ans, suit des cours et des séminaires toute l'année, utilise quotidiennement l'ordinateur au travail et a obtenu une note de « 283,7839 » en compréhension de textes suivis. Soudain, on a l'impression de mieux connaître la « personne ».

Quand on combine un grand nombre de variables sans corrélation, on trouve inévitablement des enregistrements uniques dans la totalisation croisée. À chaque répondant inscrit dans le FMGD de l'EIACA correspond un poids d'échantillonnage, qui suppose implicitement que chaque répondant n'est pas « vraiment » unique parmi l'ensemble de la population. En d'autres termes, les enregistrements uniques résultant d'une combinaison de variables nous apprennent seulement que le répondant était unique à l'intérieur de l'échantillon de l'EIACA, mais pas de l'ensemble de la population.

Il convient de souligner que l'analyse visant à déceler le risque de divulgation n'était pas limitée à un examen des poids; en effet, on a pris plusieurs mesures pour évaluer le risque. D'après les résultats de ces tests et de ces mesures, l'EIACA a adopté diverses méthodes visant à protéger la confidentialité de ses répondants. En voici quelques-unes :

- 1) **Suppression des variables** : Dans certains cas, des variables individuelles étaient si sensibles qu'il a fallu les supprimer complètement du fichier de microdonnées à grande diffusion. L'adresse et le numéro de téléphone du répondant entrent dans cette catégorie. On a également supprimé des variables pour débarrasser le FMGD de variables administratives dont la valeur analytique était douteuse.
- 2) **Recodage ou regroupement** : Dans d'autres cas, on a recodé une variable sensible sur le plan du détail en des groupements moins visibles (revenu, année d'immigration, salaire, âge, etc.) ou on a combiné les groupements existants en groupes plus grands et moins visibles. S'il s'agissait nettement de tenter d'inclure le plus de renseignements, ces préoccupations concernant les données étaient toutefois subordonnées à l'objectif supérieur de la confidentialité. En d'autres termes, on a tenté de conserver le plus de variables possible au niveau le plus détaillé possible, mais on a établi le fichier en se souciant surtout de protéger la confidentialité.
- 3) **Suppression sélective** : À mi-chemin entre les deux options, la suppression sélective consiste à supprimer complètement une variable, mais seulement pour certains répondants ou certains groupes identifiables de répondants. Cette mesure préserve le niveau de détail pour la plus grande partie de la population et n'entraîne qu'une incidence minime sur les erreurs-types des estimations utilisées avec ces variables.

Essentiellement, les variables qui ne comportaient aucun risque pour la confidentialité ont été établies telles quelles dans le FMGD de l'EIACA. Celles qui suscitaient un risque limité pour certaines sous-populations seulement ont fait l'objet de suppressions sélectives (on a supprimé les valeurs pour un groupe sélectionné de répondants). Celles qui comportaient un risque plutôt général à cause de leur niveau de détail ont été recodées ou leurs valeurs ont été combinées en groupements assez importants pour masquer tout risque pour la confidentialité. Enfin, celles qui présentaient un risque réel et incontournable pour la confidentialité ou celles qui n'offraient guère de valeur analytique ont été entièrement supprimées du fichier de microdonnées à grande diffusion de l'EIACA.

6.4.9 Création du FMGD international (étape 9)

Tout comme le FMGD national peut être considéré comme un sous-ensemble du fichier principal national, le volet canadien du fichier international de microdonnées à grande diffusion de l'ELCA est, en réalité, un sous-ensemble du fichier canadien de microdonnées à grande diffusion. Les variables comprises dans le fichier national mais non dans le FMGD international sont le statut d'Autochtone, le statut d'immigrant, la langue maternelle et l'utilisation d'une deuxième langue, la maîtrise de la langue officielle, ainsi que les groupements canadiens de variables liées à la scolarité et d'autres variables propres au Canada ou sans contrepartie comparable dans d'autres pays.

7.0 Totalisation et analyse des données

La présente section résume les lignes directrices à respecter par les utilisateurs qui totalisent, analysent, publient ou diffusent des données tirées des bandes de microdonnées de l'enquête. À l'aide de ces lignes directrices, les utilisateurs de microdonnées devraient être en mesure de produire les mêmes chiffres que ceux qui sont produits par Statistique Canada, ainsi que des chiffres encore inédits de manière conforme aux présentes lignes directrices.

7.1 Utilisation de coefficients de pondération dans la totalisation

Les enquêtes sur la littératie reposent sur des plans d'échantillonnage complexes avec stratification, plusieurs degrés de sélection et des probabilités inégales de sélection des répondants. L'utilisation de données tirées d'enquêtes aussi complexes présente des problèmes pour les analystes, car le plan d'enquête et les probabilités de sélection ont une incidence sur les méthodes à retenir pour calculer les estimations et les variances. Afin d'éviter tout biais dans les estimations et dans l'analyse des données, **on doit utiliser les coefficients de pondération de l'enquête.**

Si bon nombre de méthodes d'analyse qu'on trouve dans les programmes statistiques permettent d'utiliser des coefficients de pondération, la signification ou la définition de ces derniers diffère de celle qui convient dans le cadre d'une enquête par sondage. Par conséquent, alors que les estimations produites par le programme sont souvent exactes, les variances qui sont calculées sont peu précises. Des programmes servant à calculer les erreurs-types d'estimations simples comme des totaux, des proportions et des ratios (dans le cas de variables qualitatives) sont présentés dans la section suivante.

7.2 Définitions de l'estimation catégorique et de l'estimation quantitative

Avant d'aborder les méthodes de totalisation et d'analyse des données de l'EIACA, il convient de décrire les deux principaux types d'estimations ponctuelles des caractéristiques d'une population qu'on peut produire à partir du fichier de microdonnées de l'enquête.

Estimation catégorique

Une estimation catégorique est une estimation du nombre ou du pourcentage de la population visée par une enquête qui possède certaines caractéristiques ou qui entre dans une catégorie donnée. On peut citer comme exemple le nombre d'Albertains qui se situent au niveau 1 sur l'échelle de compréhension de textes suivis ou la proportion de Canadiens qui se situent au

niveau 4 sur l'échelle de la numératie. Dans le cas du nombre de personnes possédant une certaine caractéristique, on peut aussi parler de l'estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Regardez-vous parfois la télévision ou des vidéos dans une autre langue que le français ou l'anglais?

R : **Oui / Non**

Q : Comment coteriez-vous votre capacité de lecture en anglais, dont vous avez besoin dans la vie courante?

R : **Excellente / Bonne / Passable / Médiocre**

Estimation quantitative

Une estimation quantitative est une estimation du total, de la moyenne, de la médiane ou d'une autre mesure de la tendance centrale d'une quantité, fondée sur une partie ou sur la totalité des membres de la population visée par une enquête. Elle se présente essentiellement sous la forme X/Y où X est l'estimation d'un total quantitatif de la population visée par l'enquête et Y est une estimation du nombre de personnes au sein de la population visée par l'enquête qui font partie de ce total.

On peut citer comme exemple d'une estimation quantitative le nombre moyen d'employeurs que les travailleurs canadiens ont eus au cours des 12 derniers mois. Le numérateur est une estimation du nombre total d'employeurs que les travailleurs canadiens ont eus au cours des 12 derniers mois, et le dénominateur est le nombre de Canadiens ayant déclaré qu'ils ont travaillé au cours des 12 derniers mois.

Exemples de questions quantitatives :

Q : Combien d'employeurs différents avez-vous eus au cours des 12 derniers mois?

R : **employeur(s)**

Q : Combien d'heures par semaine travaillez-vous habituellement à cet emploi?

R : **heures**

7.2.1 Totalisation des estimations catégoriques

À partir du fichier de microdonnées, on peut obtenir des estimations du nombre de personnes qui possèdent une certaine caractéristique à l'intérieur d'un pays donné en additionnant les coefficients de pondération finaux de tous les enregistrements présentant la caractéristique en question.

On obtient, pour le pays, des proportions et des ratios de forme X/Y comme suit :

- a) en additionnant les coefficients de pondération finaux des enregistrements présentant la caractéristique d'intérêt pour le numérateur (X),
- b) en additionnant les coefficients de pondération finaux des enregistrements présentant la caractéristique d'intérêt pour le dénominateur (Y), puis
- c) en divisant l'estimation du numérateur par celle du dénominateur.

7.2.2 Totalisation des estimations quantitatives

À partir du fichier de microdonnées, on peut obtenir des estimations quantitatives en multipliant la valeur de la variable d'intérêt par le coefficient de pondération final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements d'intérêt. Par exemple, pour obtenir une estimation du nombre total d'employeurs différents que les travailleurs à temps partiel d'un pays donné ont eus au cours des 12 derniers mois, on multiplie la valeur déclarée en réponse à la question D4 (nombre d'employeurs) par le coefficient de pondération final de l'enregistrement, puis on additionne cette valeur pour tous les enregistrements comportant la valeur D5=2 (temps partiel).

Pour obtenir une moyenne pondérée de forme X/Y , on calcule le numérateur (X) selon la méthode de l'estimation quantitative et le dénominateur (Y), selon la méthode de l'estimation catégorique. Par exemple, pour estimer le nombre moyen d'employeurs que les travailleurs à temps partiel d'un pays donné ont eus au cours des 12 derniers mois :

- a) on estime le nombre total d'employeurs selon la méthode décrite ci-dessus,
- b) on estime le nombre de personnes comprises dans cette catégorie en additionnant les coefficients de pondération finaux de tous les enregistrements comportant la valeur QD5=2, puis
- c) on divise l'estimation a) par l'estimation b).

7.3 Estimations des niveaux de compétence

Le plan d'échantillonnage de l'EIACA est une adaptation du modèle logistique à trois paramètres de la théorie de la réponse d'item. Le premier paramètre (A) est la capacité de l'item de discriminer (sensibilité aux capacités) et le deuxième (B) est sa difficulté. Un troisième paramètre (C) est le paramètre inférieur de l'asymptote, qui représente la chance, possiblement non nulle, d'une bonne réponse, indépendamment de la capacité. Toutefois, comme le test de l'EIACA n'utilisait pas de questions à choix multiple, ce paramètre (C) a été fixé à zéro partout; l'équation est donc devenue, pour ainsi dire, un modèle logistique à deux paramètres. Après avoir calculé les paramètres, on attribue à chaque item une probabilité de réponse de 80 (PR80), qui correspond au niveau de capacités nécessaire à un répondant pour exécuter la tâche avec une probabilité de réussite de 80 %.

Ainsi qu'il est mentionné plus haut, on a résumé les capacités d'un répondant sur les trois échelles en utilisant les paramètres des items et les capacités du répondant conformément aux modèles d'échelonnement selon la méthode TRI. L'application différait quelque peu de la norme puisque, lors de l'EIACA, on a soumis à chaque répondant un nombre relativement restreint d'items afin de cerner plus efficacement les niveaux de capacités de la population. Toutefois, comme les données ne visent pas à estimer des niveaux de capacités individuels, il faut procéder à des analyses plus compliquées.

On a employé la méthode des valeurs plausibles pour estimer uniformément les caractéristiques clés de la population et pour en calculer d'autres avec autant d'exactitude que selon la méthode TRI standard. Cet aspect supplémentaire exige essentiellement que l'estimation des capacités soit fondée sur une série de cinq valeurs plausibles pour chacun des trois aspects de la littératie. Ces cinq valeurs plausibles—allant de prose1 à prose5 pour l'échelle de compréhension de textes suivis, de doc1 à doc5 pour celle des textes schématiques, de num1 à num5 pour celle de la numératie et de health1 à health5 pour celle

des connaissances en matière de santé—ont été recodées en fonction de niveaux plausibles ayant des valeurs de 1 à 5 qui reflètent la progression, déterminée empiriquement, des capacités et des stratégies de traitement de l'information nécessaires pour exécuter des tâches de plus en plus complexes. Le niveau 1 correspond aux notes allant de 0 à 226 (inclusivement); le niveau 2, aux notes de 226,0001 à 276; le niveau 3, aux notes de 276,0001 à 326; le niveau 4, aux notes de 326,0001 à 376 et le niveau 5, aux notes égales ou supérieures à 376,0001. Les variables sont appelées plev1 à plev5 pour l'échelle de compréhension de textes suivis, dlev1 à dlev5 pour celle des textes schématiques et nlev1 à nlev5 pour celle de la numération.

Le cadre de travail n'étant pas tout à fait le même, on a construit un peu différemment l'échelle de la résolution de problèmes. On a d'abord créé les cinq mêmes valeurs plausibles allant de 0 à 500 (prob1 à prob5), mais la définition des niveaux était légèrement différente. Par exemple, cette échelle ne définit que quatre niveaux de compétence, du niveau 1 (le plus faible) au niveau 4 (le plus élevé). S'il est nécessaire de regrouper les niveaux 4 et 5 afin de répliquer les estimations publiées (on trouve habituellement trop peu de répondants au niveau 5 pour produire des estimations fiables) pour les échelles de textes suivis, de textes schématiques, de numération et de santé, cette étape n'est toutefois pas nécessaire pour le domaine de la résolution de problèmes.

Pour calculer des estimations ponctuelles simples relatives à l'un des cinq aspects de la littératie, il suffit d'utiliser le coefficient de pondération de la population ainsi que l'une des cinq valeurs plausibles correspondantes (choisies au hasard).

Toutefois, on peut obtenir une estimation ponctuelle plus précise en établissant la moyenne des cinq estimations produites à partir de chacune des cinq valeurs plausibles, qu'on peut calculer comme suit :

$T. = (\sum_i T_i) / 5$, où T_i est un vecteur de cinq estimations pondérées établies d'après chacune des cinq valeurs plausibles.

Il faut retenir qu'en prenant la moyenne des cinq valeurs plausibles, on ne produit qu'une estimation ponctuelle valide, et non une estimation valide de la variance. On doit utiliser **les cinq** valeurs plausibles ainsi que les 30 poids de rééchantillonnage afin de calculer correctement des estimations de la variance fondées sur le plan d'échantillonnage. Ces dernières sont abordées plus en détail à la section 8.1.2. (Utilisation des valeurs plausibles et des poids de rééchantillonnage dans le calcul de l'erreur d'échantillonnage).

7.4 Arrondissement des estimations

Afin que les estimations calculées d'après le fichier de microdonnées et destinées à la publication ou à la diffusion correspondent à celles qui sont produites par Statistique Canada, les utilisateurs doivent respecter les lignes directrices suivantes concernant l'arrondissement de ces estimations :

- a) Les estimations figurant dans le corps principal d'un tableau statistique doivent être arrondies à la centaine la plus proche selon la méthode d'arrondissement normale. Autrement dit, si le premier ou le seul chiffre à supprimer est compris entre 0 et 4, le dernier chiffre à conserver reste inchangé. Si le premier ou le seul chiffre à supprimer est compris entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple,

selon la méthode d'arrondissement normale à la centaine la plus proche, si les deux derniers chiffres sont compris entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (celui des centaines) reste inchangé. Si les deux derniers chiffres sont compris entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.

- b) Les taux partiels marginaux et les totaux marginaux figurant dans les tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis être arrondis à la centaine la plus proche selon la méthode d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (numérateurs ou dénominateurs), puis être arrondis à une seule décimale selon la méthode d'arrondissement normale. Selon la méthode d'arrondissement normale à un seul chiffre, si le dernier ou le seul chiffre à supprimer est compris entre 0 et 4, le dernier chiffre à conserver reste inchangé. Si le premier ou le seul chiffre à supprimer est compris entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou ratios) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis être arrondies à la centaine la plus proche (ou à la décimale la plus proche) selon la méthode d'arrondissement normale.
- e) Dans les cas où, à cause de restrictions d'ordre technique ou autre, on utilise une autre méthode d'arrondissement que la méthode d'arrondissement normale, produisant ainsi des estimations qui diffèrent des estimations correspondantes publiées par Statistique Canada, les utilisateurs doivent préciser la raison de ces différences dans les documents destinés à la publication ou à la diffusion.
- f) Sous aucun prétexte les utilisateurs ne doivent publier ou diffuser des estimations non arrondies. Ces dernières laissent supposer que la précision est plus grande qu'elle ne l'est en réalité.

8.0 Qualité des données

On peut évaluer la qualité des données d'une enquête en examinant deux types d'erreur : l'erreur d'échantillonnage et l'erreur non liée à l'échantillonnage.

Les estimations tirées de la présente enquête sont fondées sur un échantillon de personnes. On aurait sans doute obtenu des chiffres quelque peu différents si on avait procédé à un recensement intégral faisant appel au questionnaire, aux intervieweurs, aux superviseurs, aux méthodes de traitement, etc., effectivement utilisés pour les besoins de l'enquête. L'écart entre l'estimation obtenue à partir de l'échantillon et les résultats d'un dénombrement intégral mené dans des conditions semblables est désigné **l'erreur d'échantillonnage** de l'estimation.

Par ailleurs, des erreurs qui ne sont pas dues à l'échantillonnage peuvent survenir à peu près à chaque étape de l'enquête. L'intervieweur peut mal comprendre une directive, le répondant peut se tromper au moment de répondre, les réponses peuvent être mal transcrites sur le questionnaire, et des erreurs peuvent survenir au moment du traitement et de la totalisation des données. Il s'agit là **d'erreurs non liées à l'échantillonnage**.

8.1 Erreurs d'échantillonnage

Comme les estimations d'une enquête par sondage sont forcément entachées d'une erreur d'échantillonnage, le chercheur qui tient à adopter une saine méthode statistique voudra donner aux utilisateurs une idée de l'ampleur de cette erreur. La présente section résume les évaluations de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et qu'il incite les utilisateurs à employer eux aussi s'ils produisent des estimations à partir de ce fichier de microdonnées.

Pour mesurer l'ampleur éventuelle des erreurs d'échantillonnage, on fait appel à l'erreur-type des estimations tirées des résultats d'enquête.

Par contre, étant donné la grande diversité des estimations que peut produire une enquête, l'erreur-type d'une estimation s'exprime habituellement par rapport à l'estimation à laquelle elle se rapporte. La mesure qui en découle, désignée le coefficient de variation (c.v.) de l'estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même; elle s'exprime en pourcentage de l'estimation.

Par exemple, supposons que, sur la foi des résultats de l'enquête, on estime que 16,6 % des Canadiens se situent au niveau 1 en compréhension de textes suivis et que l'erreur-type de cette estimation s'élève à 0,013. Le coefficient de variation de l'estimation se calcule donc comme suit :

$$\left(\frac{.013}{.166} \right) \times 100\% = 7.8\%$$

8.1.1 Lignes directrices en matière de diffusion du c.v.

Parmi les critères servant à déterminer si les estimations d'une enquête sont publiables, on retrouve le coefficient de variation (c.v.), qui correspond à l'erreur-type d'une estimation exprimée en pourcentage de l'estimation.

Avant de diffuser ou de publier une estimation de l'EIACA, l'utilisateur doit d'abord déterminer le niveau de qualité de l'estimation, qui peut être acceptable, marginale ou inacceptable. La qualité des données est compromise à la fois par l'erreur d'échantillonnage et par l'erreur non liée à l'échantillonnage. Pour les besoins de la diffusion, cependant, la qualité d'une estimation est déterminée uniquement d'après l'erreur d'échantillonnage reflétée par le coefficient de variation, tel qu'il ressort du tableau 8.1. L'utilisateur doit néanmoins lire la section 8 afin de mieux comprendre les caractéristiques qualitatives de ces données.

D'abord, il convient de déterminer le nombre de répondants dont les données entrent dans le calcul de l'estimation. S'il est inférieur à 30, la qualité de l'estimation pondérée doit être considérée comme inacceptable. Pour une estimation pondérée fondée sur un échantillon d'au moins 30 personnes, l'utilisateur doit déterminer le c.v. de l'estimation et adopter les lignes directrices énoncées ci-dessous, qui s'appliquent aux estimations pondérées arrondies. Toute estimation peut être considérée comme diffusable. Toutefois, celles dont la qualité est marginale ou inacceptable doivent s'accompagner d'une mise en garde à l'intention de l'utilisateur.

Tableau 8.1 Lignes directrices en matière de niveau de qualité

Niveau de qualité de l'estimation	Lignes directrices
1. Acceptable	L'estimation se caractérise par un échantillon de 30 personnes ou plus et un faible coefficient de variation, allant de 0,0 % à 16,5 %. Aucune mise en garde n'est requise.
2. Marginale	L'estimation se caractérise par un échantillon de 30 personnes ou plus et un coefficient de variation élevé, allant de 16,6 % à 33,3 %. Ces estimations doivent être signalées au moyen de la lettre M (ou d'un indicateur semblable). Elles doivent s'accompagner d'une mise en garde à l'intention de l'utilisateur concernant le haut niveau d'erreur associé aux estimations.
3. Inacceptable	L'estimation se caractérise par un échantillon inférieur à 30 personnes ou un coefficient de variation très élevé, supérieur à 33,3 %. Statistique Canada recommande de ne pas diffuser les estimations de qualité inacceptable. Par contre, si l'utilisateur choisit de le faire, les estimations doivent être signalées au moyen de la lettre U (ou d'un indicateur semblable) et s'accompagner de la mise en garde suivante : « Nous informons l'utilisateur que... (préciser la nature des données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et risquent d'être erronées. Ces données et les conclusions qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également le présent avertissement. »

8.1.2 Utilisation des valeurs plausibles et des poids de rééchantillonnage dans le calcul de l'erreur d'échantillonnage

La présente section reprend essentiellement le contenu de la documentation qui accompagne les programmes STATTOOL (SAS et SPSS) conçus par Statistique Canada pour aider les utilisateurs à manipuler les données de l'ELCA et de l'EIACA. Les programmes et les outils évoqués dans la présente section figurent sur le CD-ROM du fichier de microdonnées à grande diffusion de l'EIACA, dans le répertoire intitulé « STATTOOL ». Certains détails de la présente section sont plutôt axés sur les comparaisons internationales que permet d'établir le fichier de microdonnées à grande diffusion de l'ELCA, mais l'exposé qui suit renseignera également le lecteur sur les utilisations et les limites pratiques des données de l'EIACA.

Calcul des estimations ponctuelles

Dans la présente section, nous allons voir comment utiliser les poids d'échantillonnage (POPWT) pour obtenir des estimations de population sous forme de pourcentages (totaux) et de moyennes. (Le calcul des erreurs-types est présenté dans la section 6.)

Tous les exemples sont fondés sur une population fictive présentant les caractéristiques suivantes :

Simplifier l'analyse provisoire

L'ensemble de données de l'EIACA est certes difficile à manipuler. Les cinq valeurs plausibles correspondant aux cinq domaines (y compris les connaissances en matière de santé), ainsi que les trente poids de rééchantillonnage, compliquent l'évaluation exacte des erreurs-types.

Dans bien des cas, la simplification du processus, notamment à l'étape exploratoire, réduirait grandement le temps de traitement nécessaire à l'analyse de l'estimation des compétences.

Il est donc recommandé, à l'étape de la recherche préliminaire, d'utiliser une seule des valeurs plausibles (VP), au lieu des cinq. Cette méthode est beaucoup plus exacte que le calcul de la moyenne des cinq VP, car ainsi, la répartition pondérée de la population reflète exactement l'estimation ponctuelle. La moyenne des VP, elle, masquerait l'erreur de test et, moins la population étudiée est nombreuse, plus les estimations s'écarteraient de la répartition réelle de la population.

Naturellement, lorsqu'on est prêt à publier les résultats de la recherche, on doit utiliser les poids de rééchantillonnage et les cinq valeurs plausibles pour produire les estimations finales avec des erreurs-types exactes. Cette méthode est décrite en détail dans la section 8.1.2.3.

Type	Sexe	Répartition de la population	Échantillon	
			Répartition (non pondérée)	Répartition (pondérée)
Région rurale	Hommes	40,0	30,0	38,0
	Femmes	60,0	70,0	62,0
	Total	20,0	50,0	19,0
Région urbaine	Hommes	51,0	45,0	50,0
	Femmes	49,0	55,0	50,0
	Total	80,0	50,0	81,0
Total	Hommes	48,8	37,5	47,7
	Femmes	51,2	62,5	52,3
	Total	100,0	100,0	100,0

D'après ce tableau, il semble que la participation des hommes était inférieure à celle des femmes tant en région rurale qu'en région urbaine. Même si les hommes représentent près de 49 % de la population, on ne trouve que 37,5 % des hommes dans l'échantillon. On observe cette tendance dans les deux types de région. Il semble également que la région rurale soit surreprésentée, car 50 % de l'échantillon proviennent de cette région, contre seulement 20 % de la population.

Toutefois, lorsqu'on utilise les poids d'échantillonnage, les pourcentages sont fort comparables. Comment les calcule-t-on?

Pourcentages (totaux)

On a calculé comme suit le pourcentage pondéré des hommes vivant en région rurale :

$$\text{Weighted \%} = \frac{\sum_{i=1}^{rural,male} POPWT_i}{\sum_{i=1}^{rural} POPWT_i} = 38\% \text{ où } i \text{ représente la personne } i. \text{ Le numérateur est une}$$

estimation de la population totale des hommes vivant en région rurale, et le dénominateur, une estimation de la population totale vivant en région rurale.

On a calculé comme suit le pourcentage non pondéré :

$$\text{Unweighted \%} = \frac{\sum_{i=1}^{rural,male} 1_i}{\sum_{i=1}^{rural} 1_i} = \frac{n_{rural,male}}{n_{rural}} = 30\% \text{ où } n_{rural,male} \text{ est le nombre d'hommes vivant en}$$

région rurale qu'on trouve dans l'échantillon, et n_{rural} , le nombre total de personnes vivant en région rurale qu'on trouve dans l'échantillon.

Dans la dernière équation, chaque personne échantillonnée compte pour un, alors que dans la version pondérée, chaque unité échantillonnée a reçu un poids afin de représenter convenablement et proportionnellement les sous-groupes de l'échantillon (noter que le pourcentage pondéré est un ratio des totaux pondérés estimatifs).

Moyennes

Pour cet exemple fictif, supposons que nous avons également la note moyenne fondée sur la variable PROSE1, comme l'illustre le tableau ci-dessous :

Type	Sexe	Répartition de la population	Échantillon			
			Répartition (non pondérée)	Répartition (pondérée)	Avg. Prose1 (non pondérée)	Avg. Prose1 (pondérée)
		%	%	%		
Région rurale	Hommes	40,0	30,0	38,0	260	260,1
	Femmes	60,0	70,0	62,0	290	289,8
	Total	20,0	50,0	19,0	281,0	278,5
Région urbaine	Hommes	51,0	45,0	50,0	320	319,7
	Femmes	49,0	55,0	50,0	330	330,1
	Total	80,0	50,0	81,0	325,5	324,9
Total	Hommes	48,8,0	37,5	47,7	296,0	310,7
	Femmes	51,2,0	62,5	52,3	307,6	321,0
	Total	100,0	100,0	100,0	303,3	316,1

Ici encore, nous voyons que les moyennes pondérées sont très proches des moyennes non pondérées, pourvu qu'on neutralise le type de région. Ce n'est pas le cas des trois dernières lignes du tableau. Essayons de voir pourquoi. On a calculé comme suit la moyenne pondérée des hommes vivant en région rurale :

$$\text{Weighted mean} = \frac{\sum_{i=1}^{\text{rural,male}} \text{POPWT}_i * \text{PROSE1}_i}{\sum_{i=1}^{\text{rural,male}} \text{POPWT}_i} = 260.1 \quad \text{où } i \text{ représente la personne } i. \text{ Le}$$

numérateur est une estimation de la note totale pour tous les hommes vivant en région rurale, et le dénominateur, une estimation de l'ensemble de la population masculine vivant en région rurale.

On a calculé comme suit la moyenne non pondérée :

$$\text{Unweighted mean} = \frac{\sum_{i=1}^{\text{rural,male}} \text{PROSE1}_i}{\sum_{i=1}^{\text{rural,male}} 1_i} = \frac{\sum_{i=1}^{\text{rural,male}} \text{PROSE1}_i}{n_{\text{rural,male}}} = 260$$

Les résultats non pondérés et pondérés sont semblables lorsque les valeurs de PROSE1 ne varient guère d'une personne à l'autre ou que les valeurs de POPWT présentent le même comportement. Cet énoncé ne vaut pas pour les trois dernières lignes du tableau. On obtient la moyenne pondérée pour les hommes en résolvant l'équation suivante :

$$\begin{aligned}
 \text{Weighted mean} &= \frac{\sum_{i=1}^{male} POPWT_i * PROSE1_i}{\sum_{i=1}^{male} POPWT_i} \\
 &= \frac{(38\% * 19\% * 260.1) + (50\% * 81\% * 319.7)}{47.7\%} = 310.7
 \end{aligned}$$

alors que la moyenne non pondérée est donnée par :

$$\begin{aligned}
 \text{Unweighted mean} &= \frac{\sum_{i=1}^{male} PROSE1_i}{\sum_{i=1}^{male} 1_i} \\
 &= \frac{(30\% * 50\% * 260) + (45\% * 50\% * 320)}{37.5\%} = 296.0
 \end{aligned}$$

Dans la dernière équation, chaque personne échantillonnée compte pour un, alors que dans la version pondérée, chaque unité échantillonnée a reçu un poids afin de représenter convenablement et proportionnellement les sous-groupes de l'échantillon. Par exemple, les 30 %X50 %=15 % des hommes vivant en région rurale qu'on trouve dans l'échantillon ont été rajustés par les poids pour tenir compte des 38 %X19 %=7,22 % de l'ensemble de l'échantillon, ce qui reflète beaucoup mieux ce qu'on trouve dans l'ensemble de la population. (Il convient de noter que la vraie proportion de la population des hommes vivant en région rurale est 40 %X20 %=8 %.)

En conclusion, il faut toujours utiliser les poids d'échantillonnage pour calculer des statistiques d'après des données échantillons.

Autres poids d'échantillonnage

Comme nous l'avons vu plus haut, la somme des poids d'échantillonnage (POPWT) à l'intérieur d'un échantillon donne une estimation de la taille de la population. Il s'agit d'un poids d'échantillonnage couramment utilisé, mais la somme donne parfois un nombre très élevé et des nombres différents d'un pays à l'autre, ce qui n'est pas toujours souhaitable. Par exemple, si l'on veut calculer une estimation pondérée du rendement moyen de la population de tous les pays (ou de sous-populations à l'intérieur d'un pays), l'utilisation de la variable POPWT comme variable de pondération fait que chaque pays contribue proportionnellement à la taille de sa population, les grands pays comptant plus que les petits pays. En général, le poids POPWT n'est pas le meilleur poids à utiliser pour les analyses portant sur plusieurs pays. Une autre conséquence de l'utilisation du poids POPWT est la tendance à gonfler les résultats de tests de signification lorsque les logiciels informatiques sont incapables de calculer correctement les données pondérées. Nous allons maintenant voir deux versions possibles de poids d'échantillonnage ayant trait à ces questions en particulier. Ces versions tirent parti du fait qu'on obtient les mêmes estimations de la population sous forme de moyennes et de proportions lorsqu'on utilise une variable de pondération proportionnelle au poids de population (POPWT).

Somme donnant le poids d'échantillonnage constant (CONSTWT)

Il est possible de modifier le poids de population POPWT de sorte que tous les pays contribuent également lorsqu'on calcule la moyenne ou une proportion d'un pays à l'autre. L'équation est la suivante :

$$CONSTWT_{g,i} = POPWT_{g,i} * \left[\frac{100}{\sum_{i=1}^g POPWT_i} \right]$$

pour chaque personne du groupe d'intérêt g. La transformation des poids est différente à l'intérieur de chaque pays mais, en fin de compte, la somme de la variable CONSTWT à l'intérieur de chaque pays est de 100. La variable CONSTWT, à l'intérieur de chaque pays, est proportionnelle à POPWT multiplié par le ratio de 100 divisé par la somme de poids pour toutes les personnes du groupe d'intérêt. On peut utiliser ces poids lorsqu'on veut calculer des estimations internationales et qu'on veut que chaque pays contribue également à l'estimation internationale, sans égard à la taille du groupe d'intérêt dans le pays (voir le tableau ci-dessous).

Groupe d'intérêt		Pays	Chiffre de population (région rurale)	Estimations de population
			Mean PROSE1 (POPWT)	Mean PROSE1 (CONSTWT)
Région rurale	A	3 700 000	290	290
	B	37 000 000	260	260
	C	7 000 000	300	300
Ensemble			268	283

Somme donnant le poids d'échantillonnage de la taille de l'échantillon (SMPLWT)

Il est possible de modifier le poids de population POPWT lorsqu'on veut utiliser la taille réelle de l'échantillon pour exécuter des tests de signification (à l'intérieur de chaque pays). L'équation est la suivante :

$$SMPLWT_{g,i} = POPWT_{g,i} * \left[\frac{n_g}{\sum_{i=1}^g POPWT_i} \right]$$

pour chaque personne du groupe d'intérêt g où n_g est la taille réelle de l'échantillon du groupe g. La transformation des poids est différente à l'intérieur de chaque pays mais, en fin de compte, la somme de la variable CONSTWT à l'intérieur de chaque pays donne la taille de l'échantillon du groupe g. La variable SMPLWT, à l'intérieur de chaque pays, est proportionnelle à POPWT multiplié par le ratio de la taille de l'échantillon (n_g) divisé par la somme des poids pour toutes les personnes du groupe d'intérêt. Si certains logiciels de statistique permettent d'utiliser la taille de l'échantillon comme diviseur dans le calcul des erreurs-types, d'autres utilisent plutôt la somme des poids, ce qui entraîne de graves erreurs-types à la baisse pour les statistiques si l'on utilise POPWT comme variable de pondération. Lorsqu'on effectue des analyses à l'aide de ce genre de logiciel, il est recommandé d'utiliser une variable comme SMPLWT comme variable de pondération. En raison de l'effet de regroupement dans les

échantillons de la plupart des pays, il convient également d'appliquer à la variable SMPLWT un facteur de correction, par exemple un effet du plan de sondage.

Utilisation des valeurs plausibles pour calculer des estimations ponctuelles

Pour atteindre son objectif d'une vaste couverture des buts et des processus de la littératie, l'EIACA reposait sur un éventail d'items répartis en livrets d'évaluation. Chaque participant à l'évaluation a rempli un seul livret, ce qui a réduit au minimum le fardeau de réponse individuel. L'ELCA a utilisé un plan d'échantillonnage matriciel pour attribuer les livrets d'évaluation aux répondants afin qu'on puisse broser un tableau complet du rendement en littératie dans chaque pays à partir des composantes remplies par chaque répondant. L'ELCA a utilisé l'échelonnement selon la théorie de la réponse d'item (TRI) pour combiner les réponses individuelles en vue d'établir des estimations exactes du rendement en littératie dans la population de chaque pays. L'échelle TRI de l'ELCA utilise également la méthodologie de l'imputation multiple ou des « valeurs plausibles » pour calculer les notes de compétence en littératie de tous les répondants, même si chacun n'a répondu qu'à une partie de l'ensemble des items d'évaluation.

La plupart des tests portant sur les compétences cognitives visent à évaluer exactement le rendement de répondants individuels à des fins de diagnostic, de sélection ou de placement. On peut améliorer l'exactitude de ces mesures en augmentant le nombre d'items remis à la personne. Pour connaître la répartition des compétences dans une vaste population, toutefois, on peut obtenir des estimations plus efficaces à l'aide d'un plan d'échantillonnage matriciel. Ce type de plan sollicite un petit nombre de réponses de chaque répondant échantillonné, tout en maintenant une vaste représentation du contenu lorsque les réponses sont agrégées pour tous les répondants. Avec cette méthode, toutefois, l'avantage d'estimer les caractéristiques d'une population est plutôt contrebalancé par l'incapacité de formuler des énoncés précis au sujet des personnes; l'agrégation des notes individuelles risque donc de biaiser gravement les estimations des caractéristiques d'une population.

On a mis au point la méthodologie des valeurs plausibles afin de régler ce problème en utilisant toutes les données disponibles pour estimer directement les caractéristiques de populations et de sous-populations, puis en produisant plusieurs notes imputées (appelées valeurs plausibles) d'après ces distributions, qu'on peut utiliser dans les analyses avec un logiciel de statistique courant. Mislevy (1991) présente un examen détaillé de la méthodologie des valeurs plausibles. Les principaux points à retenir sont les suivants :

- a) Lorsqu'on veut calculer des statistiques portant sur des notes (PROSE, DOC, NUMERACY, etc.), on ne dispose pas d'une, mais de cinq valeurs de notes attribuées à chaque personne. Chaque ensemble de valeurs plausibles convient tout aussi bien pour estimer les paramètres de population.
- b) Ces statistiques fondées sur des notes doivent toujours être calculées au niveau d'une population ou d'une sous-population. Il ne faut jamais les utiliser pour établir une inférence au niveau individuel.

Utilisation des valeurs plausibles

Exemple1 : Médiane estimative de la variable PROSE dans le pays A.

Pour chaque personne, on ne dispose pas d'une, mais de cinq notes à prendre en compte, comme le montre le tableau ci-dessous.

Pays A	PROSE1	PROSE2	PROSE3	PROSE4	PROSE5
Personne 1	222	275	300	245	254
Personne 2	289	310	212	250	265
...
Personne n	285	275	243	321	312
Médiane	285	281	283	279	289

Afin d'estimer la médiane globale pour la variable PROSE dans le pays A, on estime d'abord la médiane à partir du premier ensemble de valeurs plausibles correspondant à la variable PROSE1. On répète ensuite cette première étape en utilisant les variables PROSE2 à PROSE5 pour obtenir au total cinq estimations tout aussi valables de la médiane pour le pays en question. Comme elles sont toutes aussi valables, l'étape suivante en vue d'obtenir une seule estimation de la médiane consiste à calculer la moyenne de ces cinq estimations. On obtient ainsi :

$$\text{Médiane globale} = (285 + 281 + 283 + 279 + 289) / 5 = 283,4$$

Notez qu'il ne faut pas calculer la moyenne des notes au niveau individuel. Par exemple, $(222 + 275 + 300 + 245 + 254) / 5 = 259,2$ ne constitue pas une bonne estimation de la variable PROSE pour la personne 1. Il en va de même de chacune des variables PROSEn pour une personne donnée. Les variables fondées sur les notes doivent toujours être interprétées dans le contexte de populations ou de sous-populations.

Les variables PROSE1 à PROSE5 sont des notes correspondant aux capacités brutes. Lorsqu'on regroupe ces notes en niveaux, elles donnent cinq variables de niveau appelées, d'après le domaine de la compréhension de textes suivis, PLEV1 à PLEV5 dans le FMGD. Les variables XPROSE1 à XPROSE5 représentent un recodage des variables PLEV1 à PLEV5 qu'on trouve dans le FMGD de l'EIACA, les niveaux 4 et 5 étant regroupés afin de donner un nombre suffisant de répondants à chaque niveau pour permettre une analyse exacte. On peut procéder de même pour les variables de niveau axées sur les textes schématiques, la numératie et la santé (DLEV1 à DLEV5, NLEV1 à NLEV5 et HLEV1 à HLEV5). Quant aux niveaux de la résolution de problèmes (PSLEV1 à PSLEV5), on les laisse tels quels.

Exemple 2 : Coefficients estimatifs des paramètres de régression logistique (niveaux 2 et 3 des variables dépendantes XPROSE1 à XPROSE5).

D'après les valeurs trouvées dans le tableau précédent, on obtiendrait :

Pays A	XPROSE1	XPROSE2	XPROSE3	XPROSE4	XPROSE5
Personne 1	2	3	3	2	2
Personne 2	3	4/5	2	2	2
...
Personne n	3	2	2	3	3

On remarque avant tout qu'une personne donnée peut se trouver à différents niveaux de compétence, selon l'ensemble de valeurs plausibles pris en compte. Cette caractéristique n'invalide pourtant pas la méthodologie utilisée. Comme nous l'avons expliqué plus haut, afin d'obtenir les coefficients des paramètres de régression logistique au niveau des pays, on calcule d'abord ces paramètres en fonction du premier ensemble de valeurs plausibles. On répète ensuite cette étape quatre fois en utilisant les variables XPROSE2 à XPROSE5 pour obtenir quatre autres ensembles de paramètres estimatifs, comme le montre le tableau ci-dessous :

Pays A	XPROSE1	XPROSE2	XPROSE3	XPROSE4	XPROSE5
Coordonnée à l'origine	0,124	0,129	0,122	0,125	0,120
Bêta 1	1,051	1,059	1,049	1,055	1,060
Bêta 2	0,584	0,591	0,545	0,499	0,645
Bêta 3	3,222	4,123	3,012	3,542	3,201

Comme ces ensembles de coefficients estimatifs des paramètres sont tous aussi valables les uns que les autres, l'étape suivante en vue d'obtenir un seul ensemble d'estimations consiste à calculer la moyenne de ces cinq résultats. On obtient ainsi :

Coordonnée globale à l'origine = $(0,124 + 0,129 + 0,122 + 0,125 + 0,120) / 5 = 0,124$

Bêta 1 global = $(1,051 + 1,059 + 1,049 + 1,055 + 1,060) / 5 = 1,055$

Bêta 2 global = $(0,548 + 0,591 + 0,545 + 0,499 + 0,645) / 5 = 0,566$

Bêta 3 global = $(3,222 + 4,123 + 3,012 + 3,542 + 3,201) / 5 = 3,420$

8.1.3 Estimation de la variance d'erreur dans le cadre de l'EIACA

La méthodologie de l'EIACA est une méthode éprouvée de quantification des compétences. Il existe donc déjà une importante documentation concernant les méthodes d'estimation de la variance relativement à ce genre d'études. Le texte de la présente section provient en grande partie du guide de l'utilisateur du PIRLS 2001 de l'IEA; il a été adapté au contexte de l'EIACA.

Aperçu

Lorsqu'on analyse des données obtenues au moyen d'un plan de sondage complexe comme celui de l'EIACA, il importe de calculer des estimations exactes de la variance d'erreur à l'égard des statistiques d'intérêt. Dans le cas de l'ELCA, cette variance d'erreur peut provenir de deux sources : le processus d'échantillonnage (toujours présent) et le processus d'imputation (lorsque les statistiques d'intérêt portent sur des notes de compétence). La présente section décrit les méthodes d'estimation de ces composantes de la variance d'erreur.

Estimation de la variance d'échantillonnage

Sur le plan analytique, lorsqu'on recueille des données dans le cadre d'une enquête par sondage complexe, il est souvent difficile de produire une estimation de la variance qui soit sans biais ou cohérente avec le plan de sondage. Une catégorie de techniques appelées méthodes de rééchantillonnage permet d'estimer la variance pour le type de plans d'échantillonnage complexes comme ceux qui sont utilisés dans le cas de l'ELCA.

Le principe de base du rééchantillonnage consiste à choisir successivement des sous-échantillons dans l'ensemble de l'échantillon, à calculer la statistique d'intérêt pour chaque sous-échantillon, puis à utiliser ces sous-échantillons ou poids statistiques de rééchantillonnage pour estimer la variance de la statistique fondée sur l'échantillon complet. Différentes façons de créer des sous-échantillons à partir de l'échantillon complet donnent différentes méthodes de rééchantillonnage. Les sous-échantillons sont appelés poids de rééchantillonnage et les statistiques calculées à partir de ces poids sont appelées estimations répétées.

L'une de ces méthodes est la technique de rééchantillonnage jackknife (JRR) (Wolter, 1985). Dans le cas de l'ELCA, à l'intérieur de chaque pays, l'échantillon complet a été scindé au

hasard en 30 sous-ensembles de taille égale ou presque égale, chaque sous-ensemble ressemblant à l'échantillon complet. On forme les poids de rééchantillonnage en supprimant un sous-ensemble à la fois et en multipliant les poids des autres sous-ensembles par

$$\frac{30}{29}$$

On crée ainsi 30 poids de rééchantillonnage. Cette méthode est également appelée la méthode JK1. Les poids liés à chaque poids de rééchantillonnage correspondent aux variables REPLIC01 à REPLIC30. On utilise ces poids uniquement pour calculer la variance d'échantillonnage. Les estimations ponctuelles doivent être calculées selon la méthode décrite dans la section précédente.

Calcul de la variance d'échantillonnage selon la méthode JK1

Le principe de base consiste ici à calculer l'estimation d'intérêt d'après l'échantillon complet en utilisant la variable de pondération POPWT (ou SMPLWT ou CONSTWT) ainsi que chaque poids de rééchantillonnage (en utilisant les variables REPLIC01 à REPLIC30). On utilise ensuite la variation entre les estimations fondées sur les poids de rééchantillonnage et l'estimation fondée sur l'échantillon complet pour estimer la variance d'échantillonnage pour l'échantillon complet. Supposons que $\hat{\theta}$ est l'estimation fondée sur l'échantillon complet d'un paramètre de population θ . L'estimateur de la variance d'échantillonnage $\text{var}(\hat{\theta})$ est donné par :

$$\text{var}_{\text{simpl}}(\hat{\theta}) = \frac{30 \sum_{g=1}^{30} (\hat{\theta}_{(g)} - \hat{\theta})^2}{29} \quad \text{où } \hat{\theta}_{(g)} \text{ est l'estimation de } \theta \text{ fondée sur les observations comprises dans le } g^{\text{e}} \text{ poids de rééchantillonnage.}$$

Lorsque la statistique d'intérêt porte sur des notes de compétence, on fonde habituellement la variance d'échantillonnage uniquement sur le premier ensemble de valeurs plausibles au lieu de calculer cinq fois l'expression ci-dessus et de calculer la moyenne des cinq variances d'échantillonnage estimatives.

Exemple 1 : Revenu personnel moyen dans le pays A

Le revenu personnel moyen est donné par l'expression suivante :

$$\text{Mean personal income} = \frac{\sum_{i=1}^{\text{Country A}} \text{POPWT}_i * D43_i}{\sum_{i=1}^{\text{Country A}} \text{POPWT}_i} = 27\,601$$

Afin de calculer la variance d'échantillonnage, on calcule 30 fois l'expression ci-dessus, chaque calcul étant fondé sur le poids de rééchantillonnage pertinent.

$$\text{Mean personal income}_{(1)} = \frac{\sum_{i=1}^{\text{Country A}} \text{REPLIC01}_i * D43_i}{\sum_{i=1}^{\text{Country A}} \text{REPLIC01}_i} = 26\,983$$

$$\text{Mean personal income}_{(2)} = \frac{\sum_{i=1}^{\text{Country A}} \text{REPLIC02}_i * D43_i}{\sum_{i=1}^{\text{Country A}} \text{REPLIC02}_i} = 26\,146$$

• • •

$$\text{Mean personal income}_{(30)} = \frac{\sum_{i=1}^{\text{Country A}} \text{REPLIC30}_i * D43_i}{\sum_{i=1}^{\text{Country A}} \text{REPLIC30}_i} = 28\,965$$

Puis, on applique simplement la formule de variance donnée plus haut. On obtient ainsi :

$$\text{var}_{\text{smp}}(\hat{\theta}) = \frac{29 \sum_{g=1}^{30} (\hat{\theta}_{(g)} - \hat{\theta})^2}{30} = \frac{29}{30} [(26983 - 27601)^2 + (26146 - 27601)^2 + \dots + (28965 - 27601)^2]$$

Enfin, la statistique $(\hat{\theta} - \theta) / \text{var}(\hat{\theta})^{1/2}$ fait l'objet d'une distribution de t approximative à 29 degrés de liberté.

Estimation de la variance d'imputation

Lorsque les statistiques d'intérêt portent sur des notes de compétence, il est nécessaire d'estimer la variance d'imputation. Comme nous l'avons mentionné dans la section précédente, on produit cinq notes par test pour les participants à l'EIACA. Ces différentes notes sont appelées valeurs plausibles (VP). Lors de l'EIACA, on a remis à chaque personne des blocs d'exercices. L'ensemble de blocs couvre tous les concepts à évaluer, mais chaque répondant ne répondait pas aux questions de tous les blocs. Grâce à un type d'affectation équilibrée de blocs aux répondants, on a couvert l'ensemble complet de questions en agrégeant les réponses des répondants. Pour un groupe de répondants semblables, on a estimé une distribution bayésienne a posteriori des notes. Les valeurs plausibles pour chaque répondant sont réalisées à partir de cette distribution a posteriori. Ces notes ne sont pas significatives pour un répondant individuel mais, en les combinant, on peut les utiliser pour estimer des moyennes et d'autres quantités de population.

Calcul de la variance d'imputation

La méthode générale d'estimation de la variance d'imputation au moyen des valeurs plausibles est la suivante :

- a) D'abord, on estime la statistique d'intérêt θ en utilisant chaque fois un ensemble différent de valeurs plausibles (M) et la variable POPWT (ou SMPLWT ou CONSTWT). Appelons ces cinq estimations $\hat{\theta}_1$ à $\hat{\theta}_5$. La statistique d'intérêt peut être n'importe quel élément estimable d'après les données-échantillons : une moyenne, l'écart entre des moyennes, des centiles, des coefficients de paramètre de régression, etc.
- b) Puis, on arrive à l'estimation globale en calculant la moyenne de $\hat{\theta}_m$ où $m=1, 2, \dots, 5$. Appelons cette estimation $\hat{\theta}$.
- c) La variance d'imputation est calculée comme suit :

$$Var_{imp}(\hat{\theta}) = \left[1 + \frac{1}{5}\right] \times \sum_{m=1}^5 \frac{(\hat{\theta}_m - \hat{\theta})^2}{4}$$

Estimation de la variance d'erreur globale

Dans des conditions idéales et avec des ressources informatiques illimitées, on calculerait la variance d'erreur globale comme suit :

$$Var(\hat{\theta}) = \sum_{m=1}^5 \frac{Var_{smp}(\hat{\theta}_m)}{5} + Var_{imp}(\hat{\theta})$$

Comme chacune des Var_{smp} suppose qu'on calcule les statistiques d'intérêt 30 fois (en utilisant chaque fois un poids d'échantillonnage différent), on peut utiliser plutôt cette formule abrégée consistant à estimer la variance d'échantillonnage à partir du premier ensemble de valeurs plausibles seulement.

$$Var(\hat{\theta}) = Var_{smp}(\hat{\theta}_1) + Var_{imp}(\hat{\theta})$$

Lorsque les statistiques d'intérêt ne portent sur aucune note de compétence, la formule de la variance d'erreur globale devient simplement la suivante :

$$Var(\hat{\theta}) = Var_{smp}(\hat{\theta})$$

Degrés de liberté

$(\hat{\theta} - \theta) / Var(\hat{\theta})^{1/2}$ fait l'objet d'une distribution de t approximative, les degrés de liberté (Jonhson et Rust, 1993) étant donnés par :

$$v = \frac{1}{\frac{f_m^2}{4} + \frac{(1-f_m)^2}{29}}$$

où f_m est donné par :

$$f_m = \frac{Var_{imp}(\hat{\theta})}{Var(\hat{\theta})}$$

En pratique, le nombre de degrés de liberté est fixé à 29 (ce qui est efficace lorsque f est relativement petit, par exemple inférieur à 30 %).

Établissement de comparaisons

Nous allons maintenant voir comment calculer la bonne variance d'erreur quand on compare les estimations d'enquête entre les pays, avec les estimations internationales et à l'intérieur d'un pays. Afin de simplifier le texte, nous envisageons uniquement le rendement moyen estimatif pour la variable PROSE. On devrait pouvoir généraliser directement le contenu de la présente section à n'importe quel type d'estimation d'enquête.

Comparaisons entre les pays

Quand on compare le rendement moyen estimatif pour la variable PROSE entre les pays A et B, la variance d'erreur est donnée par :

$$Var(\hat{\theta}) = Var_{Country A}(\hat{\theta}) + Var_{Country B}(\hat{\theta})$$

Par exemple, supposons que le rendement moyen estimatif est, pour le pays A, de 290 avec une variance d'erreur de 25 et, pour le pays B, de 307 avec une variance d'erreur de 30,25. L'écart entre ces deux pays est donc de $307 - 290 = 17$. La question est la suivante : cet écart de 17 points résulte-t-il d'une erreur due au fait qu'on n'a échantillonné qu'une partie de la population et qu'on n'a administré qu'une partie des items? Pour répondre à cette question, il faut d'abord calculer la statistique suivante (appelée statistique de Wald) :

$$(\hat{\theta})/Var(\hat{\theta})^{1/2} = 17/(25 + 30.25)^{1/2} = 2.287$$

Quand on compare cette valeur à la valeur critique de 95 % d'une distribution de t à 24 degrés de liberté (2,04), on conclut que les données suffisent à affirmer que ces deux pays n'ont pas le même rendement moyen estimatif.

Notez que cette méthode est également valide quand on compare les résultats de l'ELCA à ceux de l'EIACA.

Comparaisons avec les estimations internationales

Parmi les statistiques publiées, l'une des plus importantes montre le rendement moyen de votre pays par rapport à la moyenne internationale. Ici, la variance d'erreur est donnée par :

$$Var(\hat{\theta}) = \frac{(N-1)^2 Var_{smp\ A}(\hat{\theta}_A) + \sum_{k=1, k \neq A}^N Var_{smp\ k}(\hat{\theta}_k)}{N^2} + Var_{imp}(\hat{\theta}_A - \hat{\theta}_{int})$$

où N est le nombre de pays utilisés pour calculer la moyenne internationale, $\hat{\theta}_m$ représente la moyenne estimative pour le pays m , et $\hat{\theta}_{int}$ représente la moyenne internationale estimative.

Prenons, par exemple, les tableaux suivants :

	Rendement moyen	Variance d'échantillonnage
Pays A	290	20
Pays B	300	22
Pays C	286	18
Pays D	324	22
International	300	

Rendement moyen	PROSE1	PROSE2	PROSE3	PROSE4	PROSE5
Pays A	288	292	292	288	290
International	301	300	299	302	298
$\hat{\theta}_A - \hat{\theta}_{int}$	-13	-8	-7	-14	-8

Ici, on constate que le rendement moyen estimatif du pays A est de 290 avec une variance d'échantillonnage de 20 (variance d'imputation de 5) et que le rendement moyen international estimatif, fondé sur quatre pays, est de 300. L'écart entre le résultat du pays A et la moyenne internationale est donc de $290 - 300 = -10$. La question est la suivante : cet écart de 10 points résulte-t-il d'une erreur due au fait qu'on n'a échantillonné qu'une partie de la population et qu'on n'a administré qu'une partie des items? Pour répondre à cette question, il faut d'abord calculer la statistique suivante :

$$Var(\hat{\theta}) = \frac{(3)^2 20 + (22 + 18 + 22)}{4^2} + (1 + \frac{1}{5}) \frac{[(13-10)^2 + (8-10)^2 + (7-10)^2 + (14-10)^2 + (8-10)^2]}{4}$$

$$= 15.125 + 12.6 = 27.725$$

On calcule alors la statistique de Wald :

$$(\hat{\theta}) / \text{Var}(\hat{\theta})^{1/2} = 10 / (27.725)^{1/2} = 1.899$$

Quand on compare cette valeur à la valeur critique de 95 % d'une distribution de t à 24 degrés de liberté (2,04), on conclut que les données ne suffisent pas à affirmer que le rendement moyen estimatif du pays A est différent du rendement moyen international estimatif.

Comparaisons à l'intérieur d'un pays

La plupart du temps, quand on compare des sous-groupes à l'intérieur d'un pays, il n'existe pas de formule directe pour calculer la variance d'erreur globale comme c'était le cas dans les deux sections précédentes. La principale raison en est qu'habituellement, les échantillons des différents sous-groupes ne sont pas traités comme indépendants pour les besoins de tests statistiques. Par conséquent, il faut utiliser une méthode jackknife applicable à des échantillons corrélés pour estimer la variance d'échantillonnage de l'écart entre les sous-groupes. Il s'agit de calculer l'écart entre les sous-groupes une fois pour chacun des 30 échantillons répétés, et cinq fois de plus, soit une fois par ensemble de valeurs plausibles, comme nous l'avons décrit plus haut (voir les sections sur le calcul des variances d'échantillonnage et d'imputation).

Toutefois, on peut aisément utiliser des modèles de régression linéaire pour calculer ces écarts. Voici comment calculer l'écart entre les hommes et les femmes pour la variable PROSE :

- a) Créer des variables nominales pour les sous-groupes; appelons MAN la variable qui prend la valeur 1 si le répondant est un homme et 0 dans le cas contraire, et WOMAN la variable qui prend la valeur 1 si le répondant est une femme et 0 dans le cas contraire. On crée autant de variables nominales qu'il y a de sous-groupes.
- b) En utilisant POPWT, exécuter un modèle de régression linéaire avec PROSE1 comme variable dépendante et MAN ou WOMAN comme variable indépendante. (Lorsqu'il y a plus de variables nominales k et que k est supérieur à 2, choisir $k-1$ variables nominales comme variables indépendantes.) La variable nominale laissée pour compte devient le sous-groupe de référence.
- c) À l'aide des poids de rééchantillonnage, répéter l'étape b) 30 fois.
- d) À l'aide des poids POPWT, répéter l'étape b) quatre fois, soit une fois par ensemble de valeurs plausibles.
- e) Combiner les renseignements tirés des étapes b), c) et d) pour calculer l'estimation ponctuelle globale, la variance d'erreur et la statistique de Wald.

Par exemple, supposons qu'après avoir créé les variables nominales, on obtient, à la suite l'étape b, le résultat suivant :

$$\text{Mean PROSE1} = 270 + 18 * \text{WOMAN} \text{ (pondéré par POPWT)}$$

Cette expression devient simplement $\text{Mean PROSE1} = 270$ lorsque les répondants sont des hommes et $\text{Mean PROSE1} = 288$ lorsque les répondants sont des femmes. Cela signifie que le coefficient qui précède la variable WOMAN dans le modèle de régression est l'écart entre les

femmes et les hommes, alors que la coordonnée à l'origine (270) est le rendement moyen pour le niveau de référence, en l'occurrence les hommes.

À partir de l'étape c, on obtient :

Mean PROSE1 = 271 + 17*WOMAN (pondéré par REPLIC01)

Mean PROSE1 = 269 + 19*WOMAN (pondéré par REPLIC02)

Mean PROSE1 = 273 + 14*WOMAN (pondéré par REPLIC03)

...

Mean PROSE1 = 268 + 21*WOMAN (pondéré par REPLIC30)

On peut maintenant calculer comme suit la variance d'échantillonnage de l'écart entre les femmes et les hommes :

$$\text{var}_{\text{smp}}(\hat{\theta}) = \frac{29 \sum_{g=1}^{30} (\hat{\theta}_{(g)} - \hat{\theta})^2}{30} = \frac{29}{30} [(17-18)^2 + (19-18)^2 + (14-18)^2 + \dots + (21-18)^2] = 19.575$$

À partir de l'étape d, on obtient :

Mean PROSE2 = 271 + 20*WOMAN (pondéré par POPWT)

Mean PROSE3 = 269 + 19*WOMAN (pondéré par POPWT)

Mean PROSE4 = 273 + 17*WOMAN (pondéré par POPWT)

Mean PROSE5 = 268 + 16*WOMAN (pondéré par POPWT)

On peut maintenant calculer l'estimation ponctuelle globale de l'écart en calculant la moyenne des résultats pour PROSE1 à PROSE5. Cela donne : $(18+20+19+17+16) / 5 = 18$. La variance d'imputation est donnée par :

$$\begin{aligned} \text{Var}_{\text{imp}}(\hat{\theta}) &= \left[1 + \frac{1}{5}\right] \times \sum_{m=1}^5 \frac{(\hat{\theta}_m - \hat{\theta})^2}{4} \\ &= 1.2 \times \frac{[(18-18)^2 + (20-18)^2 + (19-18)^2 + (17-18)^2 + (16-18)^2]}{4} = 3.0 \end{aligned}$$

La statistique de Wald devient :

$$(\hat{\theta}) / \text{Var}(\hat{\theta})^{1/2} = 18 / (19.575 + 3.0)^{1/2} = 3.788$$

Quand on compare cette valeur à la valeur critique de 95 % d'une distribution de t à 24 degrés de liberté (2,04), on conclut que les données suffisent à affirmer que le rendement moyen des hommes est différent de celui des femmes à l'intérieur du pays A.

8.1.4 Analyse des données de l'EIACA sur SPSS

La présente section donne des exemples élémentaires d'analyses qu'on peut effectuer en utilisant les poids d'échantillonnage et les notes dont il a été question dans les sections précédentes. Elle fournit également des détails sur certains programmes SPSS servant à mener ce genre d'analyse et sur les résultats de ces analyses. Les analyses présentées ici sont de nature simple. Les programmes calculent le pourcentage de répondants dans certains sous-groupes, le rendement moyen de ces groupes et les erreurs-types correspondantes (racine carrée de la variance d'erreur totale) pour donner des statistiques sous forme de pourcentages et de moyennes.

Dans nos exemples, nous utilisons des macros rédigées en SPSS pouvant servir à effectuer toutes les analyses décrites dans la présente section. Il s'agit de méthodes générales qu'on peut utiliser à plusieurs fins, à condition de posséder une connaissance de base du macrolangage SPSS. Si vous possédez une certaine expérience de la programmation avec ce progiciel statistique, vous serez en mesure d'apporter les modifications nécessaires aux macros pour obtenir les résultats souhaités.

Les macros SPSS

Les quatre macros SPSS disponibles sont décrites ci-dessous :

JACKMEAN.SPS

On peut utiliser cette macro SPSS pour calculer des pourcentages pondérés de répondants à l'intérieur de groupes définis et leur moyenne à l'égard d'une valeur continue donnée. Cette macro calcule également les variances d'échantillonnage JRR pour les pourcentages et les estimations moyennes. La variable peut être n'importe quelle variable continue contenue dans le fichier.

JACKMEANPV.SPS

On peut utiliser cette macro en SPSS pour calculer des pourcentages pondérés de répondants à l'intérieur de groupes définis et leurs notes de rendement moyen sur une échelle de rendement en utilisant des valeurs plausibles. Cette macro utilise des valeurs plausibles pour calculer des notes de rendement moyen. Elle calcule également les variances d'échantillonnage selon la technique de rééchantillonnage jackknife (JRR) pour les pourcentages de répondants à l'intérieur de groupes donnés, ainsi que les variances JRR et les variances d'imputation pour les notes de rendement moyen. On utilise cette macro uniquement lorsqu'on utilise plusieurs valeurs plausibles dans les analyses.

JACKREG.SPS

On peut utiliser cette macro en SPSS pour calculer des coefficients de régression linéaire et les erreurs-types correspondantes à l'intérieur de groupes définis. On peut l'utiliser avec n'importe quelle variable de l'analyse, mais elle n'utilise pas les valeurs plausibles.

JACKREGPV.SPS

On peut utiliser cette macro SPSS pour calculer des coefficients de régression linéaire et les erreurs-types correspondantes lorsqu'on utilise des valeurs plausibles comme variables dépendantes à l'intérieur de groupes définis.

Moyennes et pourcentages en l'absence de valeurs plausibles

La présente section propose un exemple de code SPSS pouvant servir à calculer les erreurs-types pour des moyennes et des pourcentages de variables autres que des valeurs plausibles. Ce code est fourni sous forme d'un programme SPSS appelé **JACKMEAN.SPS** qui calcule les pourcentages de répondants à l'intérieur de sous-groupes définis par un ensemble de variables de classification, les erreurs-types de ces pourcentages, les moyennes pour les groupes à l'égard d'une variable choisie et les erreurs-types de ces moyennes. Les erreurs-types calculées par cette macro SPSS tiennent compte du plan d'échantillonnage de l'EIACA.

Si vous utilisez cette macro, vous devez spécifier un ensemble de variables de classification, une variable d'analyse, le nombre de poids de rééchantillonnage (si ce nombre est le même pour différents pays, vous pouvez fusionner les ensembles de données des pays; sinon effectuez l'analyse pays par pays), les poids de rééchantillonnage et le poids de population à utiliser pour l'analyse. Vous devez également spécifier le fichier de données qui contient les données à traiter.

Pour utiliser efficacement la macro, vous devez connaître la syntaxe SPSS de base. Vous devez d'abord inclure la macro dans le fichier de programme où vous l'utiliserez. Si vous travaillez en mode différé, vous devez exécuter la macro pour chaque lot. Si vous utilisez le SPSS en mode interactif, vous devez exécuter la macro une fois au début de la session; elle restera active pendant toute la session. Si vous mettez fin à la session ou si vous la relancez plus tard, vous devez exécuter la macro de nouveau. Une fois la macro comprise dans une session donnée, évitez d'utiliser le mot « JACKMEAN » dans ce programme, ce qui aurait pour effet d'exécuter la macro.

Cette macro comporte les paramètres suivants :

INFILE Nom du fichier de données contenant les variables nécessaires à l'analyse (si l'emplacement du chemin est compris dans le nom du fichier, le nom du fichier doit figurer entre guillemets). Inclure uniquement les cas d'intérêt pour l'analyse (par ex., il faut exclure les répondants pour lesquels il manque des variables avant d'exécuter la macro).

CVAR Liste des variables à utiliser pour classer les répondants dans le fichier de données. Il peut s'agir d'une seule variable ou d'une liste de variables. Il est recommandé de toujours inclure la variable correspondant au pays. En l'occurrence, il fallait spécifier au moins une variable (par ex., CNTRID).

DVAR Variable pour laquelle il faut calculer des moyennes. Une seule variable doit être inscrite ici.

NJKZ Indication du nombre de poids de rééchantillonnage établis dans le fichier de données. Si vous travaillez avec les données d'un seul pays, vous devez appliquer à l'argument NJKZ autant de poids de rééchantillonnage qu'il en faut dans le pays. (Si vous travaillez avec

les données de plus d'un pays, assurez-vous que tous les ensembles de données comportent le même nombre de poids de rééchantillonnage.)

RPWT Poids de rééchantillonnage compris dans les fichiers de données, généralement REPLIC01 à REPLIC30. Les poids de rééchantillonnage doivent être spécifiés comme suit : « REPLIC01 TO REPLIC30 ».

WGT Poids d'échantillonnage à utiliser dans l'analyse, généralement POPWT.

La manière la plus simple d'exécuter la macro consiste à utiliser la notation SPSS courante servant à appeler des macros. Il s'agit d'inscrire le nom de la macro, suivi par la liste correspondante d'arguments nécessaires à l'analyse, chacun étant séparé par une barre oblique. Par exemple, si vous exécutez la macro en utilisant le code suivant :

```
Include "c:\jackmean.sps".
```

```
Jackmean
```

```
Infile = temp           /  
Cvar = cntrid           /  
Dvar = d43              /  
Njkz = 30               /  
Rpwt = replic01 to replic30 /  
Wgt = popwt.
```

Elle calcule la moyenne du revenu personnel (D43) et l'erreur-type correspondante, à l'intérieur de chaque pays, en prenant la variable POPWT comme poids d'échantillonnage. Les données sont lues à partir du fichier système TEMP.

Le fichier qui contient ces résultats est appelé FINAL et est conservé dans le répertoire par défaut utilisé par le SPSS. Les variables contenues dans ce fichier sont les suivantes :

Variables de classification

Chacune des variables de classification est conservée dans le fichier résultant. Il y a une seule occurrence par combinaison spécifique de catégories de variables de classification.

Variable de pondération

Cette variable contient l'estimation de la taille de population des groupes définis par combinaison spécifique de catégories de variables de classification. Dans l'exemple ci-dessus, cette variable est appelée POPWT.

MNX

Cette variable contient les moyennes de la variable DVAR pour les groupes définis par la combinaison correspondante de catégories de variables de classification.

MNX_SE

Cette variable contient les erreurs-types des valeurs MNX calculées selon la méthode jackknife.

PCT

Cette variable contient les pourcentages de personnes dans les groupes pour la dernière variable de classification inscrite, à l'intérieur de la combinaison spécifiée des catégories définies à l'origine par les groupes. Dans notre exemple, on obtiendrait les pourcentages de répondants par pays.

PCT_SE

Cette variable contient les erreurs-types de PCT calculées selon la méthode jackknife.

On peut alors imprimer le fichier obtenu à l'aide de cette macro en utilisant la méthode SPSS de son choix. En voici un exemple :

```
get file = "x:\IALSS\IALSSDATA.sav"
  / keep = cntrid gendaa2 d43 popwt replic01 to replic30 .

select if (gendaa2=1 or gendaa2=2) and not(missing(d43)).

save outfile = respondent.

include "c:\IALSS\jackmean.sps".

jackmean infile= respondent           /
  cvar = cntrid gendaa2               /
  dvar = d43                          /
  njkz = 30                           /
  rpwt = replic01 to replic30        /
  wgt = popwt.

print formats   cntrid gendaa2 n (F6.0) popwt (f10.0) mnx mnx_se pct pct_se (f8.2).

report format=list automatic / var = cntrid gendaa2 n popwt mnx mnx_se pct pct_se.
```

CNTRID	Gender of Respondent	N	POPWT	MNX	MNX_SE	PCT	PCT_SE
22	1	170	109777	70949.93	7977.68	60.39	3.03
22	2	190	72007	45423.49	6084.25	39.61	3.03

Moyennes et pourcentages en présence de valeurs plausibles

La présente section donne un exemple de code SPSS pouvant servir à calculer les erreurs-types pour les valeurs plausibles et les pourcentages moyens. Ce code est fourni sous forme d'une macro SPSS appelée **JACKMEANPV.SPS** qui calcule les pourcentages de répondants à l'intérieur de sous-groupes définis par un ensemble de variables de classification, les erreurs-types de ces pourcentages, les moyennes des groupes sur l'une des échelles de rendement utilisant les valeurs plausibles, ainsi que les erreurs-types de ces moyennes. Les erreurs-types calculées par cette macro SPSS tiennent compte du plan d'échantillonnage de l'ELCA et des composantes de la variance d'imputation.

Si vous utilisez cette macro, vous devez spécifier un ensemble de variables de classification, le nom et le nombre des valeurs plausibles, le nombre de poids de rééchantillonnage (si ce nombre est le même pour différents pays, vous pouvez fusionner les ensembles de données des pays, sinon effectuer l'analyse pays par pays), les poids de rééchantillonnage et le poids de population à utiliser pour l'analyse. Vous devez également spécifier le fichier de données qui contient les données à traiter.

Pour utiliser efficacement la macro, vous devez connaître la syntaxe SPSS de base. Vous devez d'abord inclure la macro dans le fichier de programme où vous l'utiliserez. Si vous travaillez en mode différé, vous devez exécuter la macro pour chaque lot. Si vous utilisez le SPSS en mode interactif, vous devez exécuter la macro une fois au début de la session; elle restera active pendant toute la session. Si vous mettez fin à la session ou si vous la relancez plus tard, vous devez exécuter la macro de nouveau. Une fois la macro comprise dans une session donnée, évitez d'utiliser le mot « JACKMEANPV » dans ce programme, ce qui aurait pour effet d'exécuter la macro.

Cette macro comporte les paramètres suivants :

INFILE Nom du fichier de données contenant les variables nécessaires à l'analyse (si l'emplacement du chemin est compris dans le nom du fichier, le nom du fichier doit figurer entre guillemets). Inclure uniquement les cas d'intérêt pour l'analyse (par ex., il faut exclure les répondants pour lesquels il manque des variables avant d'exécuter la macro).

CVAR Liste des variables à utiliser pour classer les répondants dans le fichier de données. Il peut s'agir d'une seule variable ou d'une liste de variables. Il est recommandé de toujours inclure la variable correspondant au pays. En l'occurrence, il fallait spécifier au moins une variable (par ex., CNTRID).

PVS Valeurs plausibles à utiliser dans l'analyse. Les valeurs plausibles doivent être spécifiés comme suit : « Valeur plausible 1 TO Valeur plausible 5 », par exemple « PROSE1 TO PROSE5 ». Dans la plupart des cas, vous voudrez utiliser les cinq valeurs plausibles, mais le programme fonctionne également lorsqu'on en spécifie moins. Utilisez toujours au moins deux valeurs plausibles.

NPV Nombre de valeurs plausibles à utiliser pour l'analyse. En général, vous voudrez utiliser les cinq valeurs plausibles mais, dans certains cas, vous pouvez en utiliser moins (voir PVS ci-dessus).

NJKZ Indication du nombre de poids de rééchantillonnage établis dans le fichier de données. Si vous travaillez avec les données d'un seul pays, vous devez appliquer à l'argument NJKZ autant de poids de rééchantillonnage qu'il en faut dans le pays. (Si vous travaillez avec les données de plus d'un pays, assurez-vous que tous les ensembles de données comportent le même nombre de poids de rééchantillonnage.)

RPWT Poids de rééchantillonnage compris dans les fichiers de données, généralement REPLIC01 à REPLIC30. Les poids de rééchantillonnage doivent être spécifiés comme suit : « REPLIC01 TO REPLIC30 ».

WGT Poids d'échantillonnage à utiliser dans l'analyse, généralement POPWT.

La manière la plus simple d'exécuter la macro consiste à utiliser la notation SPSS courante servant à appeler des macros. Il s'agit d'inscrire le nom de la macro, suivi par la liste correspondante d'arguments nécessaires à l'analyse, chacun étant séparé par une barre oblique. Par exemple, si vous exécutez la macro en utilisant le code suivant :

```
Include "c:\jackmeanpv.sps".
```

```
Jackmeanpv
```

```
      Infile = temp           /  
      Cvar  = cntrid gendaa2  /  
      PVS   = PROSE1 to PROSE5 /  
      NPV   = 5                /  
      Njkz  = 30               /  
      Rpwt  = replic01 to replic30 /  
      Wgt   = popwt.
```

Elle calcule le rendement moyen en compréhension de textes suivis et l'erreur-type correspondante pour les hommes et les femmes à l'intérieur de chaque pays, en utilisant cinq valeurs plausibles et la variable POPWT comme poids d'échantillonnage. Elle calcule également les pourcentages d'hommes et de femmes à l'intérieur du pays et les erreurs-types correspondantes. Les données sont lues à partir du fichier système TEMP.

Le fichier qui contient ces résultats est appelé FINAL et est conservé dans le répertoire par défaut utilisé par le SPSS. Les variables contenues dans ce fichier sont les suivantes :

Variables de classification

Chacune des variables de classification est conservée dans le fichier résultant. Il y a une seule occurrence par combinaison spécifique de catégories de variables de classification.

Variable de pondération

Cette variable contient l'estimation de la taille de population des groupes définis par combinaison spécifique de catégories de variables de classification. Dans l'exemple ci-dessus, cette variable est appelée POPWT.

N

Cette variable contient le nombre de cas dans les groupes définis par combinaison spécifique de catégories de variables de classification.

MNX

Cette variable contient les moyennes de la première valeur plausible pour les groupes définis par la combinaison correspondante de catégories de variables de classification.

MNX_SE

Cette variable contient les erreurs-types de la moyenne pour la première valeur plausible pour les groupes calculée selon la méthode jackknife. Elle ne comprend pas l'erreur d'imputation.

PCT

Cette variable contient les pourcentages de personnes dans les groupes pour la dernière variable de classification inscrite, à l'intérieur de la combinaison spécifiée des catégories définies à l'origine par les groupes. Dans notre exemple, il s'agit du pourcentage d'hommes et de femmes à l'intérieur de chaque pays.

PCT_SE

Cette variable contient les erreurs-types de la variable PCT calculées selon la méthode jackknife.

MNPV

Cette variable contient les moyennes des valeurs plausibles pour les groupes définis par la combinaison correspondante de catégories de variables de classification.

MNPV_SE

Cette variable contient les erreurs-types pour la moyenne des valeurs plausibles pour les groupes calculée selon la méthode jackknife. Elle comprend les composantes échantillonnage et imputation.

On peut alors imprimer le fichier obtenu à l'aide de cette macro en utilisant la méthode SPSS de son choix. En voici un exemple :

```
get file = "x:\IALSS\IALSSdata.sav"
  / keep = cntrid gendaa2 popwt replic01 to replic30 prosel to prose5.

select if (gendaa2=1 or gendaa2=2) .

save outfile = respondent.

include "c:\IALSS\jackmeanpv.sps".

jackmeanpv      infile= respondent      /
                 cvar = cntrid gendaa2  /
                 pvs = prosel to prose5 /
                 npv=5                  /
                 njkz = 30               /
                 rpwt = replic01 to replic30 /
                 wgt = popwt.

print formats cntrid gendaa2 (F2.0) n (F4.0) popwt (f7.0)  mnpv  mnpv_se mnx mnx_se
pct pct_se (f6.2).

report format=list automatic margin(1,255)
  / var = cntrid gendaa2 n popwt mnpv mnpv_se mnx mnx_se pct pct_se.
```

CNTRID	Gender of Respondent	N	POPWT	MNPV	MNPV_SE	MNX	MNX_SE	PCT	PCT_SE
22	1	1605	1179970	230.41	1.11	230.39	.99	49.53	.06
22	2	3196	1202504	226.13	.99	226.64	.91	50.47	.06

Coefficients de régression en l'absence de valeurs plausibles

La présente section donne un exemple de code SPSS pouvant servir à calculer des coefficients de régression linéaire et les erreurs-types correspondantes. Ce code est fourni sous forme d'une macro SPSS appelée **JACKREG.SPS** qui calcule la corrélation multiple entre les variables dépendantes et indépendantes spécifiées, ainsi que les coefficients de régression et les erreurs-types correspondantes. Les erreurs-types calculées par cette macro SPSS tiennent compte du plan d'échantillonnage de l'ELCA.

Si vous utilisez cette macro, vous devez spécifier un ensemble de variables de classification, les variables dépendantes et indépendantes, le nombre de poids de rééchantillonnage (si ce nombre est le même pour différents pays, vous pouvez fusionner les ensembles de données des pays, sinon effectuer l'analyse pays par pays), les poids de rééchantillonnage et le poids de population à utiliser pour l'analyse. Vous devez également spécifier le fichier de données qui contient les données à traiter.

Pour utiliser efficacement la macro, vous devez connaître la syntaxe SPSS de base. Vous devez d'abord inclure la macro dans le fichier de programme où vous l'utiliserez. Si vous travaillez en mode différé, vous devez exécuter la macro pour chaque lot. Si vous utilisez le SPSS en mode interactif, vous devez exécuter la macro une fois au début de la session; elle restera active pendant toute la session. Si vous mettez fin à la session ou si vous la relancez plus tard, vous devez exécuter la macro de nouveau. Une fois la macro comprise dans une session donnée, évitez d'utiliser le mot « JACKREG » dans ce programme, ce qui aurait pour effet d'exécuter la macro.

Cette macro comporte les paramètres suivants :

INFILE Nom du fichier de données contenant les variables nécessaires à l'analyse (si l'emplacement du chemin est compris dans le nom du fichier, le nom du fichier doit figurer entre guillemets). Inclure uniquement les cas d'intérêt pour l'analyse (par ex., il faut exclure les répondants pour lesquels il manque des variables avant d'exécuter la macro).

CVAR Liste des variables à utiliser pour classer les répondants dans le fichier de données. Il peut s'agir d'une seule variable ou d'une liste de variables. Il est recommandé de toujours inclure la variable correspondant au pays. En l'occurrence, il fallait spécifier au moins une variable (par ex., CNTRID).

XVAR Liste de variables indépendantes (au moins une) qui, selon le modèle de régression linéaire, servent de prédicteurs de la variable dépendante spécifiée dans la variable DVAR. Il peut s'agir de variables continues ou catégoriques ou de tout autre type de variable codée.

DVAR Variable dépendante qui, selon le modèle de régression, est prévue par la ou les variables spécifiées par le paramètre XVAR. Une seule variable doit être inscrite ici.

NJKZ Indication du nombre de poids de rééchantillonnage établis dans le fichier de données. Si vous travaillez avec les données d'un seul pays, vous devez appliquer à l'argument NJKZ autant de poids de rééchantillonnage qu'il en faut dans le pays. (Si vous travaillez avec les données de plus d'un pays, assurez-vous que tous les ensembles de données comportent le même nombre de poids de rééchantillonnage.)

RPWT Poids de rééchantillonnage compris dans les fichiers de données, généralement REPLIC01 à REPLIC30. Les poids de rééchantillonnage doivent être spécifiés comme suit : « REPLIC01 TO REPLIC30 ».

WGT Poids d'échantillonnage à utiliser dans l'analyse, généralement POPWT.

La manière la plus simple d'exécuter la macro consiste à utiliser la notation SPSS courante servant à appeler des macros. Il s'agit d'inscrire le nom de la macro, suivi par la liste correspondante d'arguments nécessaires à l'analyse, chacun étant séparé par une barre oblique. Par exemple, si vous exécutez la macro en utilisant le code suivant :

```
Include "c:\jackreg.sps".
```

```
Jackreg
```

```
Infile = temp           /  
Cvar = cntrid           /  
Xvar = regsex           /  
Dvar = d43              /  
Njkz = 30               /  
Rpwt = replic01 to replic30 /  
Wgt = popwt.
```

Elle calcule l'équation de régression pour la variable REGSEX comme prédicteur du revenu personnel. Les données sont lues à partir du fichier système TEMP.

Le fichier qui contient ces résultats est appelé REG et est conservé dans le répertoire par défaut utilisé par le SPSS. Les variables contenues dans ce fichier sont les suivantes :

Variables de classification

Chacune des variables de classification est conservée dans le fichier résultant. Il y a une seule occurrence par combinaison spécifique de catégories de variables de classification.

Mult_RSQ

Carré du coefficient de corrélation multiple du modèle.

SS_Res, SS_Reg, SS_Total

Résidu, régression et somme totale des carrés du modèle à l'intérieur de chaque groupe défini par les variables de classification.

Coefficients de régression et erreurs-types (B## et B##.SE)

Il s'agit des coefficients de régression pour chacune des variables prédictives du modèle et des erreurs-types correspondantes selon la méthode jackknife. Le coefficient zéro (B00) est la coordonnée à l'origine du modèle. Les autres coefficients reçoivent un numéro d'ordre, en commençant par 01. Ce numéro correspond à l'ordre des variables sur la liste des variables spécifiées dans le paramètre XVAR.

On peut alors imprimer le fichier obtenu à l'aide de cette macro en utilisant la méthode SPSS de son choix. En voici un exemple :

```

get file = "x:\IALSS\IALSSdata.sav"
  / keep = cntrid gendaa2 d43 popwt replic01 to replic30.

select if (gendaa2=1 or gendaa2=2) .    compute regsex = gendaa2 - 1.

save outfile = respondent.

include "c:\IALSS\jackreg.sps".

jackreg          infile= respondent          /
                 cvar = cntrid              /
                 xvar = regsex              /
                 dvar = d43                 /
                 njkz = 30                  /
                 rpwt = replic01 to replic30 /
                 wgt = popwt.

print formats    cntrid (F2.0) n (F4.0) mult_RSQ (f5.3)
                 SS_Total SS_Reg SS_Res (F10.0) B00 B00.SE B01 B01.SE (f6.2) .

report format=list automatic margin(1,255)
  / var = cntrid n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE .

```

CNTRID	N	MULT_RSQ	SS_TOTAL	SS_REG	SS_RES	B00	B00.SE	B01	B01.SE
22	360	.023	1.2E+15	2.8E+13	1.18E+15	70950	7977.7	-25526	9560.6

Dans cet exemple, on crée la variable REGSEX en soustrayant un de la variable GENDAA2. Par conséquent, les hommes reçoivent un code de 0 et les femmes, un code de 1 à l'égard de cette variable. Dans ce modèle, on utilise la variable REGSEX pour prévoir les valeurs de la variable D43 (revenu personnel). Le modèle devient :

Revenu personnel = 70950(7978) pour les hommes,
 Revenu personnel = 70950(7978) – 25526(9561) pour les femmes.

Les nombres entre parenthèses sont les erreurs-types. Cela signifie que les femmes ont, en moyenne, un revenu personnel qui est de 25 526 \$ inférieur à celui des hommes et que 9 561 \$ est l'erreur-type liée à cette estimation.

Coefficients de régression en présence de valeurs plausibles

La présente section donne un exemple de code SPSS pouvant servir à calculer des coefficients de régression linéaire en prenant des valeurs plausibles comme variable dépendante et les erreurs-types correspondantes. Ce code est fourni sous forme d'une macro SPSS appelée **JACKREGPV.SPS** qui calcule la corrélation multiple moyenne entre les valeurs plausibles et les variables indépendantes spécifiées, ainsi que les coefficients de régression et les erreurs-types correspondantes. Les erreurs-types calculées par cette macro SPSS tiennent compte du plan d'échantillonnage de l'ELCA.

Si vous utilisez cette macro, vous devez spécifier un ensemble de variables de classification, les variables dépendantes et indépendantes, le nombre de poids de rééchantillonnage (si ce nombre est le même pour différent pays, vous pouvez fusionner les ensembles de données des pays, sinon effectuer l'analyse pays par pays), les poids de rééchantillonnage et le poids de

population à utiliser pour l'analyse. Vous devez également spécifier le fichier de données qui contient les données à traiter.

Pour utiliser efficacement la macro, vous devez connaître la syntaxe SPSS de base. Vous devez d'abord inclure la macro dans le fichier de programme où vous l'utiliserez. Si vous travaillez en mode différé, vous devez exécuter la macro pour chaque lot. Si vous utilisez le SPSS en mode interactif, vous devez exécuter la macro une fois au début de la session; elle restera active pendant toute la session. Si vous mettez fin à la session ou si vous la relancez plus tard, vous devez exécuter la macro de nouveau. Une fois la macro comprise dans une session donnée, évitez d'utiliser le mot « JACKREGPV » dans ce programme, ce qui aurait pour effet d'exécuter la macro.

Cette macro comporte les paramètres suivants :

INFILE Nom du fichier de données contenant les variables nécessaires à l'analyse (si l'emplacement du chemin est compris dans le nom du fichier, le nom du fichier doit figurer entre guillemets). Inclure uniquement les cas d'intérêt pour l'analyse (par ex., il faut exclure les répondants pour lesquels il manque des variables avant d'exécuter la macro).

CVAR Liste des variables à utiliser pour classer les répondants dans le fichier de données. Il peut s'agir d'une seule variable ou d'une liste de variables. Il est recommandé de toujours inclure la variable correspondant au pays. En l'occurrence, il fallait spécifier au moins une variable (par ex., CNTRID).

XVAR Liste de variables indépendantes (au moins une) qui, selon le modèle de régression linéaire, servent de prédicteurs de la variable dépendante spécifiée par les valeurs plausibles. Il peut s'agir de variables continues ou catégoriques ou de tout autre type de variable codée.

ROOTPV Préfixe servant à identifier les valeurs plausibles pour l'échelle de rendement qui nous intéresse. Par exemple, la racine des valeurs plausibles en compréhension de textes suivis est « PROSE ».

NPV Nombre de valeurs plausibles à utiliser pour l'analyse. En général, vous voudrez utiliser les cinq valeurs plausibles mais, dans certains cas, vous pouvez en utiliser moins (voir PVS ci-dessus).

NJKZ Indication du nombre de poids de rééchantillonnage établis dans le fichier de données. Si vous travaillez avec les données d'un seul pays, vous devez appliquer à l'argument NJKZ autant de poids de rééchantillonnage qu'il en faut dans le pays. (Si vous travaillez avec les données de plus d'un pays, assurez-vous que tous les ensembles de données comportent le même nombre de poids de rééchantillonnage.)

RPWT Poids de rééchantillonnage compris dans les fichiers de données, généralement REPLIC01 à REPLIC30. Les poids de rééchantillonnage doivent être spécifiés comme suit : « REPLIC01 TO REPLIC30 ».

WGT Poids d'échantillonnage à utiliser dans l'analyse, généralement POPWT.

La manière la plus simple d'exécuter la macro consiste à utiliser la notation SPSS courante servant à appeler des macros. Il s'agit d'inscrire le nom de la macro, suivi par la liste correspondante d'arguments nécessaires à l'analyse, chacun étant séparé par une barre oblique. Par exemple, si vous exécutez la macro en utilisant le code suivant :

```
Include "c:\jackregpv.sps".
```

```
Jackregpv
```

```
      Infile  = temp           /
      Cvar    = cntrid         /
      Xvar    = regsex         /
      Rootpv  = Prose          /
      NPV     = 5              /
      Njkz    = 30             /
      Rpwt    = replic01 to replic30 /
      Wgt     = popwt.
```

Elle calcule l'équation de régression pour la variable REGSEX comme prédicteur des valeurs plausibles en compréhension de textes suivis. Les données sont lues à partir du fichier système TEMP.

Le fichier qui contient ces résultats est appelé REG et est conservé dans le répertoire par défaut utilisé par le SPSS. Les variables contenues dans ce fichier sont les suivantes :

Variables de classification

Chacune des variables de classification est conservée dans le fichier résultant. Il y a une seule occurrence par combinaison spécifique de catégories de variables de classification.

Mult_RSQ

Carré du coefficient de corrélation multiple du modèle.

SS_Res, SS_Reg, SS_Total

Résidu, régression et somme totale des carrés du modèle à l'intérieur de chaque groupe défini par les variables de classification.

Coefficients de régression et Erreurs-types (B## et B##.SE)

Il s'agit des coefficients de régression pour chacune des variables prédictives du modèle et des erreurs-types correspondantes selon la méthode jackknife (avec les composantes échantillonnage et imputation). Le coefficient zéro (B00) est la coordonnée à l'origine du modèle. Les autres coefficients reçoivent un numéro d'ordre, en commençant par 01. Ce numéro correspond à l'ordre des variables sur la liste de variables spécifiées dans le paramètre XVAR.

On peut alors imprimer le fichier obtenu à l'aide de cette macro en utilisant la méthode SPSS de son choix. En voici un exemple :

```

get file = "x:\IALSS\IALSSdata.sav"
  / keep = cntrid gendaa2 popwt replic01 to replic30 prose1 to prose5.

select if (gendaa2=1 or gendaa2=2) . compute regex = gendaa2 - 1.

save outfile = respondent.

include "c:\IALSS\jackregpv.sps".

jackregpv   infile      = respondent
            cvar        = cntrid
            xvar        = regex
            rootpv      = prose
            npv         = 5
            njkz        = 30
            rpwt        = replic01 to replic30
            wgt         = popwt.

print formats      cntrid (F2.0) n (F4.0) mult_RSQ (f5.3)
                  SS_Total SS_Reg SS_Res (F12.0) B00 B00.SE B01 B01.SE (f6.2).

report format=list automatic margin(1,255)
  / var = cntrid n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE .

```

CNTRID	N	MULT_RSQ	SS_TOTAL	SS_REG	SS_RES	B00	B00.SE	B01	B01.SE
22	4801	.002	4395153088	10952856	4384200232	230.41	1.11	-4.27	1.50

Dans cet exemple, on crée la variable REGSEX en soustrayant un de la variable GENDAA2. Par conséquent, les hommes reçoivent un code de 0 et les femmes, un code de 1 à l'égard de cette variable. Dans ce modèle, on utilise la variable REGSEX pour prévoir les valeurs des valeurs plausibles en compréhension de textes suivis (PROSE). Le modèle devient :

Prose = 230,41 (1,11) pour les hommes,
 Prose = 230,41 (1,11) – 4,27(1,50) pour les femmes.

Les chiffres entre parenthèses sont les erreurs-types. Cela signifie que les femmes ont, en moyenne, une note en compréhension de textes suivis qui est de 4,27 inférieure à celle des hommes et que 1,50 est l'erreur-type liée à cette estimation.

8.1.5 Analyse des données de l'EIACA sur SAS

La présente section donne des exemples élémentaires d'analyses qu'on peut effectuer en utilisant les poids d'échantillonnage et les notes dont il a été question dans les sections précédentes. Elle fournit également des détails sur un programme SAS donné servant à mener ce genre d'analyse et sur les résultats de ces analyses. Les analyses présentées ici sont de nature simple. Le programme calcule le pourcentage de répondants dans certains sous-groupes, le rendement moyen de ces groupes, les nombres pondérés de répondants dans certains groupes, les centiles estimatifs de ces groupes, les coefficients de régression et coefficients de régression logistique ainsi que les erreurs-types correspondantes (racine carrée de la variance d'erreur totale).

Dans nos exemples, nous utilisons une macro rédigée en SAS pouvant servir à effectuer toutes les analyses décrites dans la présente section. Il s'agit de méthodes générales qu'on peut utiliser à plusieurs fins, à condition de posséder une connaissance de base du macrolangage SAS. Si vous possédez une certaine expérience de la programmation avec ce progiciel statistique, vous serez en mesure d'apporter les modifications nécessaires aux macros pour obtenir les résultats souhaités.

La macro SAS

La seule macro SAS disponible est décrite comme suit : **STATTOOL.SAS**

On peut utiliser ce macroprogramme en SAS pour calculer plusieurs statistiques : moyennes, centiles, fréquences, dénombrements, écarts (différences) et régressions (régression linéaire type, régression logistique et régression multinomiale). Ces statistiques sont calculées à l'intérieur de groupes définis en tenant compte des poids d'échantillonnage. Cette macro calcule également les erreurs-types JRR avec les composantes échantillonnage et imputation.

Analyses de base : moyennes, pourcentages, dénombrements, centiles, coefficients de régression et erreurs-types correspondantes

La présente section donne un exemple de code SAS pouvant servir à calculer des moyennes, des pourcentages, des dénombrements, des centiles, des coefficients de régression et les erreurs-types correspondantes pour n'importe quel type de variable (qu'il s'agisse ou non de valeurs plausibles). Ce code est fourni sous forme d'une macro SAS appelée STATTOOL.SAS qui calcule ces statistiques pour des répondants à l'intérieur de sous-groupes définis par un ensemble de n'importe quelle variable de classification (fondée ou non sur des valeurs plausibles). Les erreurs-types calculées par cette macro SAS tiennent compte des composantes échantillonnage et imputation.

Si vous utilisez cette macro, vous devez spécifier un ensemble de variables de classification, une analyse variable, le nombre de poids de rééchantillonnage (si ce nombre est le même pour différents pays, vous pouvez fusionner les ensembles de données des pays; sinon effectuez l'analyse pays par pays), les poids de rééchantillonnage et le poids de population à utiliser pour l'analyse. Vous devez également spécifier le fichier de données qui contient les données à traiter.

Pour utiliser efficacement la macro, vous devez connaître la syntaxe SAS de base. Vous devez d'abord inclure la macro dans le fichier de programme où vous l'utiliserez. Si vous travaillez en mode différé, vous devez exécuter la macro pour chaque lot. Si vous utilisez le SAS en mode interactif, vous devez exécuter la macro une fois au début de la session; elle restera active pendant toute la session. Si vous mettez fin à la session ou si vous la relancez plus tard, vous devez exécuter la macro de nouveau. Une fois la macro comprise dans une session donnée, évitez d'utiliser la chaîne « %STATTOOL » dans ce programme, ce qui aurait pour effet d'exécuter la macro.

Cette macro comporte les paramètres suivants :

WGT Poids d'échantillonnage à utiliser dans l'analyse, généralement POPWT.

RWGT Racine des variables spécifiant les poids de rééchantillonnage compris dans les fichiers de données, généralement REPLIC01 à REPLIC30. Les poids de rééchantillonnage doivent être spécifiés comme suit : « REPLIC ».

NREP Indication du nombre de poids de rééchantillonnage établis dans le fichier de données. Si vous travaillez avec les données d'un seul pays, vous devez appliquer à l'argument NREP autant de poids de rééchantillonnage qu'il en faut dans le pays. (Si vous travaillez avec les données de plus d'un pays, assurez-vous que tous les ensembles de données comportent le même nombre de poids de rééchantillonnage.)

NPV Nombre de valeurs plausibles à utiliser pour l'analyse. En général, vous voudrez utiliser les cinq valeurs plausibles mais, dans certains cas, vous pouvez en utiliser moins.

STUDY Nom de l'étude (ELCA).

CNTRYNO Identificateur du pays.

INFILE Nom du fichier de données contenant les variables nécessaires à l'analyse. (Si l'emplacement du chemin est compris dans le nom du fichier, le nom du fichier doit figurer entre guillemets.) Inclure uniquement les cas d'intérêt pour l'analyse (par ex., il faut exclure les répondants pour lesquels il manque des variables avant d'exécuter la macro).

METHOD Indication de la statistique à produire. « METHOD = mean » calcule les moyennes de la variable d'intérêt. Vous pouvez également spécifier « crosstabs » pour les totalisations croisées, « perc » pour les centiles, « diff » pour les écarts (différences), « popest » pour les chiffres de population, « reg » pour la régression linéaire type, « logistic » pour la régression logistique et « multinomial » pour la régression logistique multinomiale.

DVAR Variable pour laquelle on veut calculer des moyennes. Une seule variable doit être inscrite ici. Inscrivez le nom de la variable ou seulement la racine si la variable d'intérêt est tirée d'un ensemble de valeurs plausibles.

DVARPV Paramètre indiquant si la variable DVAR est tirée ou non d'un ensemble de valeurs plausibles. Ce paramètre prend la valeur 1 si la variable d'intérêt est tirée d'un ensemble de valeurs plausibles et 0 dans le cas contraire.

BYVAR Liste des variables à utiliser pour classer les répondants dans le fichier de données. Il peut s'agir d'une seule variable ou d'une liste de variables. Ce paramètre définit les sous-groupes pour lesquels on demande des moyennes de la variable DVAR.

BYVARPV Paramètre indiquant si la ou les variables BYVAR sont tirées ou non d'un ensemble de valeurs plausibles. Ce paramètre prend la valeur 1 si la variable d'intérêt est tirée d'un ensemble de valeurs plausibles et 0 dans le cas contraire.

Outre ces paramètres, on peut en utiliser trois autres : CRITER1, CRITER2 et CRITER3. Chacun contient un seul énoncé de programmation SAS.

La manière la plus simple d'exécuter la macro consiste à utiliser la notation SAS courante servant à appeler des macros. Il s'agit d'inscrire le nom de la macro, suivi par la liste

correspondante d'arguments nécessaires à l'analyse, chacun étant séparé par une virgule. Par exemple, si vous exécutez la macro en utilisant le code suivant :

```
%include "c:\IALSS\stattool.sas";

%stattool (wgt = popwt,
          rwt = replic,
          nrep = 30,
          npv = 5,
          study = ALL,
          method = mean,
          infile = in,
          dvar = prose,
          dvarpv = 1,
          byvar = gendaa2 age3,
          byvarpv = 0 0);
```

Elle calcule le rendement moyen en compréhension de textes suivis en utilisant les cinq ensembles de valeurs plausibles et l'erreur-type correspondante, à l'intérieur de chaque groupe défini par la combinaison des catégories sexe et âge, en prenant la variable POPWT comme poids d'échantillonnage. Les données sont lues à partir du fichier système TEMP.

Le fichier qui contient ces résultats est appelé FINALB et est conservé dans le répertoire par défaut utilisé par le SAS. Il existe également un fichier HTML appelé FINALB, conservé dans le lecteur C de votre ordinateur, dans le répertoire TEMP. On peut facilement accéder à ce fichier en utilisant EXCEL à partir de MICROSOFT. Les variables contenues dans ce fichier sont les suivantes :

Variables de classification

Chacune des variables de classification est conservée dans le fichier résultant. Il y a une seule occurrence par combinaison spécifique de catégories de variables de classification.

ESTIMATE

Variable contenant les moyennes de la variable DVAR pour les groupes définis par les combinaisons correspondantes de catégories de variables de classification.

STANDARD ERROR

Variable contenant les erreurs-types des valeurs de l'estimation calculées selon la méthode jackknife, y compris les composantes échantillonnage et imputation.

PROB > |T|

Variable donnant la probabilité qu'une statistique de Student soit plus grande que la valeur absolue de l'estimation observée, à l'intérieur de la combinaison spécifiée des catégories définies à l'origine par les groupes.

En voici deux exemples :

```
libname in "C:\IALSS\data";
data in;set in.IALSSdata;run;

%stattool(wgt = popwt,
          rwt = replic,
          nrep = 30,
          = 5,
          ALL,
          npv
          study =
          method = mean,
          infile = in,
          dvar = prose,
          dvarpv = 1,
          byvar = gendaa2 age3,
          byvarpv = 0 0);
```

Étude : ELCA :,

Moyennes estimatives en compréhension de textes suivis par domaine et gendaa2 age3, en fonction de cinq ensembles de valeurs plausibles et 29 degrés de liberté

Abstraction faite du domaine

Obs	Domain	GENDAA2	AGE3	estimate	Standard Error	Prob > T
1	ELCA	1	1	236,764	1,81853	0
2	ELCA	1	2	231,754	1,87374	0
3	ELCA	1	3	197,439	4,25769	0
4	ELCA	2	1	236,945	2,05748	0
5	ELCA	2	2	225,062	1,08269	0
6	ELCA	2	3	192,060	2,77765	0

Dans cet exemple, XPROSE1 à XPROSE5 servent de variables de classification. Nous estimons le revenu personnel moyen par niveau de compréhension de textes suivis.

```
libname in "C:\IALSS\data";

data in;set in.ialssdata;run;

%stattool(wgt = popwt,
          rwgt = replic,
          nrep = 30,
          npv = 5,
          study = ALL,
          method = mean,
          infile = in,
          dvar = D43,
          dvarpv = 0,
          byvar = XPROSE,
          byvarpv = 1,
          criter1 = if d43 < 99999997 );
```

Étude : ELCA : si d43 < 99999997,,

Moyennes estimatives pour d43 par domaine et xprose, en fonction de cinq ensembles de valeurs plausibles et 29 degrés de liberté

Abstraction faite du domaine

Obs	Domain	xprose	estimate	Standard Error	Prob > T
1	ELCA	1	39799,22	6167,01	0,000000463
2	ELCA	2	66658,33	9378,51	0,000000081
3	ELCA	3	103063,21	27265,80	0,000724651
4	ELCA	4	172247,60	60242,53	0,007788203

On peut modifier le contenu de la « méthode » pour produire d'autres statistiques en inscrivant :

perc = pour produire des centiles,

diff = pour produire des écarts (différences),

popest = pour produire des chiffres de population,

reg = pour produire une régression linéaire type,

logistic = pour produire une régression logistique,

multinomial = pour produire un coefficient de régression logistique multinomiale.

Tous les autres paramètres restent identiques à ceux des exemples illustrés ci-dessus, les erreurs-types liées à l'échantillonnage et l'erreur de test étant dûment calculées pour chaque mesure.

8.2 Erreurs non liées à l'échantillonnage

Lorsqu'elles sont réparties sur un grand nombre d'observations, les erreurs non liées à l'échantillonnage qui surviennent au hasard ont peu d'incidence sur les estimations tirées de l'enquête. Toutefois, les erreurs qui surviennent systématiquement entraînent des distorsions dans ces estimations. Les responsables de l'enquête ont donc consacré beaucoup de temps et d'efforts à réduire les erreurs non liées à l'échantillonnage. Pour vérifier la qualité des données, on a pris des mesures d'assurance de la qualité à chaque étape du cycle de collecte et de traitement des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, la formation intensive des intervieweurs à l'égard des méthodes d'enquête et du questionnaire, l'observation des intervieweurs pour déceler des problèmes de conception du questionnaire ou des erreurs d'interprétation des directives, une procédure visant à réduire au minimum les erreurs de saisie des données, ainsi que des contrôles du codage et de la révision pour vérifier la logique de traitement.

Malgré ces efforts, aucune enquête n'est à l'abri d'une erreur non liée à l'échantillonnage. Voici un aperçu des principales sources de ce type d'erreur et de son incidence sur les données de l'EIACA.

8.2.1 Base de sondage

L'utilisation du Recensement de 2001 a permis d'assurer que la base de sondage de l'EIACA était aussi inclusive que possible et que les moindres exclusions seraient effectivement calculées dans le plan d'enquête global.

8.2.2 Non-réponse

L'incidence de la non-réponse sur les résultats d'une enquête constitue une importante source d'erreurs non liées à l'échantillonnage. Il peut s'agir d'une non-réponse partielle (défaut de répondre à une ou à quelques questions seulement) ou d'une non-réponse totale.

Il y a non-réponse totale lorsque l'intervieweur ne peut communiquer avec le répondant, qu'aucun membre du ménage ne peut fournir les renseignements demandés ou que le répondant refuse de participer à l'enquête. Le taux national de non-réponse à l'EIACA se situait autour de 34 %. Toutefois, l'analyse des caractéristiques des non-répondants à l'EIACA donne à entendre que ces derniers ne se concentrent pas dans un groupe donné (ainsi, la non-réponse semble avoir été aléatoire). On a réglé le problème de la non-réponse totale en rajustant le poids des ménages ayant répondu à l'enquête, compensant ainsi pour ceux qui n'ont pas répondu.

Dans la plupart des cas, il y a non-réponse partielle à une enquête lorsque le répondant ne comprend pas ou interprète mal une question, refuse de répondre à une question ou ne se rappelle pas les renseignements demandés. En général, la fréquence des non-réponses partielles à l'EIACA était faible.

8.2.3 Erreur dans la réponse

Il convient de mentionner d'autres sources possibles d'erreur non liée à l'échantillonnage qui sont propres à l'EIACA. D'abord, certains répondants ont pu trouver le test intimidant, ce qui peut avoir eu une incidence négative sur leur rendement. Contrairement aux études « habituelles », il existe de « bonnes » et de « mauvaises » réponses aux items de test de l'EIACA. De plus, pour de nombreux répondants, c'était la première fois qu'ils subissaient un test depuis bon nombre d'années. En outre, même si les intervieweurs n'imposaient pas de limite de temps pour répondre aux questions, le fait que quelqu'un observait et attendait la réponse a peut-être imposé une contrainte de temps involontaire. Par conséquent, même si les items ont été choisis avec soin pour correspondre étroitement à des tâches courantes, il est possible que les réponses au test ne révèlent pas pleinement les compétences en littératie des répondants en raison du contexte propre à un test. Enfin, comme il s'agissait d'un test, les répondants devaient accomplir les activités de manière complètement indépendante alors que, dans la réalité, on a souvent l'occasion de parcourir des textes imprimés avec des parents, des amis et des collègues. Il est donc possible que les compétences mesurées par l'enquête ne reflètent pas l'étendue complète des aptitudes de certains répondants dans un contexte ordinaire.

8.2.4 Notation

Dans le cadre de l'EIACA, une autre source possible d'erreur non liée à l'échantillonnage a trait à la notation des items du test, notamment ceux qui ont été notés sur une échelle (par exemple, les items qui exigeaient que les répondants rédigent une réponse). Pour réduire au minimum la fréquence des erreurs de notation, on a déployé des efforts particuliers, dont la centralisation de la notation et la vérification par échantillonnage.

9.0 Les tables de coefficients de variation

9.1 Seuils pour la diffusion des estimations pour l'enquête sur l'alphabétisation et les compétences des adultes

Le tableau ci-dessous fournit une indication de la précision des estimations des prévisions démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section 8.1.1. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 9.9.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 10 000 personnes possédant une caractéristique donnée dans les provinces de l'Atlantique est médiocre.

Veuillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 9.3 ainsi que l'exemple 4 à la section 9.4 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Régions	CV acceptable 0,0 à 16,5 %	CV médiocre 16,6 à 33,3 %	CV inacceptable > 33,3 %
Provinces de l'Atlantique	26 500 et plus	26 500 à < 6 600	moins de 6 600
Québec	93 100 et plus	93 100 à < 23 100	moins de 23 100
Ontario	218 500 et plus	218 500 à < 54 600	moins de 54 600
Provinces de l'Ouest	82 800 et plus	82 800 à < 20 500	moins de 20 500
Territoires du Nord	1 300 et plus	1 300 à < 300	moins de 300
Canada	134 900 et plus	134 900 à < 33 300	moins de 33 300

9.2 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75^e percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA).

Régions	Effet du plan	Taille de l'échantillon	Population
Provinces de l'Atlantique	1.82	4,682	1,890,046
Québec	1.79	4,166	5,994,042
Ontario	3.13	4,946	9,621,290
Provinces de l'Ouest	2.07	6,657	7,336,591
Territoires du Nord	1.67	2,587	59,872
Canada	3.42	23,038	24,901,841

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais ou au moyen des 30 poids de rééchantillonnage jackknife (voir la section 8.1). Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation **médiocre** pourrait devenir **acceptable** si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

Nota : Dans le cas des estimations relatives aux francophones des provinces de l'Ouest, les effets du plan d'échantillonnage sont particulièrement élevés. Les CV approximatifs concernant cette sous-population doivent être multipliés par un facteur de 1,7.

9.3 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, de femmes ayant terminé des études postsecondaires est plus fiable que le nombre estimé de ces femmes. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'un sexe ou d'un groupe d'âge particulier, on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}}/\hat{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de femmes et le numérateur, le nombre de femmes ayant terminé des études postsecondaires.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de femmes ayant terminé des études postsecondaires comparativement au nombre d'hommes ayant terminé des études postsecondaires, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement. Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à la sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

9.4 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'EIACA et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime que 5 369 686 femmes ont terminé des études postsecondaires. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA à la section 9.9.
- 2) L'agrégat estimé 5 369 686 ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 5 000 000.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 2,4 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 2,4 %. Le résultat selon lequel il y avait 5 369 686 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 7.4) femmes ayant terminé des études postsecondaires, peut être publié sans réserve.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime que $5\,369\,686 / 12\,668\,933 = 42,4\%$ des femmes ont terminé des études postsecondaires. Comment l'utilisateur détermine-t-il le coefficient de variation de cette

estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA à la section 9,9.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les femmes ayant déclaré leur niveau de scolarité), il faut utiliser à la fois le pourcentage (42,4 %) et la portion numérateur du pourcentage (5 369 686) pour déterminer le coefficient de variation.
- 3) Le numérateur, 5 369 686, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 5 000 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 40,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 2,1 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 2,1 %. Le résultat selon lequel 42,4 % des femmes ont terminé des études postsecondaires, peut être publié sans réserve *).

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que $5\,369\,686 / 12\,668\,933 = 42,4\%$ la proportion de femmes ayant terminé des études postsecondaires, par rapport à $5\,351\,237 / 12\,223\,350 = 43,8\%$ la proportion d'hommes ayant terminé des études postsecondaires. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation du CANADA de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes de 2,1 % et un CV de l'estimation pour les hommes de 2,1 %.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1 (femmes), \hat{X}_2 est l'estimation 2 (hommes) et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,424 - 0,438 = -0,014$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,424)(0,021)]^2 + [(0,438)(0,021)]^2} \\ &= \sqrt{(0,0000793) + (0,0000846)} \\ &= 0,013\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,013 / 0,014 = 0,93$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 93 %. La

différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 7 299 911 le nombre de personnes de l'Ontario qui lisent des journaux ou utilisent l'information qui y est contenue au moins une fois par semaine (G3A=1), par rapport à 4 561 049 personnes de l'Ontario qui lisent des livres ou utilisent l'information qui y est contenue au moins une fois par semaine (G3C=1). L'utilisateur est intéressé à comparer l'estimation des personnes qui lisent des journaux à celle des personnes qui lisent des livres sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de personnes de l'Ontario qui lisent des journaux ou utilisent l'information qui y est contenue au moins une fois par semaine. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de personnes de l'Ontario qui lisent des livres ou utilisent l'information qui y est contenue au moins une fois par semaine.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA à la section 9.9
- 3) Le numérateur de cette estimation de rapport est 7 299 911. Le chiffre qui s'en rapproche le plus est 7 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 0,9 %.
- 4) Le dénominateur de cette estimation de rapport est 4 561 049. Le chiffre qui s'en rapproche le plus est 5 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 1,9 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,009)^2 + (0,019)^2} \\ &= \sqrt{0,000081 + 0,000361} \\ &= 0,021 \end{aligned}$$

- 6) Le rapport obtenu entre les personnes de l'Ontario qui lisent des journaux ou utilisent l'information qui y est contenue et celles qui lisent des livres ou utilisent l'information qui y est contenue est 7 299 911 / 4 561 049, c'est-à-dire 1,60 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 7.4). Le coefficient de variation de cette estimation est 2,1 %, ce qui fait qu'on peut la diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les hommes de l'Ontario qui lisent des journaux ou utilisent l'information qui y est contenue au moins une fois par semaine et ceux qui lisent des livres ou utilisent l'information qui y est contenue au moins une fois par semaine s'établit à 2,08, tandis qu'il est de 1,30 pour les femmes. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord, calculez le coefficient de variation approximatif pour le rapport des hommes (\hat{R}_1) et le rapport des femmes (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport des hommes est 5,6 % et 4,6 % pour celui des femmes.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{R}_1 - \hat{R}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{R}_1 \alpha_1)^2 + (\hat{R}_2 \alpha_2)^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 2,08 - 1,30 = 0,78$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(2,08)(0,056)]^2 + [(1,30)(0,046)]^2} \\ &= \sqrt{(0,0136) + (0,00358)} \\ &= 0,13\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,13 / (0,78) = 0,167$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 16,7 %. L'estimation est considérée de qualité médiocre. Elle devrait être signalée par la lettre M (ou un quelconque identificateur similaire). Elle devrait être accompagnée d'une mise en garde avertissant les utilisateurs subséquents du niveau élevé d'erreur associé à l'estimation.

9.5 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation soient beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée de femmes ayant terminé des études postsecondaires (d'après l'exemple 2 à la section 9.4 serait calculé comme suit :

$$\hat{X} = 42,4 \% \text{ (ou exprimé sous forme de proportion } 0,424)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 2,1 \% \text{ (0,021 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$IC_{\hat{x}} = \{0,424 - (2) (0,424) (0,021), 0,424 + (2) (0,424) (0,021)\}$$

$$IC_{\hat{x}} = \{0,424 - 0,018, 0,424 + 0,018\}$$

$$IC_{\hat{x}} = \{0.406, 0.442\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 40,5 % et 44,2 % des femmes ont terminé des études postsecondaires.

9.6 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisés pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elle sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de femmes ayant terminé des études postsecondaires et la proportion d'hommes ayant terminé des études postsecondaires. D'après l'exemple 3 à la section 9.4, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,013. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,424 - 0,438}{0,013} = \frac{-0,014}{0,013} = -1,08$$

Puisque $t = -1,08$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est justifiée au niveau de signification de 5 %.

9.7 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'EIACA sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total d'heures consacrées à un programme d'études serait supérieur au coefficient de variation de la proportion correspondante du nombre de personnes inscrites à un

programme d'études. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable). Des coefficients de variation de telles estimations peuvent être calculés au moyen des 30 poids de rééchantillonnage jackknife (voir la section 8.1).

9.8 Coefficients de variation pour les estimations du niveau de compétence

Tel qu'il est expliqué à la section 8.1.3.5, les cinq valeurs plausibles servant à estimer les niveaux de compétence ajoutent à la variance une autre composante, appelée variance d'imputation. On a calculé un facteur servant à extrapoler les CV obtenus d'après les tables des CV pour les estimations du niveau de compétence. On a déterminé le facteur d'extrapolation en calculant d'abord l'apport en pourcentage de la variance d'imputation à la variance globale pour un grand nombre d'estimations du niveau de compétence, puis en choisissant le 75^e centile.

Règle 6 : Estimations du niveau de compétence

On calcule les CV approximatifs des estimations du niveau de compétence en multipliant la valeur obtenue d'après la table des CV par un facteur de 1,4. On utilise cette règle conjointement avec les règles 1 à 5 (section x). Pour les estimations comportant plus d'une variable, on utilise d'abord la règle 6 pour calculer le CV approximatif de chaque variable de niveau de compétence, puis on applique les règles 1 à 5 pour combiner les CV et calculer un CV global. En voici deux exemples :

Exemple 6.1 : Estimations de proportions avec des niveaux de compétence (règle 2 et règle 6)

Supposons qu'un utilisateur estime que $2\,470\,476 / 12\,230\,514 = 20,2\%$ des hommes se sont classés au niveau 1 en compréhension de textes suivis. Comment l'utilisateur détermine-t-il le coefficient de variation de cette proportion?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA à la section x.
- 2) Selon la règle 2, le CV obtenu d'après la table des CV est 3,8 %.
- 3) En multipliant 3,8 % par le facteur d'extrapolation de 1,4, on obtient un CV de 5,3 %.
- 4) Le coefficient de variation approximatif de la proportion est donc 5,3 %, et l'estimation peut être publiée sans réserve.

Exemple 6.2 : Estimations de rapports avec des niveaux de compétence (règle 6 et règle 4)

Supposons qu'un utilisateur estime que 4 965 226 personnes se sont classées au niveau 1 en compréhension de textes suivis, alors que 6 402 351 personnes n'ont pas terminé leurs études secondaires. L'utilisateur veut comparer l'estimation de ces deux quantités sous forme de rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de ce rapport?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de *personnes qui se sont classées au niveau 1 en compréhension de textes*

suivis. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de *personnes qui n'ont pas terminé leurs études secondaires*.

- 2) Reportez-vous à la table des coefficients de variation pour le Canada à la section x.
- 3) Le numérateur de cette estimation de rapport est 4 965 226. Le CV obtenu d'après les tables des CV pour le numérateur est 2,4 %. Comme le numérateur est un niveau de compétence, la règle 6 s'applique; on multiplie le CV par un facteur de 1,4, ce qui donne un CV approximatif pour le numérateur de 3,4 %.
- 4) Le dénominateur de cette estimation de rapport est 6 402 351. Le CV obtenu d'après les tables des CV pour le dénominateur est 2,1 %. Comme le dénominateur n'est pas un niveau de compétence, la règle 6 ne s'applique pas.
- 5) Le coefficient de variation approximatif de l'estimation de rapport est donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation respectifs de \hat{X}_1 et \hat{X}_2 . C'est-à-dire que :

$$\alpha_{\hat{R}} = \sqrt{(0,034)^2 + (0,021)^2} = 0,040$$

- 6) Le rapport $4\,965\,226 / 6\,402\,351 = 0,78$ a donc un coefficient de variation approximatif de 4,0 %, et l'estimation peut être publiée sans réserve.

Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003
Tables de variabilité d'échantillonnage approximative - Provinces de l'Atlantique

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ														
	0,1 %	1,0 %	2,0 %	5,0 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %	35,0 %	40,0 %	50,0 %	70,0 %	90,0 %	
1	85,6	85,2	84,7	83,4	81,2	78,9	76,6	74,1	71,6	69,0	66,3	60,5	46,9	27,1	
2	*****	60,2	59,9	59,0	57,4	55,8	54,1	52,4	50,6	48,8	46,9	42,8	33,2	19,1	
3	*****	49,2	48,9	48,2	46,9	45,6	44,2	42,8	41,4	39,8	38,3	34,9	27,1	15,6	
4	*****	42,6	42,4	41,7	40,6	39,5	38,3	37,1	35,8	34,5	33,2	30,3	23,4	13,5	
5	*****	38,1	37,9	37,3	36,3	35,3	34,2	33,2	32,0	30,9	29,7	27,1	21,0	12,1	
6	*****	34,8	34,6	34,1	33,2	32,2	31,3	30,3	29,2	28,2	27,1	24,7	19,1	11,1	
7	*****	32,2	32,0	31,5	30,7	29,8	28,9	28,0	27,1	26,1	25,1	22,9	17,7	10,2	
8	*****	30,1	30,0	29,5	28,7	27,9	27,1	26,2	25,3	24,4	23,4	21,4	16,6	9,6	
9	*****	28,4	28,2	27,8	27,1	26,3	25,5	24,7	23,9	23,0	22,1	20,2	15,6	9,0	
10	*****	26,9	26,8	26,4	25,7	25,0	24,2	23,4	22,6	21,8	21,0	19,1	14,8	8,6	
11	*****	25,7	25,6	25,2	24,5	23,8	23,1	22,4	21,6	20,8	20,0	18,3	14,1	8,2	
12	*****	24,6	24,5	24,1	23,4	22,8	22,1	21,4	20,7	19,9	19,1	17,5	13,5	7,8	
13	*****	23,6	23,5	23,1	22,5	21,9	21,2	20,6	19,9	19,1	18,4	16,8	13,0	7,5	
14	*****	22,8	22,6	22,3	21,7	21,1	20,5	19,8	19,1	18,4	17,7	16,2	12,5	7,2	
15	*****	22,0	21,9	21,5	21,0	20,4	19,8	19,1	18,5	17,8	17,1	15,6	12,1	7,0	
16	*****	21,3	21,2	20,9	20,3	19,7	19,1	18,5	17,9	17,3	16,6	15,1	11,7	6,8	
17	*****	20,7	20,6	20,2	19,7	19,1	18,6	18,0	17,4	16,7	16,1	14,7	11,4	6,6	
18	*****	20,1	20,0	19,7	19,1	18,6	18,0	17,5	16,9	16,3	15,6	14,3	11,1	6,4	
19	*****	19,4	19,1	18,6	18,1	17,6	17,0	16,4	15,8	15,2	13,9	10,8	6,2	6,2	
20	*****	19,0	18,7	18,2	17,6	17,1	16,6	16,0	15,4	14,8	13,5	10,5	6,1	6,1	
21	*****	18,5	18,2	17,7	17,2	16,7	16,2	15,6	15,1	14,5	13,2	10,2	5,9	5,9	
22	*****	18,1	17,8	17,3	16,8	16,3	15,8	15,3	14,7	14,1	12,9	10,0	5,8	5,8	
23	*****	17,7	17,4	16,9	16,5	16,0	15,5	14,9	14,4	13,8	12,6	9,8	5,6	5,6	
24	*****	17,3	17,0	16,6	16,1	15,6	15,1	14,6	14,1	13,5	12,4	9,6	5,5	5,5	
25	*****	16,9	16,7	16,2	15,8	15,3	14,8	14,3	13,8	13,3	12,1	9,4	5,4	5,4	
30	*****	15,5	15,2	14,8	14,4	14,0	13,5	13,1	12,6	12,1	11,1	8,6	4,9	4,9	
35	*****	14,3	14,1	13,7	13,3	12,9	12,5	12,1	11,7	11,2	10,2	7,9	4,6	4,6	
40	*****	13,2	12,8	12,8	12,5	12,1	11,7	11,3	10,9	10,5	9,6	7,4	4,3	4,3	
45	*****	12,4	12,1	11,8	11,4	11,1	10,7	10,3	9,9	9,9	9,0	7,0	4,0	4,0	
50	*****	11,8	11,5	11,2	10,8	10,5	10,1	9,8	9,4	9,4	8,6	6,6	3,8	3,8	
55	*****	11,3	11,0	10,6	10,3	10,0	9,7	9,3	8,9	8,9	8,2	6,3	3,7	3,7	
60	*****	10,8	10,5	10,2	9,9	9,6	9,2	8,9	8,6	8,6	7,8	6,1	3,5	3,5	
65	*****	10,3	10,1	9,8	9,5	9,2	8,9	8,6	8,2	8,2	7,5	5,8	3,4	3,4	
70	*****	10,0	9,7	9,4	9,2	8,9	8,6	8,2	7,9	7,9	7,2	5,6	3,2	3,2	
75	*****	9,6	9,4	9,1	8,8	8,6	8,3	8,0	7,7	7,7	7,0	5,4	3,1	3,1	
80	*****	9,3	9,1	8,8	8,6	8,3	8,0	7,7	7,4	7,4	6,8	5,2	3,0	3,0	
85	*****	9,1	8,8	8,6	8,3	8,0	7,8	7,5	7,2	7,2	6,6	5,1	2,9	2,9	
90	*****	8,8	8,6	8,3	8,1	7,8	7,5	7,3	7,0	7,0	6,4	4,9	2,9	2,9	
95	*****	8,3	8,1	7,9	7,7	7,4	7,1	6,8	6,5	6,5	6,2	4,8	2,8	2,8	
100	*****	8,1	7,9	7,7	7,4	7,2	6,9	6,6	6,3	6,3	6,1	4,7	2,7	2,7	
125	*****	7,3	7,1	6,8	6,6	6,4	6,2	5,9	5,7	5,7	5,4	4,2	2,4	2,4	
150	*****	6,6	6,4	6,3	6,1	5,8	5,6	5,4	5,2	5,2	4,9	3,8	2,2	2,2	
200	*****	5,6	5,4	5,2	5,1	4,9	4,7	4,5	4,4	4,4	4,3	3,3	1,9	1,9	
250	*****	5,0	4,8	4,7	4,5	4,4	4,2	4,1	4,0	4,0	3,8	3,0	1,7	1,7	
300	*****	4,4	4,3	4,1	4,0	3,8	3,7	3,5	3,4	3,4	3,2	2,5	1,4	1,4	
350	*****	4,1	4,0	3,8	3,7	3,5	3,4	3,3	3,2	3,2	3,0	2,3	1,4	1,4	
400	*****	3,7	3,6	3,5	3,3	3,2	3,1	3,0	2,9	2,9	2,7	2,1	1,3	1,3	
450	*****	3,5	3,4	3,3	3,1	3,0	2,9	2,7	2,6	2,6	2,4	1,7	1,0	1,0	
500	*****	3,2	3,1	3,0	2,9	2,7	2,6	2,4	2,3	2,3	2,1	1,5	0,9	0,9	
750	*****	2,4	2,2	2,1	2,0	1,9	1,7	1,6	1,5	1,5	1,3	1,0	0,7	0,7	
1 000	*****	2,2	2,1	2,0	1,9	1,7	1,6	1,5	1,4	1,4	1,2	0,9	0,6	0,6	
1 500	*****	2,0	1,9	1,8	1,7	1,5	1,4	1,3	1,2	1,2	1,0	0,7	0,5	0,5	

Nota : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Tables de variabilité d'échantillonnage approximative - Québec

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1 %	1,0 %	2,0 %	5,0 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %	35,0 %	40,0 %	50,0 %	70,0 %	90,0 %
1	160,3	159,6	158,8	156,4	152,2	147,9	143,5	138,9	134,2	129,3	124,3	113,4	87,9	50,7
2	113,4	112,9	112,3	110,6	107,6	104,6	101,5	98,2	94,9	91,5	87,9	80,2	62,1	35,9
3	92,6	92,2	91,7	90,3	87,9	85,4	82,8	80,2	77,5	74,7	71,7	65,5	50,7	29,3
4	80,2	79,8	79,4	78,2	76,1	74,0	71,7	69,5	67,1	64,7	62,1	56,7	43,9	25,4
5	71,7	71,4	71,0	69,9	68,1	66,1	64,2	62,1	60,0	57,8	55,6	50,7	39,3	22,7
6	*****	65,2	64,8	63,8	62,1	60,4	58,6	56,7	54,8	52,8	50,7	46,3	35,9	20,7
7	*****	60,3	60,0	59,1	57,5	55,9	54,2	52,5	50,7	48,9	47,0	42,9	33,2	19,2
8	*****	56,4	56,1	55,3	53,8	52,3	50,7	49,1	47,5	45,7	43,9	40,1	31,1	17,9
9	*****	53,2	52,9	52,1	50,7	49,3	47,8	46,3	44,7	43,1	41,4	37,8	29,3	16,9
10	*****	50,5	50,2	49,4	48,1	46,8	45,4	43,9	42,4	40,9	39,3	35,9	27,8	16,0
11	*****	48,1	47,9	47,1	45,9	44,6	43,3	41,9	40,5	39,0	37,5	34,2	26,5	15,3
12	*****	46,1	45,8	45,1	43,9	42,7	41,4	40,1	38,7	37,3	35,9	32,7	25,4	14,6
13	*****	44,3	44,0	43,4	42,2	41,0	39,8	38,5	37,2	35,9	34,5	31,5	24,4	14,1
14	*****	42,7	42,4	41,8	40,7	39,5	38,3	37,1	35,9	34,6	33,2	30,3	23,5	13,6
15	*****	41,2	41,0	40,4	39,3	38,2	37,0	35,9	34,7	33,4	32,1	29,3	22,7	13,1
16	*****	39,9	39,7	39,1	38,0	37,0	35,9	34,7	33,6	32,3	31,1	28,4	22,0	12,7
17	*****	38,7	38,5	37,9	36,9	35,9	34,8	33,7	32,6	31,4	30,1	27,5	21,3	12,3
18	*****	37,6	37,4	36,9	35,9	34,9	33,8	32,7	31,6	30,5	29,3	26,7	20,7	12,0
19	*****	36,6	36,4	35,9	34,9	33,9	32,9	31,9	30,8	29,7	28,5	26,0	20,2	11,6
20	*****	35,7	35,5	35,0	34,0	33,1	32,1	31,1	30,0	28,9	27,8	25,4	19,6	11,3
21	*****	34,8	34,7	34,1	33,2	32,3	31,3	30,3	29,3	28,2	27,1	24,8	19,2	11,1
22	*****	34,0	33,9	33,3	32,4	31,5	30,6	29,6	28,6	27,6	26,5	24,2	18,7	10,8
23	*****	33,3	33,1	32,6	31,7	30,8	29,9	29,0	28,0	27,0	25,9	23,7	18,3	10,6
24	*****	32,6	32,4	31,9	31,1	30,2	29,3	28,4	27,4	26,4	25,4	23,2	17,9	10,4
25	*****	31,9	31,8	31,3	30,4	29,6	28,7	27,8	26,8	25,9	24,9	22,7	17,6	10,1
30	*****	29,1	29,0	28,5	27,8	27,0	26,2	25,4	24,5	23,6	22,7	20,7	16,0	9,3
35	*****	27,0	26,8	26,4	25,7	25,0	24,3	23,5	22,7	21,9	21,0	19,2	14,9	8,6
40	*****	25,2	25,1	24,7	24,1	23,4	22,7	22,0	21,2	20,5	19,6	17,9	13,9	8,0
45	*****	23,8	23,7	23,3	22,7	22,0	21,4	20,7	20,0	19,3	18,5	16,9	13,1	7,6
50	*****	22,6	22,5	22,1	21,5	20,9	20,3	19,6	19,0	18,3	17,6	16,0	12,4	7,2
55	*****	21,5	21,4	21,1	20,5	19,9	19,3	18,7	18,1	17,4	16,8	15,3	11,8	6,8
60	*****	20,5	20,2	19,6	19,1	18,5	17,9	17,3	16,7	16,0	15,3	14,6	11,3	6,5
65	*****	19,7	19,4	18,9	18,3	17,8	17,2	16,6	16,0	15,4	14,8	14,1	10,9	6,3
70	*****	19,0	18,7	18,2	17,7	17,2	16,6	16,0	15,5	14,9	14,3	13,6	10,5	6,1
75	*****	18,3	18,1	17,6	17,1	16,6	16,0	15,5	14,9	14,3	13,7	13,1	10,1	5,9
80	*****	17,8	17,5	17,0	16,5	16,0	15,5	15,0	14,5	13,9	13,3	12,7	9,8	5,7
85	*****	17,2	17,0	16,5	16,0	15,6	15,1	14,6	14,0	13,5	12,9	12,3	9,5	5,5
90	*****	16,7	16,5	16,0	15,6	15,1	14,6	14,1	13,6	13,1	12,5	12,0	9,3	5,3
95	*****	16,3	16,0	15,6	15,2	14,7	14,3	13,8	13,3	12,7	12,2	11,6	9,0	5,2
100	*****	15,9	15,6	15,2	14,8	14,3	13,9	13,4	12,9	12,4	11,9	11,3	8,8	5,1
125	*****	14,0	13,6	13,2	12,8	12,4	12,0	11,6	11,1	10,6	10,1	9,6	7,9	4,5
150	*****	12,8	12,4	12,1	11,7	11,3	10,9	10,5	10,1	9,7	9,3	8,9	7,2	4,1
200	*****	11,1	10,8	10,5	10,1	9,8	9,5	9,1	8,8	8,4	8,0	7,6	6,2	3,6
250	*****	9,9	9,6	9,4	9,1	8,8	8,5	8,2	7,9	7,6	7,2	6,9	5,6	3,2
300	*****	8,8	8,5	8,3	8,0	7,7	7,5	7,2	6,9	6,6	6,3	6,0	5,1	2,9
350	*****	8,1	7,9	7,7	7,4	7,2	6,9	6,6	6,3	6,0	5,7	5,4	4,7	2,7
400	*****	7,6	7,4	7,2	6,9	6,7	6,5	6,2	5,9	5,7	5,4	5,1	4,4	2,5
450	*****	7,2	7,0	6,8	6,5	6,3	6,1	5,9	5,6	5,4	5,1	4,8	4,1	2,4
500	*****	6,8	6,6	6,4	6,2	6,0	5,8	5,6	5,4	5,2	5,0	4,8	4,1	2,3
750	*****	5,4	5,2	5,1	4,9	4,7	4,5	4,3	4,1	3,9	3,7	3,5	3,2	1,9
1 000	*****	4,5	4,4	4,2	4,1	3,9	3,7	3,5	3,3	3,2	3,0	2,8	2,6	1,6
1 500	*****	3,5	3,3	3,2	3,0	2,9	2,7	2,6	2,4	2,3	2,1	2,0	1,8	1,3
2 000	*****	2,9	2,8	2,7	2,5	2,4	2,3	2,1	2,0	1,9	1,7	1,6	1,4	1,1
3 000	*****	2,5	2,4	2,3	2,1	2,0	1,9	1,7	1,6	1,5	1,4	1,2	1,1	0,9
4 000	*****	2,1	2,0	1,9	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1,0	0,8	0,8
5 000	*****	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1,0	0,9	0,8	0,7	0,6	0,7

Nota : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Tables de variabilité d'échantillonnage approximative - Ontario

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1 %	1,0 %	2,0 %	5,0 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %	35,0 %	40,0 %	50,0 %	70,0 %	90,0 %
1	246,6	245,5	244,2	240,4	234,0	227,4	220,6	213,6	206,4	198,9	191,1	174,4	135,1	78,0
2	174,3	173,6	172,7	170,0	165,5	160,8	156,0	151,1	145,9	140,6	135,1	123,3	95,5	55,2
3	142,4	141,7	141,0	138,8	135,1	131,3	127,4	123,3	119,2	114,8	110,3	100,7	78,0	45,0
4	123,3	122,7	122,1	120,2	117,0	113,7	110,3	106,8	103,2	99,4	95,5	87,2	67,6	39,0
5	110,3	109,8	109,2	107,5	104,7	101,7	98,7	95,5	92,3	88,9	85,5	78,0	60,4	34,9
6	100,7	100,2	99,7	98,2	95,5	92,9	90,1	87,2	84,3	81,2	78,0	71,2	55,2	31,8
7	93,2	92,8	92,3	90,9	88,5	86,0	83,4	80,7	78,0	75,2	72,2	65,9	51,1	29,5
8	87,2	86,8	86,3	85,0	82,7	80,4	78,0	75,5	73,0	70,3	67,6	61,7	47,8	27,6
9	82,2	81,8	81,4	80,1	78,0	75,8	73,5	71,2	68,8	66,3	63,7	58,1	45,0	26,0
10	*****	77,6	77,2	76,0	74,0	71,9	69,8	67,6	65,3	62,9	60,4	55,2	42,7	24,7
11	*****	74,0	73,6	72,5	70,6	68,6	66,5	64,4	62,2	60,0	57,6	52,6	40,7	23,5
12	*****	70,9	70,5	69,4	67,6	65,7	63,7	61,7	59,6	57,4	55,2	50,4	39,0	22,5
13	*****	68,1	67,7	66,7	64,9	63,1	61,2	59,3	57,2	55,2	53,0	48,4	37,5	21,6
14	*****	65,6	65,3	64,3	62,5	60,8	59,0	57,1	55,2	53,2	51,1	46,6	36,1	20,8
15	*****	63,4	63,1	62,1	60,4	58,7	57,0	55,2	53,3	51,4	49,3	45,0	34,9	20,1
16	*****	61,4	61,1	60,1	58,5	56,9	55,2	53,4	51,6	49,7	47,8	43,6	33,8	19,5
17	*****	59,5	59,2	58,3	56,8	55,2	53,5	51,8	50,1	48,2	46,3	42,3	32,8	18,9
18	*****	57,9	57,6	56,7	55,2	53,6	52,0	50,4	48,6	46,9	45,0	41,1	31,8	18,4
19	*****	56,3	56,0	55,2	53,7	52,2	50,6	49,0	47,4	45,6	43,8	40,0	31,0	17,9
20	*****	54,9	54,6	53,8	52,3	50,9	49,3	47,8	46,2	44,5	42,7	39,0	30,2	17,4
21	*****	53,6	53,3	52,5	51,1	49,6	48,1	46,6	45,0	43,4	41,7	38,1	29,5	17,0
22	*****	52,3	52,1	51,3	49,9	48,5	47,0	45,5	44,0	42,4	40,7	37,2	28,8	16,6
23	*****	51,2	50,9	50,1	48,8	47,4	46,0	44,5	43,0	41,5	39,8	36,4	28,2	16,3
24	*****	50,1	49,8	49,1	47,8	46,4	45,0	43,6	42,1	40,6	39,0	35,6	27,6	15,9
25	*****	49,1	48,8	48,1	46,8	45,5	44,1	42,7	41,3	39,8	38,2	34,9	27,0	15,6
30	*****	44,8	44,6	43,9	42,7	41,5	40,3	39,0	37,7	36,3	34,9	31,8	24,7	14,2
35	*****	41,5	41,3	40,6	39,6	38,4	37,3	36,1	34,9	33,6	32,3	29,5	22,8	13,2
40	*****	38,8	38,6	38,0	37,0	36,0	34,9	33,8	32,6	31,4	30,2	27,6	21,4	12,3
45	*****	36,6	36,4	35,8	34,9	33,9	32,9	31,8	30,8	29,6	28,5	26,0	20,1	11,6
50	*****	34,7	34,5	34,0	33,1	32,2	31,2	30,2	29,2	28,1	27,0	24,7	19,1	11,0
55	*****	33,1	32,9	32,4	31,6	30,7	29,8	28,8	27,8	26,8	25,8	23,5	18,2	10,5
60	*****	31,7	31,5	31,0	30,2	29,4	28,5	27,6	26,6	25,7	24,7	22,5	17,4	10,1
65	*****	30,4	30,3	29,8	29,0	28,2	27,4	26,5	25,6	24,7	23,7	21,6	16,8	9,7
70	*****	29,3	29,2	28,7	28,0	27,2	26,4	25,5	24,7	23,8	22,8	20,8	16,1	9,3
75	*****	28,3	28,2	27,8	27,0	26,3	25,5	24,7	23,8	23,0	22,1	20,1	15,6	9,0
80	*****	27,4	27,3	26,9	26,2	25,4	24,7	23,9	23,1	22,2	21,4	19,5	15,1	8,7
85	*****	26,6	26,5	26,1	25,4	24,7	23,9	23,2	22,4	21,6	20,7	18,9	14,7	8,5
90	*****	25,9	25,7	25,3	24,7	24,0	23,3	22,5	21,8	21,0	20,1	18,4	14,2	8,2
95	*****	25,2	25,1	24,7	24,0	23,3	22,6	21,9	21,2	20,4	19,6	17,9	13,9	8,0
100	*****	24,4	24,0	23,4	22,7	22,1	21,4	20,6	19,9	19,1	17,4	13,5	7,8	
125	*****	21,8	21,5	20,9	20,3	19,7	19,1	18,5	17,8	17,1	15,6	12,1	7,0	
150	*****	19,9	19,6	19,1	18,6	18,0	17,4	16,9	16,2	15,6	14,2	11,0	6,4	
200	*****	17,0	16,5	16,1	15,6	15,1	14,6	14,1	13,5	12,9	12,3	9,6	5,5	
250	*****	15,2	14,8	14,4	14,0	13,5	13,1	12,6	12,1	11,6	11,0	8,5	4,9	
300	*****	13,9	13,5	13,1	12,7	12,3	11,9	11,5	11,0	10,6	10,1	7,8	4,5	
350	*****	12,9	12,5	12,2	11,8	11,4	11,0	10,6	10,2	9,7	9,3	7,2	4,2	
400	*****	12,0	11,7	11,4	11,0	10,7	10,3	9,9	9,6	9,2	8,7	6,8	3,9	
450	*****	11,3	11,0	10,7	10,4	10,1	9,7	9,4	9,0	8,6	8,2	6,4	3,7	
500	*****	10,5	10,2	9,9	9,6	9,2	8,9	8,5	8,1	7,7	7,3	5,7	3,5	
750	*****	8,5	8,3	8,1	7,8	7,5	7,3	7,0	6,7	6,4	6,1	4,7	3,0	
1 000	*****	7,2	7,0	6,8	6,5	6,3	6,0	5,7	5,4	5,1	4,8	3,5	2,5	
1 500	*****	5,7	5,5	5,3	5,1	4,9	4,7	4,5	4,3	4,1	3,9	3,0	2,0	
2 000	*****	4,8	4,6	4,4	4,3	4,1	3,9	3,7	3,5	3,3	3,1	2,5	1,7	
3 000	*****	3,6	3,5	3,4	3,2	3,1	2,9	2,8	2,6	2,5	2,3	1,8	1,4	
4 000	*****	2,8	2,7	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,5	1,2	
5 000	*****	1,9	1,8	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1,0	0,8	0,6	
6 000	*****	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1,0	0,9	0,8	0,6	0,5	
7 000	*****	1,5	1,4	1,3	1,2	1,1	1,0	0,9	0,8	0,7	0,6	0,5	0,4	
8 000	*****	1,3	1,2	1,1	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	

Nota : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Tables de variabilité d'échantillonnage approximative - Provinces de l'Ouest

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1 %	1,0 %	2,0 %	5,0 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %	35,0 %	40,0 %	50,0 %	70,0 %	90,0 %
1	150,9	150,2	149,5	147,1	143,2	139,2	135,0	130,7	126,3	121,7	116,9	106,8	82,7	47,7
2	106,7	106,2	105,7	104,1	101,3	98,4	95,5	92,5	89,3	86,1	82,7	75,5	58,5	33,8
3	87,1	86,7	86,3	85,0	82,7	80,4	78,0	75,5	72,9	70,3	67,5	61,6	47,7	27,6
4	75,4	75,1	74,7	73,6	71,6	69,6	67,5	65,4	63,2	60,9	58,5	53,4	41,3	23,9
5	67,5	67,2	66,8	65,8	64,1	62,2	60,4	58,5	56,5	54,4	52,3	47,7	37,0	21,4
6	61,6	61,3	61,0	60,1	58,5	56,8	55,1	53,4	51,6	49,7	47,7	43,6	33,8	19,5
7	57,0	56,8	56,5	55,6	54,1	52,6	51,0	49,4	47,7	46,0	44,2	40,3	31,3	18,0
8	*****	53,1	52,8	52,0	50,6	49,2	47,7	46,2	44,7	43,0	41,3	37,7	29,2	16,9
9	*****	50,1	49,8	49,0	47,7	46,4	45,0	43,6	42,1	40,6	39,0	35,6	27,6	15,9
10	*****	47,5	47,3	46,5	45,3	44,0	42,7	41,3	39,9	38,5	37,0	33,8	26,1	15,1
11	*****	45,3	45,1	44,4	43,2	42,0	40,7	39,4	38,1	36,7	35,3	32,2	24,9	14,4
12	*****	43,4	43,1	42,5	41,3	40,2	39,0	37,7	36,5	35,1	33,8	30,8	23,9	13,8
13	*****	41,7	41,5	40,8	39,7	38,6	37,5	36,3	35,0	33,8	32,4	29,6	22,9	13,2
14	*****	40,1	39,9	39,3	38,3	37,2	36,1	34,9	33,8	32,5	31,3	28,5	22,1	12,8
15	*****	38,8	38,6	38,0	37,0	35,9	34,9	33,8	32,6	31,4	30,2	27,6	21,4	12,3
16	*****	37,6	37,4	36,8	35,8	34,8	33,8	32,7	31,6	30,4	29,2	26,7	20,7	11,9
17	*****	36,4	36,2	35,7	34,7	33,8	32,8	31,7	30,6	29,5	28,4	25,9	20,1	11,6
18	*****	35,4	35,2	34,7	33,8	32,8	31,8	30,8	29,8	28,7	27,6	25,2	19,5	11,3
19	*****	34,5	34,3	33,8	32,9	31,9	31,0	30,0	29,0	27,9	26,8	24,5	19,0	11,0
20	*****	33,6	33,4	32,9	32,0	31,1	30,2	29,2	28,2	27,2	26,1	23,9	18,5	10,7
21	*****	32,8	32,6	32,1	31,3	30,4	29,5	28,5	27,6	26,6	25,5	23,3	18,0	10,4
22	*****	32,0	31,9	31,4	30,5	29,7	28,8	27,9	26,9	26,0	24,9	22,8	17,6	10,2
23	*****	31,3	31,2	30,7	29,9	29,0	28,2	27,3	26,3	25,4	24,4	22,3	17,2	10,0
24	*****	30,7	30,5	30,0	29,2	28,4	27,6	26,7	25,8	24,8	23,9	21,8	16,9	9,7
25	*****	30,0	29,9	29,4	28,6	27,8	27,0	26,1	25,3	24,3	23,4	21,4	16,5	9,5
30	*****	27,4	27,3	26,9	26,1	25,4	24,7	23,9	23,1	22,2	21,4	19,5	15,1	8,7
35	*****	25,4	25,3	24,9	24,2	23,5	22,8	22,1	21,4	20,6	19,8	18,0	14,0	8,1
40	*****	23,8	23,6	23,3	22,6	22,0	21,4	20,7	20,0	19,2	18,5	16,9	13,1	7,5
45	*****	22,4	22,3	21,9	21,4	20,7	20,1	19,5	18,8	18,1	17,4	15,9	12,3	7,1
50	*****	21,2	21,1	20,8	20,3	19,7	19,1	18,5	17,9	17,2	16,5	15,1	11,7	6,8
55	*****	20,3	20,2	19,8	19,3	18,8	18,2	17,6	17,0	16,4	15,8	14,4	11,2	6,4
60	*****	19,4	19,3	19,0	18,5	18,0	17,4	16,9	16,3	15,7	15,1	13,8	10,7	6,2
65	*****	18,6	18,5	18,3	17,8	17,3	16,7	16,2	15,7	15,1	14,5	13,2	10,3	5,9
70	*****	18,0	17,9	17,6	17,1	16,6	16,1	15,6	15,1	14,5	14,0	12,8	9,9	5,7
75	*****	17,3	17,0	16,5	16,1	15,6	15,1	14,6	14,1	13,5	13,0	11,9	9,5	5,5
80	*****	16,7	16,5	16,0	15,6	15,1	14,6	14,1	13,6	13,1	12,6	11,5	9,2	5,3
85	*****	16,2	16,0	15,5	15,1	14,6	14,2	13,7	13,2	12,7	12,2	11,1	8,9	5,2
90	*****	15,8	15,5	15,1	14,7	14,2	13,8	13,3	12,8	12,3	11,8	10,7	8,7	5,0
95	*****	15,3	15,1	14,7	14,3	13,9	13,4	13,0	12,5	12,0	11,5	10,4	8,5	4,9
100	*****	14,9	14,7	14,3	13,9	13,5	13,1	12,6	12,2	11,7	11,2	10,1	8,3	4,8
125	*****	13,4	13,2	12,8	12,4	12,1	11,7	11,3	10,9	10,5	10,1	9,0	7,4	4,3
150	*****	12,0	11,7	11,4	11,0	10,7	10,3	9,9	9,5	9,1	8,7	7,8	6,4	3,9
200	*****	10,4	10,1	9,8	9,5	9,2	8,9	8,6	8,3	8,0	7,7	7,0	5,8	3,4
250	*****	9,3	9,1	8,8	8,5	8,3	8,0	7,7	7,4	7,1	6,8	6,2	5,2	3,0
300	*****	8,5	8,3	8,0	7,8	7,5	7,3	7,0	6,8	6,5	6,2	5,6	4,8	2,8
350	*****	7,9	7,7	7,4	7,2	7,0	6,8	6,5	6,3	6,1	5,8	5,3	4,4	2,6
400	*****	7,2	7,0	6,8	6,5	6,3	6,1	5,8	5,6	5,4	5,2	4,7	4,1	2,4
450	*****	6,8	6,6	6,4	6,2	6,0	5,8	5,6	5,4	5,2	5,0	4,5	3,9	2,3
500	*****	6,4	6,2	6,0	5,8	5,6	5,4	5,2	5,0	4,8	4,6	4,1	3,7	2,1
750	*****	5,1	4,9	4,8	4,6	4,4	4,3	4,1	3,9	3,7	3,5	3,0	2,6	1,7
1 000	*****	4,4	4,3	4,1	4,0	3,8	3,7	3,5	3,4	3,2	3,0	2,5	2,1	1,5
1 500	*****	3,4	3,3	3,1	3,0	2,8	2,7	2,6	2,4	2,3	2,1	1,7	1,4	1,2
2 000	*****	2,8	2,7	2,6	2,4	2,3	2,1	2,0	1,9	1,8	1,6	1,3	1,1	1,1
3 000	*****	1,9	1,8	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1,0	0,8	0,7	0,9
4 000	*****	1,3	1,2	1,1	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,8
5 000	*****	1,2	1,1	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,7
6 000	*****	0,6	0,5	0,4	0,3	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,6

Nota

1. Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.
2. Dans le cas des estimations relatives aux francophones des provinces de l'Ouest, les effets du plan d'échantillonnage sont particulièrement élevés. Les CV approximatifs concernant cette sous-population doivent être multipliés par un facteur de 1,7.

Enquête internationale sur l'alphabétisation et les compétences des adultes, 2003

Tables de variabilité d'échantillonnage approximative - Territoires du Nord

NUMÉRATEUR DU POURCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1 %	1,0 %	2,0 %	5,0 %	10,0 %	15,0 %	20,0 %	25,0 %	30,0 %	35,0 %	40,0 %	50,0 %	70,0 %	90,0 %
1	*****	19,0	18,7	18,2	17,7	17,2	16,7	16,1	15,5	14,9	13,6	10,5	6,1	
2	*****		13,3	12,9	12,5	12,2	11,8	11,4	11,0	10,5	9,6	7,4	4,3	
3	*****			10,5	10,2	9,9	9,6	9,3	9,0	8,6	7,9	6,1	3,5	
4	*****				9,1	8,9	8,6	8,3	8,0	7,8	7,4	6,8	5,3	3,0
5	*****					8,2	7,9	7,7	7,4	7,2	6,9	6,1	4,7	2,7
6	*****						7,2	7,0	6,8	6,6	6,3	6,1	5,6	4,3
7	*****							6,7	6,5	6,3	6,1	5,9	5,6	4,0
8	*****								6,3	6,1	5,9	5,5	4,8	3,7
9	*****									5,7	5,6	5,4	5,2	5,0
10	*****										5,4	5,3	5,1	4,9
11	*****											5,2	5,0	4,9
12	*****												4,8	4,6
13	*****													4,6
14	*****													
15	*****													
16	*****													
17	*****													
18	*****													
19	*****													
20	*****													
21	*****													
22	*****													
23	*****													
24	*****													
25	*****													
30	*****													
35	*****													
40	*****													
45	*****													
50	*****													

Nota : Pour bien utiliser ces



10.0 Dispositions d'enregistrement et variables uniques



Veillez referer au document «EIACA_manueldecode_F.pdf» pour les dispositions d'enregistrement et les fréquences univariés de l'EIACA.

11.0 Personnes-ressources

Services à la clientèle du CTCSE

Services à la clientèle du CTCSE
Statistique Canada
150, promenade du Pré Tunney
Ottawa ON K1A 0T6
Tél: 1-800-307-3382
Télécopieur: 613-951-4441
Courriel: educationstats@statcan.ca

RHDCC

Ressources humaines et Développement sociales Canada :
Service à la clientèle
Direction de la politique sur l'apprentissage
Politique stratégique
140, Promenade du Portage
Phase IV, 3e étage
Gatineau (Québec) K1A 0J9
Tél. : (819) 994-4566
Télécopieur : (819) 997-5433
Courriel : NC-SP-PS-LPDR-DGPAR-GD@hrsdc-rhdcc.gc.ca

NLS

Diana Kaan
Agente principale de politique
Secrétariat national à l'alphabétisation, RHDCC
Gouvernement du Canada
360, rue Albert, Suite 1510, Ottawa (ON) K1A 0J9
Tél: 613-946-5722
Télécopieur: 613-946-5882
Courriel: diana.kaan@hrsdc-rhdcc.gc.ca
http://www.hrsdc.gc.ca/en/gateways/nav/top_nav/program/nls.shtml