



Guide de l'utilisateur des microdonnées

ENQUÊTE LONGITUDINALE NATIONALE SUR LES ENFANTS ET LES JEUNES

Fichier secondaire

Cycle 4

Septembre 2000 à Juin 2001



Statistics
Canada

Statistique
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	6
3.0	Objectifs	8
4.0	Concepts et définitions	9
4.1	Définitions	9
4.2	Variables dérivées relatives à la famille	10
4.3	Répondant	10
5.0	Méthodologie de l'enquête – Échantillon	12
6.0	Collecte des données	13
6.1	Collecte auprès des ménages	13
6.1.1	Composante «Enfant»	13
6.2	Collecte dans les écoles	14
6.2.1	Questionnaire de l'enseignant	14
6.2.2	Questionnaire du directeur	14
7.0	Traitement des données	16
7.1	Vérification	16
7.2	Saisie des données	16
7.3	Nettoyage	16
7.4	Préédition	17
7.5	Vérification de l'enchaînement des questions	17
7.6	Codage	17
7.7	Vérification de la cohérence	17
7.8	Nomenclature et structure de codage des variables de l'ELNEJ	18
7.9	Siglaion des sections du questionnaire	19
7.10	Exemples de noms de variables	21
7.11	Structure de codage des variables de l'ELNEJ	22
7.12	Indicateurs d'imputation	23
7.13	Variables dérivées	23

8.0	Contenu de l'enquête	24
8.1	Système de traitement de l'enquête	24
8.2	Variables démographiques	24
8.3	Questionnaire de l'adulte	25
8.4	Questionnaire de l'enfant	25
8.5	Avertissement	27
9.0	Validation des échelles de l'ELNEJ	29
9.1	Validation des données d'échelle	29
9.1.1	Définition de l'échelle	29
9.1.2	Échelles et calculs	29
9.1.3	Évaluation des données exprimées selon une échelle	29
9.2	Analyse factorielle	30
9.2.1	Analyse factorielle pour les échelles	30
9.2.2	Transformation des données au moyen de l'échelonnement optimal	31
9.2.3	Analyse factorielle axée sur les données pondérées	31
9.3	Calcul des scores et imputation des réponses	32
9.3.1	Calcul des scores pour chaque facteur	32
9.3.2	Exemple de calcul des scores de facteurs	32
9.3.3	Poids factoriel négatif	33
9.3.4	Code de non-réponse	33
9.3.5	Items bruts	33
9.4	Mesures de fiabilité des échelles	33
9.4.1	Coefficient alpha de Cronbach	34
9.4.2	Interprétation du coefficient alpha de Cronbach	34
9.4.3	Qu'est-ce qu'un niveau satisfaisant de fiabilité?	34
9.5	Échelles de l'éducation du cycle 4	35
10.0	Méthodologie d'enquête – Taux de réponse et couverture de la composante Éducation	37
10.1	Sous-couverture de l'échantillon	37
10.2	Non-réponse totale	39
10.2.1	Définitions	39
10.3	Les consentements	40
10.4	La non réponse des enseignants et des directeurs	40

11.0	Imputation	42
11.1	Imputation des réponses des directeurs d'école	42
11.2	Imputation des réponses des enseignants.....	43
12.0	Pondération, traitement de la non-réponse et limites.....	44
12.1	Ajustement de la sous couverture	44
12.2	Ajustement du consentement des parents.....	45
12.3	Ajustement du consentement des commissions scolaires	46
12.4	Ajustement pour non-réponse des enseignants ou directeurs	46
12.5	Résumé des ajustements.....	48
12.6	Limitations	48
13.0	Qualité des données et recouvrement de l'enquête principal.....	50
14.0	Lignes directrices concernant la totalisation, l'analyse et la diffusion de données.....	51
14.1	Lignes directrices pour l'arrondissement	51
14.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	52
14.3	Définitions des estimations de type nominal et des estimations quantitatives.....	53
14.3.1	Estimations de type nominal.....	53
14.3.2	Estimations quantitatives	53
14.3.3	Totalisation d'estimations de type nominal	54
14.3.4	Totalisation d'estimations quantitatives	54
14.4	Lignes directrices pour l'analyse statistique.....	55
14.5	Lignes directrices pour la diffusion en fonction du coefficient de variation.....	56
14.6	Lignes directrices relatives au niveau de qualité	57
15.0	Calcul de la variance	58
15.1	Importance du calcul de la variance.....	58
15.2	Calcul de la variance et du coefficient de variation	58
15.2.1	Tables de variabilité d'échantillonnage approximative	59
15.2.2	Calcul des coefficients de variation	59
15.2.3	Allocation de l'échantillon	59
15.2.4	Effets du plan, des tailles de l'échantillon et des chiffres de population.....	60
15.2.5	Nature approximative des tables de variabilité.....	61
15.3	Comment utiliser les tables c.v. pour des estimations de type nominal	61
15.4	Exemples d'utilisation des tables de c.v. et du chiffrier pour des estimations nominales	63

15.5	Comment utiliser les tables de c.v. pour calculer des intervalles	65
15.6	Comment utiliser les tables de c.v. pour effectuer un test t	66
15.7	Coefficients de variation d'estimations quantitatives	67
15.8	Seuils pour la diffusion des estimations relatives à la composante Éducation de l'ELNEJ67	
15.9	Programmes SAS et SPSS pour calculer la variance, en utilisant les poids bootstrap69	
16.0	Direct Assessment	70
17.0	Enjeux analytiques.....	71
17.1	Impact de l'utilisation d'un plan de sondage complexe sur l'analyse.....	71
17.2	L'unité d'analyse.....	71
17.3	Type d'analyse : transversale, longitudinale ou répétée	71
17.3.1	Analyse longitudinale	71
17.3.2	Analyse transversale	72
17.3.3	Analyse répétée	72
17.4	Non-réponse totale et partielle.....	73
17.4.1	Non-réponse totale	73
17.4.2	Non-réponse partielle	73
17.5	Traitement des données	73
17.6	Calcul de la variance avec certains logiciels	74
18.0	Annexe I - Clichés d'enregistrement.....	75
18.1	Fichier de l'éducation.....	75
18.2	Fichier secondaire.....	81

1.0 Introduction

Les données du cycle 4 de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) – composante de l'éducation – ont été recueillies d'avril à juin 2001 par Statistique Canada, en partenariat avec Développement des ressources humaines Canada. La composante de l'éducation est l'une des parties de la deuxième diffusion des données du cycle 4. L'autre partie de la diffusion comprend chacune des variables utilisées pour établir les scores des questionnaires sur les étapes du développement administrés aux enfants âgés de 3 à 71 mois.

Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées de la composante de l'éducation, pour décrire la qualité des données et pour expliquer d'autres questions d'analyse. La première diffusion des données du cycle 4 a eu lieu en juin 2003, et on y renvoie souvent le lecteur pour des renseignements supplémentaires tout au long de ce manuel.

Ce cycle est le dernier cycle de l'ELNEJ pour lequel de l'information est recueillie des enseignants et des directeurs d'école.

Toute question au sujet de l'ensemble de données ou de son utilisation devrait être adressée à :

Statistique Canada

Services à la clientèle
Division des enquêtes spéciales
Téléphone : (613) 951-3321 ou appelez sans frais : 1 800 461-9050
Télécopieur : (613) 951-4527
Courriel : des@statcan.ca

Division des enquêtes spéciales, Statistique Canada
2500 - R, Édifice Main
Tunney's Pasture
Ottawa, Ontario K1A 0T6

Développement des ressources humaines Canada

Direction générale de la recherche appliquée
Développement des ressources humaines Canada
Place du Portage-Phase II, 7e étage
165, rue Hôtel de Ville
Hull, Québec
K1A 0J2
Téléphone : (819) 953-3465
Télécopieur : (819) 953-8868

2.0 Contexte

L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) est une étude à long terme qui vise principalement à observer le développement et le bien-être des enfants au Canada dans leur cheminement de la petite enfance à l'âge adulte. L'enquête a débuté en 1994 et elle est menée conjointement par Statistique Canada et Développement des ressources humaines Canada.

L'enquête a été conçue pour recueillir des renseignements sur les facteurs qui influent sur le développement social et émotionnel ainsi que sur le comportement des enfants et des jeunes. Elle permet en outre de suivre les conséquences de ces facteurs sur leur développement au fil du temps.

L'enquête porte sur un large éventail de sujets y compris la santé, le développement physique, l'apprentissage et le comportement des enfants ainsi que des données sur leur environnement social (famille, amis, école et communauté). Les données de l'ELNEJ sont utilisées par un grand nombre de personnes à tous les niveaux du gouvernement, dans les universités et les organisations qui élaborent les politiques.

La composante de l'éducation comprend trois questionnaires distincts qui ne font pas partie de la composante principale de l'ELNEJ : le questionnaire du directeur, le questionnaire de l'enseignant unique et le questionnaire des différents enseignants.

Au cours de la collecte du cycle 4 de l'ELNEJ, Statistique Canada a eu une charge de travail plus lourde que prévu en matière d'interviews en raison du nombre d'enquêtes simultanées sur le terrain. Au bout de deux mois, il est devenu évident que la collecte de la composante des écoles était compromise. On a examiné un certain nombre de possibilités et on a pris les décisions suivantes : on ferait passer le test de mathématiques dans le ménage, on abandonnerait le test de lecture afin de diminuer le fardeau du répondant et de réduire les coûts au minimum et enfin, la composante des écoles se limiterait au sous-échantillon de répondants pour qui les données scolaires pourraient être recueillies et traitées à temps pour l'envoi par la poste aux écoles avant la fin de l'année scolaire. Cette dernière décision a eu comme résultat un taux plus faible de retour des questionnaires des enseignants et des directeurs d'écoles. Pour compenser la non-réponse, on a créé un poids spécialement pour le fichier de l'éducation. Ce poids sert à l'estimation transversale et il est le seul poids qui a été produit spécifiquement pour le fichier d'éducation. Pour de plus amples renseignements au sujet de la pondération, on peut se reporter au chapitre 12.

Population de l'enquête

Au cycle 4, un échantillon représentatif d'enfants canadiens âgés de 0 à 17 ans a été suivi afin de produire des estimations longitudinales et transversales. La composante de l'éducation comprend seulement les enfants de 6 à 15 ans qui fréquentent l'école publique. Note: Il y a 5 cas pour lesquels le parent a dit que l'enfant fréquentait une école publique alors que le directeur d'école a dit que l'établissement était une école privée.

Population cible

Les objectifs de l'ELNEJ sont de produire à la fois des estimations longitudinales et transversales. Par conséquent, plusieurs populations cibles sont visées par l'échantillon du cycle 4 :

- Transversalement, l'échantillon du cycle 4 représente les enfants de 0 à 17 ans au 1^{er} janvier 2001.
- Longitudinalement, nous avons trois cohortes :
 - Une première cohorte représente les enfants de 0 à 11 ans en 1994-1995; ces enfants sont maintenant âgés de 6 à 17 ans au cycle 4.

- Une deuxième cohorte représente les enfants âgés de 0 à 1 an en 1996-1997; ces enfants sont âgés de 4 à 5 ans au cycle 4.
- Une troisième cohorte représente les enfants âgés de 0 à 1 an en 1998-1999; ces enfants sont âgés de 2 à 3 ans au cycle 4.
- Une quatrième cohorte représente les enfants âgés de 0 à 1 an en 2000-2001. Ces enfants seront suivis jusqu'à l'âge de 5 ans.

La population cible pour le fichier de l'éducation diffère légèrement de celle de la composante principale. Tandis que la composante principale de l'enquête a pour but de représenter tous les enfants dans leurs cohortes d'âge respectives, celle de l'éducation vise seulement les enfants qui fréquentent l'école publique.

Cycles de collecte

La collecte de données se fait à intervalles de deux ans.

Cycle	Début de la collecte	Fin de la collecte
1	février 1995	avril 1995
2	avril 1997	juin 1997
3	avril 1999	juin 1999
4	avril 2001	juin 2001

3.0 Objectifs

Les objectifs de l'ELNEJ sont :

- déterminer la fréquence de divers facteurs de risque et de protection chez les enfants et les jeunes;
- comprendre comment ces facteurs influent, tout comme les événements de la vie, sur le développement de l'enfance;
- mettre cette information au service de l'élaboration de politiques et de programmes destinés à aider les enfants et les jeunes;
- recueillir des renseignements sur une grande diversité de sujets d'ordre biologique, social et économique;
- recueillir des renseignements sur le milieu où grandit l'enfant, qu'il s'agisse de la famille, des pairs, de l'école ou de la collectivité;

Les renseignements proviennent de diverses sources (parents, enfants et enseignants), ainsi que de mesures directes (EPIV, tests de calcul et de lecture, etc.).

Stratégie de communication des données

Les fichiers de microdonnées de la composante principale du cycle 4 ont été diffusés en juin 2003, en même temps que la rediffusion des fichiers mis à jour des cycles 1, 2 et 3. Le fichier de microdonnées de la composante de l'éducation fait partie d'une diffusion secondaire du cycle 4, qui comprendra également un fichier de microdonnées distinct contenant les variables utilisées pour calculer les scores des étapes du développement (pour des renseignements sur cette mesure, voir le guide de l'utilisateur du fichier principal).

Qualité des données pour les questionnaires sur les étapes du développement

Il est important que les utilisateurs soient informés des taux de non-réponse particulièrement élevés pour la composante sur les étapes du développement. En présence de non-réponse, les utilisateurs devraient toujours faire une analyse de la non-réponse, afin de déterminer l'impact de celle-ci. Ceci est particulièrement vrai lorsque la non-réponse est importante. Cependant, si la non-réponse est trop importante, il peut devenir très difficile, voire impossible de produire des estimations fiables. Comme règle de base, les taux de réponse devraient être d'au moins 67 %. Lorsque les taux de réponse sont inférieurs à 67 %, ajuster pour la non-réponse peut devenir problématique.

L'Initiative de démocratisation des données (IDD) fournit aux établissements d'enseignement post-secondaire canadien un accès illimité à des fins exclusives de recherche et d'enseignement, aux fichiers de microdonnées à grande diffusion et aux bases de données. Les fichiers pour le quatrième cycle de l'ELNEJ seront disponibles sur la page Web IDD au lien hypertexte suivant: http://www.statcan.ca/francais/Dli/Data/Ftp/nlscy_f.htm

4.0 Concepts et définitions

Nombre de variables et de concepts sont essentiels aux analyses des données de l'ELNEJ. Une explication des concepts clés de l'ELNEJ est présentée ci-dessous. Les variables dérivées sont celles qui ne proviennent pas directement des renseignements fournis par les répondants, mais qui sont plutôt calculées à partir de ces renseignements.

Le contenu de chaque section des divers questionnaires de l'ELNEJ est présenté dans le guide de l'utilisateur du fichier principal.

Dans le cadre de l'ELNEJ, l'unité d'analyse est l'enfant ou le jeune. Le chapitre 5 décrit la conception de l'enquête.

Seuls les concepts pertinents au fichier de l'éducation seront inclus ici. On peut trouver des renseignements relatifs aux concepts du fichier principal dans le guide de l'utilisateur du fichier principal.

4.1 Définitions

Composantes

L'ELNEJ est constituée de diverses composantes, celles-ci sont des ensembles de questions générés par l'application selon l'âge de l'enfant. Les principales composantes sont : Adulte, Enfant, Adolescent, Questionnaire auto-administré, Tests de mesure directe, Exercices de calcul et Mesures cognitives. La composante de l'éducation constitue une composante supplémentaire; les questionnaires de cette composante sont remplis par le directeur et les enseignants des enfants dont les parents et le conseil scolaire ont donné leur consentement et qui fréquentent une école publique ou catholique.

Âge effectif (DMMCQ01)

L'âge de l'enfant détermine, dans la plupart des cas, les questions qui seront posées à son sujet. Plutôt que d'utiliser l'âge réel de l'enfant, on fait appel dans l'ELNEJ à un âge calculé appelé « âge effectif », pour que l'enfant reste dans le groupe d'âge auquel il a été associé, peu importe que la collecte ait lieu avant ou après son anniversaire. Pour le cycle 4, l'âge effectif est calculé en fonction de l'année 2000 : c'est-à-dire 2000 moins année de naissance. Par exemple, un enfant né 1998 aurait un âge effectif de 2 ans (2000-1998). Notez que l'âge actuel de l'enfant au moment de l'interview est parfois différent de l'âge effectif.

Personne la mieux renseignée et conjoint

Dans chaque ménage de l'ELNEJ, pour chaque enfant sélectionné, on a demandé au ménage quelle était la personne qui connaissait le mieux cet enfant. Cette personne a été désignée comme la « personne la mieux renseignée sur cet enfant » (PMR). Celle-ci fournit des renseignements sur tous les enfants sélectionnés au sein du ménage ainsi que des renseignements sur elle-même et sur son conjoint. En de rares occasions, il aurait été préférable de désigner deux PMR différentes dans un même ménage.

Par exemple, dans le cas d'une famille reconstituée, il aurait peut-être été plus indiqué de désigner la mère comme PMR pour un enfant et le père pour un autre enfant.

Cependant, pour simplifier la procédure d'interview, seulement une PMR a été désignée par ménage.

Certains ménages ne comptent pas de PMR. Lorsque l'enfant sélectionné est âgé de 16 ans et plus et qu'il ne vit plus avec un parent ou un tuteur, aucune PMR n'est désignée au sein du ménage.

La ventilation des données sur les liens entre les PMR et les enfants de l'ELNEJ pour le quatrième cycle s'établit comme suit :

Pour 89,8 % des enfants, la PMR était la mère (la mère biologique dans 88,6 % des cas et la mère par alliance, adoptive ou de famille d'accueil dans 1,2 % des cas).

Pour 8,7 % des enfants, la PMR était le père et pour 1,5 % des enfants, la PMR n'était pas l'un des parents.

Dans la majorité des cas où la PMR n'était pas un des parents, l'enfant vivait dans le même ménage qu'un de ses parents, mais ce dernier n'avait pas été retenu comme PMR. Cette situation se présente généralement lorsque la mère était très jeune et vivait elle-même avec ses parents, c'est-à-dire les grands-parents de l'enfant, et que la grand-mère avait été désignée comme la PMR.

Lorsque la PMR avait un partenaire qui résidait sous le même toit au moment de l'interview, cette personne était désignée comme le conjoint. Étaient considérés comme conjoints les partenaires mariés et les partenaires en union libre. On a recueilli des renseignements socioéconomiques détaillés au sujet du conjoint afin de décrire la situation de la famille de l'enfant.

Changement de la « personne la mieux renseignée » d'un cycle à l'autre

Pour diverses raisons, la PMR et son conjoint peuvent être des personnes différentes de celles désignées au cycle précédent. C'est pourquoi les analystes doivent agir avec prudence lorsqu'ils comparent les données sur la PMR d'un cycle à l'autre.

4.2 Variables dérivées relatives à la famille

Pour obtenir des renseignements supplémentaires au sujet de ces variables, on peut se reporter au guide de l'utilisateur du fichier principal.

- Familles monoparentales
- Famille intacte, reconstituée ou recomposée
- Fratrie

4.3 Répondant

Un répondant transversal est un enfant dont la composante adulte ou sa composante enfant ou jeune est complète. Ces enfants représentent la population cible de janvier 2001.

Un ménage répondant est un ménage où une composante adulte ou une composante enfant ou jeune a été complétée.

Un enfant répondant est un enfant pour lequel une composante adulte ou sa composante enfant ou jeune a été complétée. Un ménage répondant sans composante adulte complète pourrait avoir 1 enfant répondant et 1 enfant non-répondant. Dix-neuf enfants se retrouvent dans cette situation.

Une fois que l'état de la réponse était déterminé au niveau des ménages, les répondants étaient sélectionnés parmi tous les enfants sous-échantillonnés et admissibles du fichier de l'éducation. Pour cette sélection, les conditions suivantes devaient être remplies : l'enfant devait fréquenter une école publique confessionnelle ou non par l'intermédiaire de laquelle nous avions reçu le consentement des parents, le consentement du conseil scolaire ainsi qu'un questionnaire de l'enseignant (d'un enseignant unique ou de plus d'un enseignant) ou du directeur de l'école. Ces répondants étaient ensuite évalués quant à leur ressemblance à la population cible.

5.0 Méthodologie de l'enquête – Échantillon

Le plan d'échantillonnage de l'ELNEJ est un plan probabiliste qui est en grande partie déterminé par le plan d'échantillonnage de l'Enquête sur la population active (EPA). Tous les répondants du fichier de l'éducation proviennent du panel longitudinal initial de l'ELNEJ, qui ont été introduits dans l'échantillon en 1994 et y sont demeurés jusqu'au cycle de collecte actuel. On pourra trouver des renseignements plus détaillés sur l'ensemble de la méthode d'échantillonnage dans le guide de l'utilisateur du fichier principal.

Comme il est expliqué au chapitre précédent, les répondants du fichier de l'éducation constituent un sous-échantillon de l'échantillon admissible de l'ELNEJ. L'inclusion du répondant dans le sous-échantillon était déterminée par l'état complet de la composante du ménage au moment de l'échéance fixée pour l'admissibilité et, par conséquent, ne constitue pas un véritable sous-échantillonnage probabiliste. Ce type d'échantillonnage est souvent appelé « échantillonnage défini par un seuil d'inclusion ». Même si pour chaque élément on peut calculer la probabilité de sélection, il y a exclusion délibérée d'une partie de la population cible.

Le sous-échantillon d'enfants admissibles représente environ 70 % de l'échantillon total. Ce plus petit échantillon ainsi que la réduction du fardeau de réponse pour la composante des écoles ont eu un effet positif sur le nombre de questionnaires retournés. Le chapitre 10 présente une analyse complète des problèmes de couverture découlant du sous-échantillonnage et une analyse de la non-réponse afin d'évaluer le biais possible.

Le lecteur doit noter que bien qu'on ait tout mis en œuvre pour qu'il y ait ajustement proportionnel itératif du sous-échantillon aux caractéristiques de l'échantillon complet, il se peut que les nouveaux poids ne rendent pas compte complètement de certains biais. Pour de plus amples renseignements sur la pondération, on peut se reporter au chapitre 12.

6.0 Collecte des données

La collecte des données pour le quatrième cycle de l'ELNEJ s'est déroulée de l'automne 2000 au printemps 2001. Elle a été effectuée dans deux environnements principaux : les ménages et les écoles.

6.1 Collecte auprès des ménages

L'enquête combine les méthodes d'interview assistée par ordinateur et l'usage des questionnaires papier. Selon la composition du ménage et la nature des composantes requises, l'interview s'effectuera en tout ou en partie par téléphone et /ou visite sur place.

Vous trouverez ci-après une brève description des composantes assistées par ordinateur et des questionnaires papier utilisés pour la collecte de l'ELNEJ.

6.1.1 Composante «Enfant»

Une composante «Enfant» est créée pour chaque enfant choisi qui est âgé de 0 à 17 ans.

La personne la mieux renseignée sur les enfants et les jeunes (PMR) répond aux questions de la composante «Enfant». Il s'agit normalement de la mère de l'enfant, mais il peut également s'agir du père, d'un beau parent ou d'un parent adoptif qui habite dans le même logement. Seule la PMR ou son conjoint peut répondre aux questions de cette composante.

À la fin de cette composante, on demandera au répondant de fournir le nom, l'adresse et le numéro de téléphone de deux personnes (amis, parents) qui pourraient nous aider à dépister la famille dans deux ans, lorsque l'enquête sera répétée.

Pour les enfants à l'école primaire (enfants âgés jusqu'à 15 ans, sauf ceux à la maternelle) et en préparation pour la collecte de données dans les écoles, on demandera à la personne la mieux renseignée son consentement à ce que des informations soient recueillies auprès du professeur et du directeur de l'enfant sélectionné.

Liste des sujets traités:

- Scolarité
- Santé
- Renseignements médicaux et biologiques
- Travail de la mère après la naissance
- Développement de l'enfant
- Tempérament
- Alphabétisation
- Communication
- Activités
- Comportement
- Comportement positif
- Habitudes de sommeil
- Développement moteur et social
- Relations
- Rôle parental

- Garde des enfants
- Attentes (Aspirations)
- Caractéristiques sociodémographiques

Les composantes suivantes font également partie de l'ELNEJ et elles sont décrites en détail dans le guide de l'utilisateur du fichier principal :

- Composante Adulte
- Composante Jeune
- Composante Peabody (échelle de vocabulaire en images de Peabody – EVIP)
- Composante Connaissance des nombres
- Composante Écran de contrôle générique
- Questionnaires sur les étapes du développement
- Questionnaire Qui suis-je ? pour les 4 et 5 ans
- Test de mathématiques (exercice de calcul)
- Questionnaires auto-administrés – 10 à 11, 12 à 13, 14 à 15 et 16 à 17 ans
- Test cognitif pour les 16 et 17 ans

6.2 Collecte dans les écoles

Cette étape de la collecte des données a été menée en milieu scolaire d'avril à juin 2001.

Pour les enfants de l'échantillon qui fréquentaient l'école à un niveau supérieur à la maternelle et dont l'âge était inférieur ou égal à 15ans, on a demandé aux Personnes les mieux renseignées sur l'enfant leur consentement afin de recueillir des données auprès de l'enseignant et du directeur de l'école de l'enfant. La collecte à l'école comportait trois questionnaires qui ont été envoyés par la poste aux enseignants et aux directeurs. On a demandé à ces derniers de les remplir et de les retourner par la poste à Statistique Canada dans des enveloppes fournies à cette fin. Si un élève avait plusieurs enseignants, la trousse devait être remise à l'enseignant qui le connaissait le mieux, idéalement un professeur de langue ou de mathématiques.

6.2.1 Questionnaire de l'enseignant

Deux questionnaires ont été développés, un premier pour les cas où toutes les matières de base sont enseignées par le même enseignant et un second pour les cas où l'enfant a des enseignants différents pour les matières de base. Afin de déterminer quel questionnaire devait être envoyé au professeur, nous avons demandé aux parents si leur enfant avait le même professeur pour la plupart des matières de base (i.e. mathématiques, langue, arts et science). Les questionnaires de l'enseignant portaient sur le rendement et le comportement de l'enfant à l'école ainsi que sur les méthodes pédagogiques et le climat en classe.

6.2.2 Questionnaire du directeur

Le questionnaire du directeur permettait de recueillir de l'information sur les méthodes pédagogiques en vigueur à l'école, la disponibilité des ressources éducatives et le climat social de l'école. Ainsi, le questionnaire du directeur

servait à recueillir des données sur les politiques de l'école et le climat d'enseignement, plutôt qu'au sujet d'un enfant en particulier.

7.0 Traitement des données

7.1 Vérification

Le principal produit de l'ELNEJ est un fichier principal de données « nettoyées ». Nous décrivons brièvement ici certaines des étapes de traitement nécessaires à la production de ce fichier.

La vérification des questionnaires papier a été effectuée à l'aide d'un environnement de traitement généralisé. Il s'agit d'un système générique qui suit les étapes suivantes pour « nettoyer » un fichier du début à la fin :

Nettoyage
Préédition
Vérification de l'enchaînement des questions
Codage
Vérification de la cohérence
Variables dérivées
Fichier de traitement final

7.2 Saisie des données

Saisie des données pour les questionnaires d'interview papier et crayon

On a fait la saisie de ces questionnaires au bureau central de Statistique Canada.

On fait la saisie de tout document contenant au moins une question remplie par le répondant et on remet un fichier contenant tous les enregistrements au bureau central pour la suite du traitement. De plus, on a intégré dans ce processus de saisie, certaines vérifications de qualité qui marquent les entrées inusitées afin de mettre les opérateurs en garde contre les risques d'entrées erronées.

Si le répondant avait coché plusieurs réponses, il devait accepter la première. Les erreurs restantes devaient être corrigées à une étape ultérieure.

7.3 Nettoyage

Exigences définitives

L'objet de cette étape est de retirer les enregistrements en double complets, de séparer les enregistrements d'identification en double pour examen et de diviser les données en réponse et non-réponse, selon les critères donnés.

On a examiné les questionnaires des répondants et des non-répondants et on a établi des spécifications fondées sur cette analyse pour déterminer les enregistrements qui seraient éliminés en raison de la non-réponse. Essentiellement, s'il manquait des données essentielles ou s'il n'y avait pas de réponse à plus de la moitié des questions sur un enregistrement, il était éliminé du fichier.

7.4 Prédiction

Pour tous les enregistrements où il manquait des valeurs (blancs), on a inséré la valeur « 9, 99, 999... » pour indiquer qu'aucune information n'avait été recueillie. De même, les réponses suivant l'instruction « cochez toutes les réponses qui s'appliquent » ont été modifiées pour en faire des réponses oui/non. Enfin, toutes les réponses de texte ont été enlevées du fichier de traitement et mises de côté pour être traitées séparément.

7.5 Vérification de l'enchaînement des questions

Cette vérification consiste à reproduire le cheminement des questions du questionnaire. Les variables qui sont sautées en suivant l'enchaînement des questions sont changées de *Non déclaré* à *Enchaînement valide*.

On utilise des tables de décision pour préciser les vérifications. Le logiciel LogiPlus est utilisé pour générer le code SAS en utilisant l'information fournie par les tables de décision. Un rapport comprenant les comptes avant et après des variables est produit, de même qu'un rapport comprenant le nombre de décisions prises.

Lorsque l'instruction « passez à » suit une question à laquelle on a répondu, toutes les questions sautées sont indiquées comme *Enchaînement valide*. Lorsque l'instruction « passez à » suit une réponse *Ne sait pas* ou *Refus*, toutes les questions sautées sont indiquées comme *Non déclaré*.

7.6 Codage

Lors du processus de codage, on attribue des codes standard comme le code SCIAN aux champs des réponses afin qu'on puisse mieux analyser les données. De plus, les champs *Autre, précisez* sont examinés et recodés.

Le codage a été fait manuellement par le personnel de la division préposé au traitement. Les questions pour lesquelles il y a plusieurs catégories de réponses contiennent souvent *Autre, précisez* comme dernière catégorie. La réponse à la question *Autre, précisez* est saisie. Ces réponses sont examinées et peuvent être recodées à l'une des catégories existantes. Si la réponse correspond à l'une des catégories existantes, cette catégorie est indiquée comme *Oui* et *Autre, précisez* est indiquée comme *Non*.

7.7 Vérification de la cohérence

La vérification de la cohérence consiste à examiner la cohérence entre différentes variables. La première étape a été d'apparier les questionnaires du directeur et des enseignants et ceux des répondants de l'enquête principale afin de s'assurer qu'ils faisaient partie de l'échantillon. De plus, tout enregistrement concernant un enfant dont le parent ou le conseil scolaire avait refusé qu'il participe à l'enquête était retiré.

La deuxième étape consistait à comparer le niveau scolaire déclaré de l'enfant. Dans les cas où le niveau scolaire déclaré par le parent ne correspondait pas à celui déclaré par l'enseignant, on acceptait comme exact le niveau déclaré par l'enseignant.

La troisième et dernière étape a été de vérifier la cohérence entre les réponses à l'intérieur du questionnaire. Quand les réponses étaient incompatibles, on acceptait la première réponse comme étant la bonne.

7.8 Nomenclature et structure de codage des variables de l'ELNEJ

La documentation du fichier de microdonnées de l'ELNEJ applique certaines normes de désignation et de codage des valeurs des variables, le but étant de rendre l'interprétation des données plus facile à l'utilisateur.

Nomenclature des variables

On a conçu une nomenclature pour les diverses variables du fichier de données de l'ELNEJ pour renseigner plus précisément l'utilisateur sur ces dernières. Tous les noms de variables sont d'au plus huit caractères, ce qui permet d'utiliser facilement ces désignations avec des progiciels d'analyse comme SAS ou SPSS.

Nomenclature des variables du fichier de l'éducation

"D" désigne le cycle 4 de l'ELNEJ

"EDU" désigne le fichier de l'éducation

"T" désigne des questions provenant des deux questionnaires d'enseignant

"S" désigne des questions provenant du questionnaire d'enseignant unique

"M" désigne des questions provenant du questionnaire d'enseignant multiple

"P" désigne des questions provenant du questionnaire du directeur

Nomenclature des variables des fichiers primaire et secondaire

D SE C Q nnx ou D SE C d Q nnx

D désigne le cycle de l'ELNEJ

A indique le premier cycle

B le deuxième

C le troisième, et ainsi de suite

SE - désigne la section du questionnaire où la question a été posée ou dont la variable a été tirée.

C - désigne l'unité de collecte ou l'unité à laquelle se rapporte la variable.

Il y a cinq possibilités¹ :

C indique l'enfant,

P la PMR,

S le conjoint,

H le ménage

Y l'adolescent

d - la lettre minuscule désigne le cycle de l'ELNEJ où la variable a figuré pour la première fois au dossier ou le cycle dans lequel des changements ont été apportés à une question.

Exemple :

¹ Il convient de noter que, bien qu'il y ait des variables pour une diversité d'unités d'analyse (PMR, conjoint et ménage), on ne pourra produire que des « estimations pour les enfants » à partir du fichier de microdonnées de l'ELNEJ. On peut se reporter aux caractéristiques de la PMR, du conjoint et du ménage pour décrire les caractéristiques de l'enfant. Ainsi, il sera possible d'estimer le nombre d'enfants vivant dans un ménage à faible revenu ou dont la PMR a reçu une note élevée à l'échelle de la dépression. Il sera toutefois impossible de produire des estimations du nombre de ménages à faible revenu ou de PMR déprimées.

-b indique que la variable était nouvelle au cycle 2. Dans les cycles postérieurs, la lettre minuscule qui désigne le cycle indiquera les variables nouvelles. On aura un c pour les nouvelles variables du cycle 3, un d pour celles du cycle 4, et ainsi de suite. Entre les cycles, on a pu réviser le contenu du questionnaire. Si la révision a changé le sens ou les valeurs d'une question, la variable est considérée comme nouvelle et se voit accoler la lettre d.

Q désigne la variable qui est liée à une question directement posée dans un des questionnaires de l'ELNEJ.

S désigne un résultat calculé pour une des échelles utilisées dans le questionnaire.

D signifie que la variable a été calculée à partir des données d'autres questions posées dans le questionnaire.

I signifie que la variable est un indicateur d'imputation.

Z signifie que la variable est un indicateur d'incohérence des données déclarées entre le cycle en cours et les cycles antérieurs.

« nnx » désigne la question ou la variable; en général, nn est un numéro séquentiel attribué à une variable et x, un indicateur alphabétique séquentiel pour une suite de variables du même ordre.

7.9 Siglaison des sections du questionnaire

Le tableau qui suit indique les acronymes employés pour les diverses sections des questionnaires de l'ELNEJ. Ces acronymes font partie du nom des variables du fichier de données de l'ELNEJ. Ils occupent les deuxième et troisième positions du nom de variable.

	Variable	Recueillie ou générée à partir de :
GE	Renseignements géographiques	données d'échantillon
HH	Ménage	Caractéristiques du logement
MM	Variables d'observation de la liste du ménage	variables démographiques de base pour chaque membre du ménage; celles-ci figurent au fichier de données de l'ELNEJ pour l'enfant, la PMR et le conjoint.
DM	Variables démographiques calculées qui décrivent les conditions de vie de l'enfant	données de la liste du ménage et de la grille des liens
SD	Renseignements sociodémographiques	l'enfant au questionnaire de l'enfant et de la PMR et conjoint au questionnaire de l'adulte
HL	Santé	la PMR et conjoint au questionnaire de l'adulte; de l'enfant au questionnaire de l'enfant; et de l'adolescent au questionnaire de l'adolescent.
CH	États chroniques des adultes	La PMR et conjoint à la section sur la santé du questionnaire de l'adulte
RS	Limitations d'activités	La PMR et conjoint à la section sur la santé du questionnaire de l'adulte

DP	Échelle de dépression	Questionnaire des parents (test d'échelle que subit la PMR)
ED	Scolarité	L'enfant de 4 à 15 ans au questionnaire de l'enfant; la PMR et conjoint au questionnaire de l'adulte; et de l'adolescent au questionnaire de l'adolescent
LF	Activité professionnelle	La PMR et conjoint au questionnaire de l'adulte; et de l'adolescent au questionnaire de l'adolescent.
IN	Revenu	revenu du ménage et revenu personnel de la PMR au questionnaire de l'adulte; et celui de l'adolescent au questionnaire de l'adolescent
FN	Échelle de fonctionnement de la famille	Questionnaire de l'adulte (administré à la PMR ou au conjoint)
MD	Renseignements médicaux et biologiques	Questionnaire de l'enfant (de 0 à 3 ans)
TM	Tempérament	Questionnaire de l'enfant (3 mois à 35 ans)
LT	Alphabétisation	Questionnaire de l'enfant (0 à 9 ans)
AA	Activités	Questionnaire de l'enfant (0 à 15 ans) et de l'adolescent au questionnaire de l'adolescent
BE	Comportement	Questionnaire de l'enfant (0 à 11 ans)
MS	Développement moteur et social	Questionnaire de l'enfant (0 à 47 mois)
RL	Rapports sociaux	Questionnaire de l'enfant (4 à 9 ans)
PR	Style parental	Questionnaire de l'enfant (0 à 15 ans)
CR	Garde des enfants	Questionnaire de l'enfant (0 à 13 ans)
PP	Test EPIV	enfants de 4 à 6 ans (enfants de plus de 6 ans s'ils sont en 1 ^{re} année ou moins)
FF	Amis et famille	Questionnaires auto-administrés des 10 - 17 ans
SC	École	Questionnaires auto-administrés des 10 - 17 ans
AM	Ce que je suis, moi	Questionnaires auto-administrés des 10 - 17 ans
FB	Sentiments et comportements	Questionnaires auto-administrés des 10 - 17 ans
PM	Mes parents et moi	Questionnaires auto-administrés des 10 - 17 ans
PU	Puberté	Questionnaires auto-administrés des 10 - 17 ans
DR	Usage du tabac, de l'alcool et des drogues	Questionnaires auto-administrés des 10 - 17 ans
AT	Activités	Questionnaires auto-administrés des 10 - 17 ans
HT	Santé	Questionnaires auto-administrés des 10 - 17 ans
WK	Travail et sources d'argent	Questionnaires auto-administrés des 10 - 17 ans
DA	Sorties	Questionnaires auto-administrés des 10 - 17 ans
MA	Test de calcul	enfants en 2 ^e année et plus et Mesure cognitive au 16-17 ans
SF	Sécurité du voisinage	Questionnaire de l'adulte (section administrée à la PMR au conjoint(e))

SP	Soutien social	Questionnaire de l'adulte (section administrée à la PMR au conjoint(e))
SL	Sommeil	Questionnaire de l'enfant (0 à 13 ans)
PB	Comportement positif	Questionnaire de l'enfant (3 à 5 ans)
AS	Aspirations	Questionnaire de l'enfant (16 à 17 ans)
AG	Étapes du développement	Questionnaire des parents (3 à 71 mois)
WM	Qui suis-je?	Mesure directe (4 à 5 ans)
KN	Connaissance des nombres	Mesure directe (3 à 5 ans)
OB	Évaluation de l'observation	Tous les enfants qui ont reçu la mesure directe
TS	Enseignant unique	Questionnaire de l'enseignant unique
TM	Enseignants différents	Questionnaire des enseignants différents
PC	Directeur	Questionnaire du directeur

7.10 Exemples de noms de variables

Voici des exemples qui illustrent la nomenclature des variables au fichier de données de l'ELNEJ :

Nom de la variable	Objet :
DLFSQ2	Q2 de la section sur l'activité professionnelle pour le conjoint
D	Variable du cycle 4
LF	Section sur l'activité professionnelle
S	Conjoint
Q	Question directement posée dans le questionnaire
2	Identificateur de l'élément

Nom de la variable	Objet :
DPRCS03	Note d'interaction positive à l'échelle du rôle parental pour un enfant de 2 à 15 ans
D	Variable du cycle 4
PR	Section sur le rôle parental
C	Enfant
S	Note ou résultat
03	Identificateur de la variable

7.11 Structure de codage des variables de l'ELNEJ

On a conçu des normes pour la structure de codage des variables de l'ELNEJ afin de décrire certaines situations en toute cohérence pour toutes les variables. Voici ces diverses situations avec leurs codes descriptifs respectifs.

Refus

Sur le fichier principal de données, le code 8 est attribué à une question à laquelle le répondant a refusé de répondre. Une variable à un chiffre reçoit le code 8, une variable à deux chiffres, le code 98, une variable à trois chiffres, le code 998, etc. Comme les réponses sont inscrites par le répondant sur le questionnaire papier, ce code particulier n'est pas utilisé pour les variables de ces questionnaires.

Cas « ne sait pas »

Le répondant peut ignorer la réponse à une question. Là encore, le système IAO comporte une touche de fonction qui permet de décrire cette situation.

Dans le fichier de données de l'ELNEJ, l'ignorance d'une réponse reçoit le code 7. Une variable à un chiffre se voit attribuer le code 7, une variable à deux chiffres, le code 97, une variable à trois chiffres, le code 997, et ainsi de suite.

Cas « sans objet »

Dans certains cas, une question ne s'applique pas au répondant. Dans le fichier de données, on se sert des codes 6, 96, 996, etc., pour indiquer qu'une question ou une variable dérivée est sans objet.

Dans certains cas, une question ou une série de questions était sans objet. Par exemple, la question sur les styles d'enseignement de la lecture ne s'applique qu'aux enseignants des enfants à qui on enseigne cette matière. Autrement, le code 6 sera attribué à cette question.

Cas « non déclaré »

Dans le cadre du traitement qui se fait au bureau central, on code parfois comme « non déclaré » la réponse à une question. On indique par là que la réponse à la question est inconnue. On attribue de tels codes pour quelques raisons :

- Dans certains cas, il était impossible de remplir un questionnaire ni même parfois de le commencer. Par exemple, il se peut qu'il y ait eu une interruption

de quelque sorte et que le questionnaire partiellement rempli ait été retourné par la poste. Si on avait recueilli suffisamment de données sur le répondant, on attribuait le code « non déclaré » à toutes les autres questions sans réponse du questionnaire. Comme seule exception, si on savait qu'une section ou un questionnaire ne s'appliquait pas, on codait ces éléments comme « sans objet ».

- Dans une autre situation, on a recouru à des codes « non déclaré » à la suite de vérifications de cohérence. Si on décelait une erreur de cohérence entre des groupes de variables, une ou plusieurs des variables en question se voyaient attribuer le code « non déclaré ».

Dans le cas des variables dérivées, si une ou plusieurs des variables d'entrée avaient reçu le code « refus », « ne sait pas » ou « non déclaré », on faisait de même pour la variable de sortie.

7.12 Indicateurs d'imputation

Variables manquantes

Certaines variables de la composante Éducation peuvent avoir des données manquantes du fait de la non-réponse partielle. En effet, certains répondants peuvent ne pas avoir répondu à toutes les questions. Cette même situation se retrouve également dans la composante principale.

Imputation

La composante Éducation comporte des questions posées aux enseignants et aux directeurs d'école. Si l'une de ces composantes était répondue et l'autre totalement non-répondue, la composante non-répondue a été imputée entièrement. Par contre, dans le cas où les deux composantes sont entièrement non-répondues, l'enregistrement est considéré comme non-répondant et aucune imputation n'a été faite.

L'imputation consiste à déterminer des valeurs vraisemblables pour des réponses manquantes (voir chapitre 11). Les deux indicateurs d'imputation du fichier de microdonnées de la composante Éducation de l'ELNEJ sont FLAG_TE et FLAG_PR. FLAG_TE = 1 indique que les données de l'enseignant (TEacher) sont imputées. Elle vaut 0 sinon. La variable FLAG_PR = 1 indique que les données du directeur (PRincipal) sont imputées. La variable vaut 0 sinon.

Veuillez noter que si une ou quelques unes des variables d'une composante répondue présentent de la non-réponse, cette ou ces variables ne sont pas imputées.

7.13 Variables dérivées

Combinaison de questions

Afin de faciliter le travail d'analyse, des variables ont été dérivées à partir des réponses obtenues à des groupes de questions spécifiques. Par exemple, les questions au sujet des années et des mois d'expérience d'un enseignant ou d'un directeur ont été combinées en une seule variable indiquant le nombre total de mois d'expérience déclaré.

Nom des variables dérivées

Toutes les variables dérivées dans le fichier de microdonnées de l'ELNEJ reçoivent un D en cinquième position dans leur nom.

8.0 Contenu de l'enquête

L'ELNEJ a été conçue en fonction d'une approche écologique ou holistique du développement de l'enfant. L'enquête permet de saisir toute la diversité et la dynamique des facteurs ayant une incidence sur les enfants. Pour assurer une couverture adéquate de tous les domaines pertinents qui influent sur le développement de l'enfant, une consultation multidisciplinaire a eu lieu dès la conception de l'enquête. Le choix des priorités, des questions d'enquête et des domaines particuliers a résulté en grande partie d'un effort concerté soutenu par la contribution et les conseils :

- du groupe consultatif expert de l'ELNEJ, composé de spécialistes du développement de l'enfant et des sciences sociales;
- de ministères fédéraux;
- de représentants provinciaux et territoriaux responsables des programmes liés au développement de l'enfant.

Le groupe a recommandé que l'ELNEJ porte sur une large gamme de caractéristiques et de facteurs influant sur la croissance et le développement de l'enfant. On a recueilli une foule de renseignements au sujet de l'enfant, de son ou ses parents, des caractéristiques de la famille et du quartier ainsi que de l'école de l'enfant et sa vie à l'école. La présente section donne un aperçu du contenu de chacune des sections du questionnaire ayant servi à la collecte des données de l'ELNEJ.

8.1 Système de traitement de l'enquête

Le système de traitement de l'ELNEJ comprend certaines vérifications de base de la qualité qui sont effectuées pour chaque section du questionnaire. Toutes les questions ayant donné lieu à un taux élevé de non-réponse ou à de nombreux rejets par suite de la vérification ont été examinées de façon détaillée. Au besoin, des comparaisons ont été faites avec des sources extérieures de données et des analyses ont été effectuées pour cerner les raisons possibles des écarts par rapport à ces autres sources. Les préoccupations relatives aux éventuels problèmes de qualité des données pour quelques questions d'une section particulière sont traitées dans ce chapitre.

La question de la validation des scores est abordée au chapitre 9.

8.2 Variables démographiques

Les variables démographiques sont recueillies dans le questionnaire du ménage. À partir de ce questionnaire, certains renseignements démographiques de base (p. ex., l'âge, le sexe, l'état matrimonial) sont recueillis pour tous les membres du ménage de l'enfant. Ce questionnaire vise en outre à recueillir des données sur la grille des liens, c'est-à-dire les liens entre tous les membres du ménage et la PMR. Ces données ont permis de créer un vaste ensemble de variables pour décrire la situation familiale de l'enfant. La plupart de ces variables dérivées sont essentielles à l'analyse des données de l'ELNEJ et sont décrites au guide de l'utilisateur du fichier principal.

8.3 Questionnaire de l'adulte

Pour des plus amples renseignements, voir le guide de l'utilisateur du fichier principal.

Éducation (parent) - La section sur le niveau de scolarité est remplie pour la PMR et son conjoint/partenaire. L'objectif est de recueillir des renseignements sur le nombre d'années d'études terminées, le plus haut niveau de scolarité atteint ainsi que la fréquentation actuelle d'un établissement d'enseignement.

Activité sur le marché du travail - La section relative à l'activité sur le marché du travail est remplie à la fois pour la PMR et son conjoint/partenaire. La stabilité d'emploi a un effet sur le milieu familial, tant sur le plan du revenu que sur celui du niveau de stress. Certaines recherches montrent que le chômage des parents peut avoir un effet néfaste sur les résultats des enfants.

Revenu - Dans la section de l'enquête portant sur le revenu, on recueille des renseignements sur les sources et le montant du revenu pour chacun des ménages. Des données sur la fourchette de revenu sont également recueillies pour la PMR et son conjoint. Ces renseignements constituent un indicateur de la situation économique de la famille, une composante essentielle du cadre de vie de l'enfant.

Santé - Cette section comporte des questions adressées aux PMR et à leurs conjoints au sujet de l'état général de santé, les états chroniques, les limitations d'activités, le tabagisme et la consommation d'alcool. On a inclus les questions sur le tabagisme parce que des recherches ont indiqué que le comportement des parents à cet égard pourrait permettre de prévoir le tabagisme chez les enfants. La consommation d'alcool est abordée en raison de son incidence potentielle sur la santé physique et mentale de l'adulte, la situation économique de la famille et les relations familiales.

Sécurité du quartier - Cette section permet de recueillir des renseignements sur la satisfaction des répondants quant à leur quartier en tant que milieu de vie pour les enfants, y compris leurs perceptions relativement à l'importance des dangers et des problèmes, à la cohésion sociale ou sentiment de « voisinage ».

Caractéristiques sociodémographiques - La section sur les caractéristiques sociodémographiques vise à recueillir des renseignements sur l'immigration, l'origine ethnique et le profil linguistique des membres du ménage. Ces données permettent l'analyse des diverses composantes de la population canadienne et l'identification des personnes appartenant à une minorité visible. La section comporte en outre une question sur l'appartenance religieuse. Ces questions se rapportent à la PMR, à son conjoint et à l'enfant.

En raison de problèmes techniques, on n'a posé aucune question sociodémographique aux répondants des cycles 3 et 4 ayant participé aux cycles antérieurs. Les données du fichier représentent les réponses fournies par les répondants au cycle 2. Il convient d'aviser les utilisateurs que l'information sur la ou les langues parlées, tout particulièrement, peut avoir changé depuis la dernière collecte de données sur le sujet. Il convient donc d'utiliser les données associées à ces variables avec prudence.

8.4 Questionnaire de l'enfant

Éducation (enfant) - Cette section vise à obtenir des renseignements pour tous les groupes d'âge. On recueille les renseignements de base sur les sujets suivants : le

niveau scolaire, le genre d'école fréquentée et la langue d'enseignement, l'attrait qu'exerce l'école sur l'enfant, l'absentéisme, le nombre de changements d'école et de déménagements.

Dans le cas des enfants de première année ou plus avancés, on pose des questions supplémentaires sur d'autres aspects, comme le saut d'une classe et le redoublement scolaire, le rendement scolaire et l'éducation spécialisée.

Santé (enfant) - Cette section vise à recueillir des renseignements sur l'état de santé physique de l'enfant – état général de santé, blessures, limitations et états chroniques – de même que sur le recours aux services de santé et aux médicaments.

Renseignements médicaux et biologiques - La section des renseignements médicaux et biologiques vise les enfants de 0 à 3 ans. L'objectif principal consiste à recueillir des renseignements sur certains facteurs, comme l'âge gestationnel et le poids à la naissance, dont il a été démontré qu'ils ont un effet direct sur la croissance et le développement de l'enfant. Par exemple, à long terme, les bébés ayant un poids insuffisant courent des risques plus élevés d'être en mauvaise santé et d'éprouver des problèmes de développement persistants.

Travail de la mère après la naissance d'un enfant - On a posé ces questions pour déterminer la période écoulée entre la naissance d'un enfant et le retour au travail de la mère ainsi que le niveau de participation des mères au marché du travail après leur retour.

Principales étapes du développement - Ces questions ont été ajoutées au cycle 4 pour obtenir une meilleure mesure du développement des jeunes enfants. Ensemble, les principales étapes du développement (p. ex., les premiers mots ou les premiers pas) donnent une indication générale du développement de l'enfant.

Tempérament - Cette section permet de mesurer le tempérament des jeunes enfants selon la perception qu'ont les parents du degré de difficulté que présente leur enfant. Cette mesure est fondée sur l'hypothèse que le tempérament de l'enfant est conditionné non seulement par des facteurs biologiques, mais aussi par les perceptions des parents quant au degré de difficulté de l'enfant.

Alphabétisation - Cette section permet de mesurer le contact qu'on les enfants avec les livres, leur intérêt pour la lecture et les activités d'apprentissage auxquelles les parents s'adonnent avec leurs enfants. La section porte principalement sur la stimulation des jeunes enfants à la maison.

Communications - Ces questions sont adaptés à partir de l'étude « *Competent Children* » de la Nouvelle-Zélande. Elles portent sur les capacités de l'enfant de comprendre les messages oraux, de transmettre un message à autrui et de communiquer verbalement. La dernière question, portant sur la bonne compréhension des messages verbaux, n'est posée qu'au sujet des enfants de 3 ans.

Activités - Cette section permet de mesurer la participation des enfants à diverses activités non scolaires ainsi que les responsabilités qu'assument les enfants âgés de 10 et 11 ans à la maison. Ces questions servent à établir l'échelle des responsabilités familiales (DACCS6) qui mesure le degré de responsabilités assumées à la maison. Cette section donne certaines indications quant à l'emploi du temps de l'enfant, ses intérêts personnels et son degré d'interaction avec ses pairs.

Comportement - L'objectif de cette échelle est d'évaluer certains aspects du comportement des enfants de 2 ans et plus ainsi que les habitudes alimentaires des

enfants âgés de 1 an à 3 ans. Les questions de cette section servent à mesurer la fréquence de comportements tels que l'hyperactivité et l'agressivité physique.

Comportement positif - Cette nouvelle section vise à évaluer le comportement positif des enfants âgés de 3 à 5 ans, notamment la persévérance et l'autonomie. Selon l'étude « *Competent Children* » de la Nouvelle-Zélande, la persévérance et l'autonomie figurent parmi l'ensemble des compétences constituant de bons indicateurs du rendement général de l'enfant. Les questions sont inspirées de l'étude néo-zélandaise et des questions de l'ELNEJ se rapportant au comportement des enfants d'autres groupes d'âge.

Sommeil - Les recherches révèlent que les troubles du sommeil permettent de prévoir des difficultés éventuelles chez les enfants. À l'inverse, l'absence de tels troubles est corrélée avec un tempérament facile et un développement positif. Plusieurs nouvelles questions s'inspirant de l'Étude longitudinale du développement des enfants du Québec (ELDEQ), de Santé Québec ont été ajoutées au cycle 4. Ces questions portent sur les heures de sommeil, les heures de sommeil ininterrompu par nuit, combien de fois le sommeil des parents a été interrompu par l'enfant, etc.

Développement moteur et social - L'échelle du développement moteur et social (DMS) mesure divers aspects du développement moteur, social et cognitif des enfants âgés de 0 à 3 ans; les questions varient selon l'âge des enfants. Trois scores (DMSCS01, DMSCS02 et DMSCdS03) sont établis à partir de ces questions.

Relations - L'objectif de cette section consiste à recueillir des renseignements sur les relations que les enfants entretiennent avec les autres. Les relations positives avec d'autres enfants et des adultes peuvent contribuer à neutraliser les effets d'autres facteurs qui entraînent des risques pour l'enfant. Les questions touchant les activités avec les amis et les relations avec les parents, les enseignants et les amis sont fondées sur celles de l'Étude sur la santé des jeunes ontariens.

Rôle parental - On estime que le rôle parental exerce une influence importante sur le comportement et le développement de l'enfant. L'objectif de cette section est de mesurer certains comportements parentaux. On a créé des échelles à partir des questions de cette section.

8.5 Avertissement

Contrairement aux cycles précédents le fichier de l'éducation du cycle 4 a été repondéré en raison des faibles taux de retour des questionnaires. On a créé un poids transversal afin de contrebalancer la réduction de l'échantillon. Une analyse a été effectuée à l'aide d'une sélection de variables de la composante principale, qui ont été incluses sur le fichier de l'éducation. En effectuant un ajustement proportionnel de ces variables sélectionnées, on a créé un poids ajusté.

Grâce à l'inclusion de variables de la composante principale sur ce fichier, on espère que les utilisateurs pourront effectuer la plus grande partie de leur analyse transversale sans avoir recours à la composante principale. Si les utilisateurs ont besoin du fichier principal, ils doivent noter que leurs estimations peuvent différer pour les variables qui n'ont pas été utilisées pour le calibrage, étant donné que tout écart de couverture entre le fichier principal et le fichier du sous-échantillon ne serait pas pris en compte par le poids ajusté de l'éducation. De même, pour toute analyse longitudinale, on aura besoin du fichier principal afin d'obtenir le poids longitudinal; les données de l'éducation ne s'appliqueront

qu'à un sous-ensemble du fichier complet, et il n'y aura pas d'ajustement pour tenir compte des données manquantes.

9.0 Validation des échelles de l'ELNEJ

9.1 Validation des données d'échelle

9.1.1 Définition de l'échelle

Pour certains des concepts dont la mesure a été jugée importante pour l'ELNEJ, on a décidé qu'il conviendrait mieux d'utiliser une échelle. Une échelle est constituée simplement d'un groupe de questions dont les réponses mises ensemble permettent de mesurer certains concepts.

Par exemple, à partir du questionnaire de l'enfant, on a déterminé qu'il serait important d'évaluer certains comportements parentaux. L'échelle vise à mesurer trois concepts ou facteurs liés au rôle parental : l'interaction positive, l'inefficacité parentale et la cohérence des pratiques parentales.

9.1.2 Échelles et calculs

Pour chaque facteur mesuré au moyen d'une échelle, on calcule un score (note). Le score attribué à un facteur particulier peut servir à classer les personnes. Par exemple, pour les enfants ayant les scores les plus élevés sur l'échelle d'interaction positive, les PMR ont déclaré avoir plus de rapports positifs avec l'enfant (ils riaient ensemble davantage, l'enfant était félicité plus souvent, etc.). Le score pour un facteur particulier est généralement fondé sur une série de questions, étant donné qu'une seule question ne peut mesurer le facteur ou le concept avec suffisamment de précision.

Au moment d'élaborer l'ELNEJ, lorsqu'on a pensé aux échelles particulières qui devraient servir à mesurer un concept donné, on a choisi autant que possible des échelles utilisées dans le cadre d'autres études. On connaissait ainsi les propriétés psychométriques des mesures produites et on disposait de références complètes.

9.1.3 Évaluation des données exprimées selon une échelle

Dans nombre de cas, on a modifié le libellé de certaines questions et, dans certains cas, on a ajouté de nouvelles questions. Parfois, l'échelle utilisée n'avait jamais servi à des enquêtes sur les enfants au Canada ou n'avait été appliquée qu'à de très petits échantillons. Compte tenu de ces préoccupations et d'autres relatives aux conditions dans lesquelles se dérouleraient les interviews, on était d'avis que les structures des facteurs des échelles utilisées dans le cadre de l'ELNEJ pourraient différer de celles décrites dans les ouvrages spécialisés. Par conséquent, l'équipe chargée du projet a décidé qu'il fallait procéder à une évaluation approfondie des données exprimées selon une échelle, afin de s'assurer que les propriétés psychométriques présentées dans les autres études s'appliquaient à l'ELNEJ.

L'analyse des données exprimées selon une échelle s'est faite en trois grandes étapes. Tout d'abord, une nouvelle analyse factorielle a été effectuée pour toutes les échelles, afin de déterminer les concepts ou facteurs inhérents à chacune. Puis les scores ont été calculés selon cette structure factorielle. Enfin, des mesures de fiabilité ont été produites. Les procédures employées pour chacune de ces étapes sont décrites en détail ci-dessous.

NOTE : Bon nombre des échelles ont été élaborées et validées au cycle 1. Au cours des cycles subséquents, on a imposé la structure factorielle émanant de l'analyse du cycle 1. Cette façon de procéder assure la cohérence des échelles au fil du temps et permet les analyses longitudinales ainsi que les comparaisons transversales. Pour chaque échelle, une note indique à quel cycle la validation a été effectuée.

9.2 Analyse factorielle

9.2.1 Analyse factorielle pour les échelles

La détermination de la structure factorielle de chacune des échelles s'est faite à partir des données du premier cycle. La structure factorielle imposée aux échelles employées au cycle 1, puis réutilisées aux cycles 2, 3 et 4 de l'enquête est le résultat d'analyses des données du cycle 1. On a repris l'analyse après le cycle 4, et les résultats de cette analyse sont résumés ci-dessous. Les résultats de l'analyse factorielle du premier cycle sont présentés dans le guide de l'utilisateur du cycle 1.

1. Pour chaque échelle (et groupe d'âge, si l'échelle utilisait des questions différentes pour les divers groupes d'âge), l'échantillon de répondants a été divisé au hasard en deux sous-échantillons. On a procédé ainsi pour déterminer si des échantillons différents produiraient les mêmes résultats.
2. Pour chacun des sous-échantillons, une analyse des composantes principales a été effectuée afin de déterminer le nombre de facteurs à extraire dans le cadre de l'analyse factorielle subséquente. En principe, on s'attendait à arriver au même nombre de facteurs que ceux décrits dans les ouvrages spécialisés. En pratique, toutefois, certaines échelles ont fait ressortir un nombre différent de facteurs parce que, dans certains cas, des facteurs se sont combinés, tandis que dans d'autres, de nouveaux facteurs sont apparus.
3. L'analyse factorielle a été effectuée pour chaque sous-échantillon, et la structure des facteurs ainsi que les items de chaque facteur ont été comparés.
4. Pour effectuer les analyses, on a utilisé la valeur des items attribuée à chaque enfant du groupe d'âge approprié, multipliée par le poids normalisé de l'enfant. On normalise le poids d'un individu en divisant son poids (DWTWC01C) par la moyenne des poids de tous les individus. Ainsi, la somme des poids normalisés est égale à la taille de l'échantillon.

5. Une fois les structures factorielles établies et les items inclus dans chaque facteur déterminé, on procède au calcul des scores. Pour produire les scores, on a retranché 1 de la valeur de chacun des items afin que la plus faible valeur possible pour le score soit 0. Un score de 0 indique l'absence de problème chez l'enfant pour tout facteur de l'échelle de comportement à l'exception du facteur « prosocial », pour lequel un score de 0 indique l'absence de comportement prosocial. Certains items ont été imputés. Les valeurs à imputer ont été déterminées grâce à une procédure (la procédure PRINQUAL du progiciel SAS) qui indique parmi les valeurs possibles d'un item celle qui semble la plus plausible pour un individu étant donné son profil de réponse, le profil de réponse des autres individus dans l'échantillon ainsi que le nombre de facteurs considérés dans l'analyse.
6. On calcule le score pour chacun des facteurs de l'échelle en faisant la somme des valeurs des items qui composent les facteurs (en tenant compte des valeurs imputées). Le score est désigné « manquant » si trop de valeurs pour un item compris dans un facteur ne sont pas déclarées. Une valeur peut être manquante si le parent n'a pas voulu répondre ou s'il ne connaissait pas la réponse à la question.

9.2.2 Transformation des données au moyen de l'échelonnement optimal

Avant de procéder à l'analyse factorielle des échelles de l'ELNEJ, les données ont été transformées au moyen de l'échelonnement optimal. La méthode utilisée était l'une de celles proposées par Young et plusieurs collaborateurs (Young, 1981), qui constitue une variante de la méthode d'échelonnement optimal de Fisher. La méthode est présentée comme un moyen de transformer des données qui sont essentiellement nominales ou ordinales, en intervalles ou en ratios, afin que les techniques statistiques s'appliquant à ces derniers puissent être utilisées.

9.2.3 Analyse factorielle axée sur les données pondérées

L'analyse factorielle nécessite que les données aient des propriétés d'intervalle ou de ratio, c'est-à-dire que la distance entre chaque catégorie de réponse à la question soit la même. Par exemple, dans les échelles dont les choix de réponse sont : « jamais », « parfois », « souvent » et « toujours », on doit partir du principe que l'écart entre « jamais » et « parfois » est le même que celui entre « parfois » et « souvent », du point de vue du répondant. On estime que cela n'est pas toujours le cas pour les échelles utilisées dans l'ELNEJ, et des mesures ont été prises pour neutraliser cette difficulté (transformation des données).

9.3 Calcul des scores et imputation des réponses

9.3.1 Calcul des scores pour chaque facteur

Les résultats des analyses factorielles ont été utilisés pour déterminer les items « rattachés » à chaque facteur. L'étape suivante a consisté à calculer un score pour chaque facteur. À cette fin, on a additionné les valeurs de chaque question faisant partie du facteur. Dans certains cas, on a rééchélonné des valeurs avant de calculer le score final. Voici un exemple du calcul des scores des facteurs.

9.3.2 Exemple de calcul des scores de facteurs

Dans le cas de l'échelle relative au rôle parental dans le questionnaire de l'enfant, un des concepts qui est ressorti des analyses de facteurs est celui de l'inefficacité parentale (pour les enfants âgés de 2 à 11 ans). L'analyse factorielle des données du cycle 1 a cerné sept questions s'y rattachant.

DPRCQ04	À quelle fréquence vous arrive-t-il d'être contrarié par une parole ou un geste que [votre enfant] n'est pas censé/ée dire ou faire?
DPRCQ08	Lorsque vous parlez à [votre enfant] de son comportement, dans quelle proportion du temps le/la félicitez-vous?
DPRCQ09	Lorsque vous lui parlez de sa conduite, dans quelle proportion du temps le/la désapprouvez-vous?
DPRCQ13	À quelle fréquence vous mettez-vous en colère lorsque vous punissez [votre enfant]?
DPRCQ14	À quelle fréquence croyez-vous que le genre de punition que vous lui donnez dépend de votre humeur?
DPRCQ15	À quelle fréquence croyez-vous avoir de la difficulté à vous y prendre avec lui/elle en général?
DPRCQ18	À quelle fréquence devez-vous le/la punir à maintes reprises pour la même chose?

Ces questions sont associées à des catégories de réponse de deux genres :

- 1 - jamais
- 2 - environ une fois par semaine ou moins
- 3 - quelques fois par semaine
- 4 - une ou deux fois par jour
- 5 - plusieurs fois par jour

- 1 - jamais
- 2 - moins de la moitié du temps
- 3 - environ la moitié du temps
- 4 - plus de la moitié du temps
- 5 - tout le temps

Au moment du calcul du score de ce facteur d'inefficacité parentale, les catégories ont été rééchélonnées de 0 à 4 (la catégorie « jamais » correspondant à 0, la catégorie « environ une fois par semaine ou moins/moins de la moitié du temps » à 1,... et la catégorie « plusieurs fois par jour/tout le temps » à 4). Pour le calcul du score, ces valeurs ont été additionnées pour les sept questions comprises dans le facteur avec un résultat de 0 à 28 pour l'inefficacité parentale.

Un score de 0 indique un style parental plus efficace, tandis qu'un score élevé de 28 indique un style parental moins efficace. Pour la plupart des scores calculés dans le cadre de l'ELNEJ, un score de 0 correspond à une absence de problème. Toutefois, il existe des exceptions à cette règle, lesquelles sont signalées dans la documentation relative à chaque échelle.

9.3.3 Poids factoriel négatif

Il convient de souligner que la deuxième question incluse dans le facteur d'inefficacité parentale, DPRCQ08 (Lorsque vous parlez à [votre enfant] de son comportement, dans quelle proportion du temps le/la félicitez-vous?), va dans la direction opposée de celle des autres questions. En réalité, la question a une incidence « négative » sur le facteur. Par conséquent, dans le calcul du score, les valeurs pour cette question ont été inversées, les réponses « tout le temps » correspondant à 0, « plus de la moitié du temps » à 1... et « jamais » à 4. Dans la documentation relative à chaque échelle, toutes les questions dont l'algorithme de notation a été inversé en raison d'un poids factoriel négatif comportent une indication à cet effet.

9.3.4 Code de non-réponse

Le score pour le facteur d'inefficacité parentale est étiqueté DPRCS04 dans le cliché d'article du fichier de microdonnées. Le « S » qui figure au cinquième caractère du nom de la variable indique qu'il s'agit d'un score.

Lors du calcul du score pour chaque facteur, il est arrivé que l'une ou plusieurs des questions prises en compte dans le calcul comportent un code de non-réponse (ne sait pas, refus ou non déclaré). Si le nombre de questions comportant un code de non-réponse se situait au-dessus d'un certain seuil, le score du facteur était établi à « non déclaré ». En général, le seuil a été fixé à 10 % des réponses.

9.3.5 Items bruts

Il convient de souligner qu'outre les scores, les items bruts pour chaque échelle figurent dans le fichier de microdonnées. Cela permettra aux chercheurs d'opter pour d'autres structures de facteurs s'ils le souhaitent. Dans le cas des items bruts, les valeurs originales (variant de 1 à 5 pour l'échelle relative au rôle parental), avant le rééchelonnement ou le renversement des valeurs, ont été retenues.

9.4 Mesures de fiabilité des échelles

La fiabilité désigne la précision, la fidélité, la cohérence ou la capacité de reproduire une échelle donnée. En termes plus techniques, la fiabilité indique jusqu'à quel point les scores d'échelle sont exempts d'erreurs de mesure. Il existe de nombreuses façons de mesurer la fiabilité.

9.4.1 Coefficient alpha de Cronbach

Un des coefficients de fiabilité les plus courants est le coefficient alpha de Cronbach (Cronbach, 1951). Le coefficient alpha mesure la cohérence interne des items compris dans un facteur. Il est fondé sur la covariance moyenne de ceux-ci. On part du principe que les items compris dans un facteur comportent une corrélation positive les uns avec les autres, étant donné qu'ils visent à mesurer, jusqu'à un certain point, une entité ou un concept commun.

9.4.2 Interprétation du coefficient alpha de Cronbach

Le coefficient alpha de Cronbach donne lieu à plusieurs interprétations. Il peut être considéré comme la corrélation entre une échelle ou un facteur et toutes les autres échelles possibles comprenant le même nombre de questions, lesquelles pourraient être tirées d'un univers hypothétique de questions mesurant la caractéristique visée. Pour ce qui est du facteur de l'inefficacité parentale, par exemple, les sept questions incluses dans l'échelle peuvent être considérées comme un échantillon de l'univers de toutes les questions possibles. On aurait pu aussi demander aux parents : « Combien de fois élevez-vous la voix lorsque vous grondez votre enfant? » ou « Combien de fois menacez-vous de punir votre enfant sans le faire? ». Le coefficient alpha de Cronbach indique le niveau de corrélation prévu entre l'échelle utilisée et toutes les autres échelles possibles à sept questions qui mesurent le même phénomène.

Le coefficient de Cronbach peut aussi être interprété comme le carré de la corrélation entre le score obtenu par une personne pour un facteur en particulier (score observé) et le score que cette personne aurait obtenu si on lui avait posé toutes les questions possibles de l'univers (score réel). Étant donné que le coefficient est interprété comme un coefficient de corrélation, sa valeur se situe entre 0 et 1.

On a démontré que le coefficient est en général la limite inférieure de fiabilité d'une échelle de n questions (Novick et Lewis, 1967). Autrement dit, dans la plupart des cas, le coefficient donne une estimation prudente de la fiabilité d'un score.

9.4.3 Qu'est-ce qu'un niveau satisfaisant de fiabilité?

Il est difficile de préciser un niveau unique qui pourrait s'appliquer à toutes les situations. Certains chercheurs sont d'avis que la fiabilité ne devrait pas être inférieure à 0,8 pour les échelles largement utilisées. À ce niveau, les corrélations sont peu touchées par l'erreur de mesure aléatoire. Cependant, il est souvent très coûteux, en argent et en temps, d'obtenir un coefficient plus fiable. Il convient de souligner que dans le cas de certains des facteurs pour lesquels des scores ont été calculés dans le cadre de l'ELNEJ, la fiabilité est inférieure à ce niveau. Le coefficient alpha de Cronbach figure dans la documentation relative à chaque score calculé. Les chercheurs peuvent déterminer eux-mêmes si le score est suffisamment fiable pour les fins auxquelles ils le destinent.

Enfin, il importe de noter que pour l'ELNEJ, le coefficient alpha de Cronbach de chaque score de facteur a été calculé au moyen de SAS. En règle générale, les coefficients calculés par SAS sont inférieurs à ceux calculés à l'aide de SPSS.

9.5 Échelles de l'éducation du cycle 4

Les échelles provenant du fichier principal ne sont pas décrites ici; pour de plus amples renseignements, on peut se reporter au guide de l'utilisateur du fichier principal.

Les échelles relatives aux enseignants

Cette section particulière de l'Enquête longitudinale nationale sur les enfants et les jeunes a été conçue dans le but de mieux connaître les effets des variables relatives à l'éducation sur les résultats des enfants. Des questionnaires ont été remplis par les enseignants de chaque enfant d'âge scolaire admissible à l'enquête pour lequel les parents ainsi que le conseil scolaire ont donné leur consentement. Les diverses échelles comprises dans cette section ont été élaborées par le Centre de la statistique de l'éducation de Statistique Canada en consultation avec l'équipe de projet de l'ELNEJ, de même qu'avec D. Willms et F. Echols, de la University of British Columbia. Il faut également mentionner que pour le score relatif à l'indice des compétences sociales, la liste des compétences sociales/personnelles et des habitudes de travail est tirée de la recherche de Freeman et Hatch (1989) sur les commentaires souvent formulés sur les bulletins scolaires.

La partie suivante comprend la liste des échelles contenues dans les questionnaires des enseignants de l'ELNEJ et fournit des précisions sur les cycles où ces échelles ont été utilisées et sur les questions qui ont été ajoutées ou enlevées. Comme l'âge des enfants de l'échantillon augmentait, on a dû élaborer des questions et des échelles qui tiennent compte de ce fait. Ainsi, au cycle 1, on a dû élaborer un questionnaire pour les enseignants de la maternelle. Au cycle 2, le questionnaire a été adapté pour tenir compte des enfants qui avaient un seul enseignant pour différentes matières et des enfants qui avaient différents enseignants pour différentes matières. Au cycle 3, pour augmenter encore davantage la précision de l'enquête, on a donné un questionnaire distinct aux professeurs de langues et de mathématiques des enfants qui avaient des enseignants différents pour différentes matières. Au cycle 4, on a utilisé des questionnaires semblables à ceux du cycle 2.

Peu de changements ont été apportés aux échelles utilisées dans les questionnaires des enseignants de l'enquête. La variable **DETCQ17G** (à quelle fréquence l'élève fait-il preuve de confiance en lui-même) a été ajoutée à l'indice des compétences sociales pour les cycles 2 et 3. Le score relatif à « l'expérience scolaire » a été renommé score relatif à la « maturité scolaire » pour le cycle 3.

Les scores des diverses échelles relatives aux enseignants

L'analyse qui suit a été effectuée à l'aide des données du cycle 3 de l'ELNEJ recueillies en 2000, mais les noms des variables commencent par la lettre « D » afin de refléter le cycle courant.

Score relatif à l'indice des compétences sociales :

Ce score a été calculé à l'aide des items suivants : DETCQ17A, DETCQ17B, DETCQ17C, DETCQ17D, DETCQ17E, DETCQ17F, DETCQ17H, DETCQ17I et DETCQ17J. Aucune imputation n'a été effectuée pour ce score. Le score varie entre 0 et

36. Un score élevé indique que l'enfant démontre de très bonnes compétences sociales et personnelles en classe. Le coefficient alpha de Cronbach est de 0,93.

Score relatif à l'expérience scolaire/à la maturité scolaire :

Ce facteur a été calculé à l'aide des items suivants : DETCQ26A, DETCQ26B, DETCQ26C, DETCQ26D et DETCQ26E. Aucune imputation n'a été effectuée pour ce score. Le score varie entre 0 et 20. Un score élevé indique un faible niveau de maturité scolaire. Le coefficient alpha de Cronbach est de 0,83.

Score relatif aux troubles de la conduite et à l'agression physique

Ce facteur a été calculé à l'aide des items suivants : DETCQ27G, DETCQ27X, DETCQ27AA, DETCQ27FF, DETCQ27JJ et DETCQ27NN. Aucune imputation n'a été effectuée pour ce score. Le score total varie entre 0 et 12. Un score élevé indique la présence de troubles de la conduite et d'agression physique. Le coefficient alpha de Cronbach est de 0,887.

Score relatif à l'agression indirecte

Ce facteur a été calculé à l'aide des items suivants : DETCQ27J, DETCQ27R, DETCQ27Z, DETCQ27LL et DETCQ27SS. Aucune imputation n'a été effectuée. Le score varie entre 0 et 10. Un score élevé indique la présence d'agression indirecte. Le coefficient alpha de Cronbach est de 0,886.

Score relatif au caractère participatif du milieu scolaire

Ce facteur a été calculé à l'aide des items suivants : DETCQ50A, DETCQ50B, DETCQ50C, DETCQ50D, DETCQ50E, DETCQ50F et DETCQ50G. Aucune imputation n'a été effectuée pour ce score. Le score varie entre 0 et 28. Un score élevé indique un milieu scolaire où il y a une forte participation des enseignants. Le coefficient alpha de Cronbach est de 0,88.

Score relatif au degré de soutien du milieu scolaire

Ce facteur a été calculé à l'aide des items suivants : DETCQ50H, DETCQ50I, DETCQ50J, DETCQ50K et DETCQ50L. Aucune imputation n'a été effectuée pour ce score. Le score varie entre 0 et 20. Un score élevé indique un haut degré de soutien du milieu scolaire. Le coefficient alpha de Cronbach est de 0,91.

Score relatif au climat disciplinaire

Ce facteur a été calculé à l'aide des items suivants : DETCQ51A, DETCQ51B, DETCQ51C et DETCQ51D. Aucune imputation n'a été effectuée pour ce score. Le score varie entre 0 et 16. Un score élevé indique la présence d'un climat scolaire hautement disciplinaire. Le coefficient alpha de Cronbach est de 0,81.

Échelle relative au directeur

Il n'y avait qu'une seule échelle comprise dans le questionnaire du directeur dans le cadre de l'ELNEJ. C'était le **score relatif au climat disciplinaire**. Ce facteur a été calculé à l'aide des items suivants : DEPCQ15A, DEPCQ15B, DEPCQ15C, DEPCQ15D, DEPCQ15E, DEPCQ15F, DEPCQ15G, DEPCQ15H, DEPCQ15I, DEPCQ15J, DEPCQ15K et DEPCQ15L. Aucune imputation n'a été effectuée pour ce score. Les scores varient entre 0 et 46. Un score élevé indique la présence de problèmes disciplinaires dans l'école. On n'a noté aucun coefficient alpha de Cronbach.

10.0 Méthodologie d'enquête – Taux de réponse et couverture de la composante Éducation

Dans la composante Éducation, on considère 2 types de non-réponse : la non-réponse totale où l'unité sélectionnée (l'enseignant et le directeur d'école) ne répond à aucune question et la non-réponse partielle où l'unité sélectionnée (l'enseignant ou le directeur d'école) répond à suffisamment de questions pour être considérée répondant, sans toutefois répondre à toutes les questions.

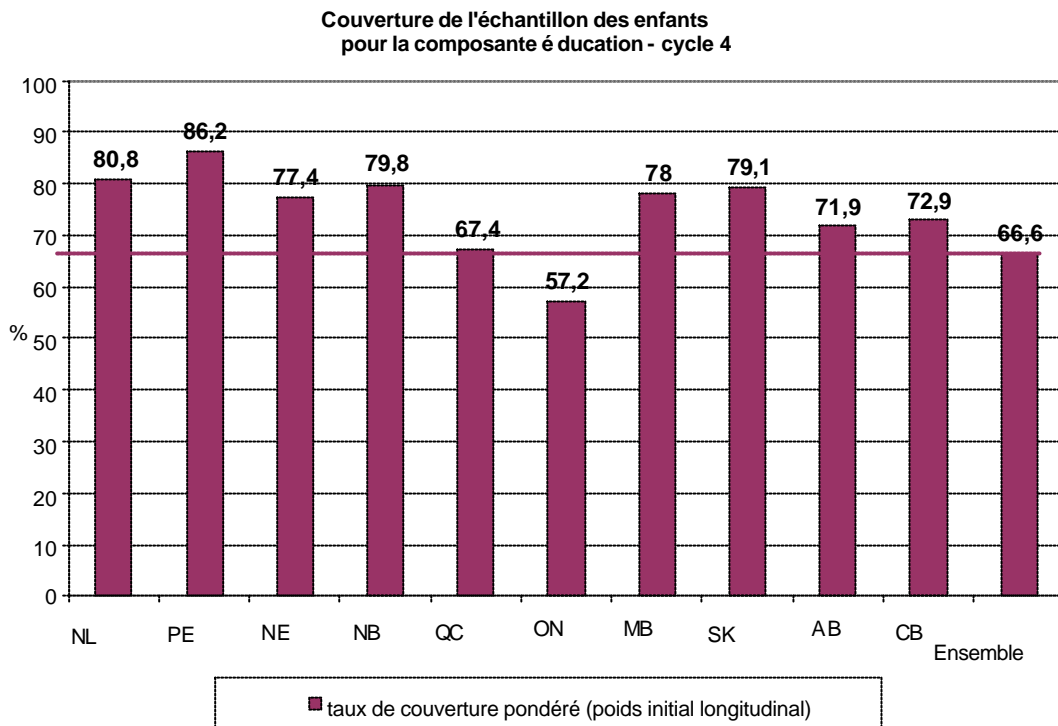
En plus de la non-réponse nous faisons place à un problème de sous-couverture causé par des difficultés opérationnelles lors de la collecte.

Ce chapitre traite de la non-réponse totale. Plus de 10 000 enfants sont au départ éligibles à l'enquête. Des problèmes opérationnels ont réduit l'échantillon et abouti à une sous couverture pour la composante Éducation de l'ELNEJ au cycle 4. La non-réponse est alors à plusieurs niveaux : les parents consentent ou non à ce que nous contactions les commissions scolaires, les commissions scolaires elles-mêmes consentent ou non à ce que nous contactions les enseignants de l'enfant et le directeur de son école. Et enfin les enseignants et/ou directeurs d'école peuvent ne pas répondre au questionnaire relatif à l'enfant ou à l'école de l'enfant.

Les taux de réponse dans une enquête sont utilisés pour mesurer l'efficacité du processus de collecte et sont également un bon indicateur de la qualité des estimations produites. La non-réponse partielle est traitée dans le chapitre suivant sur l'imputation.

10.1 Sous-couverture de l'échantillon

Seule une partie (70 %) des enfants éligibles (habitant une des 10 provinces, fréquentant une école publique, nés entre 1985 et 1994) a été échantillonnée du fait de problèmes opérationnels. Cette couverture est toutefois différente en fonction de la province de résidence ou de la taille du secteur de résidence ou encore de l'âge de l'enfant. Il n'y a pas de différence de couverture en fonction du sexe de l'enfant.



Taux de couverture par taille du secteur de résidence

Taille du secteur de résidence	taux
région rurale	68,8
région urb < 30 000	68,1
région urb 30 000 à 99 999	66,4
région urb 100 000 à 499 999	68,1
région urb + de 500 000	63,7
Non déclaré	74,7
<i>Ensemble</i>	66,3

Taux de couverture par âge

Age de l'enfant	Taux
6 ans	59,4
7 ans	53,8
8 ans	58,6
9 ans	66,3
10 ans	67,8
11 ans	68,5
12 ans	70,2
13 ans	72,7
14 ans	72,3
15 ans	74,6
<i>Ensemble</i>	66,3

10.2 Non-réponse totale

10.2.1 Définitions

Certaines définitions sont essentielles à la compréhension des différents tableaux.

Un **ménage répondant** est un ménage où une composante adulte ou une composante enfant ou jeune a été complétée.

Un **enfant répondant** est un enfant pour lequel une composante adulte ou sa composante enfant ou jeune a été complétée. Un ménage répondant sans composante adulte complète pourrait avoir 1 enfant répondant et 1 enfant non-répondant. Dix-neuf enfants se retrouvent dans cette situation.

Un **ménage non-éligible** est un ménage où tous les enfants sont décédés ou habitent maintenant à l'extérieur d'une des 10 provinces canadiennes.

Le **taux de réponse** est défini comme le nombre de répondants sur le nombre d'enfants éligibles.

Les **refus** ont été identifiés lors du traitement des fichiers de réponse. Les notes des intervieweurs ont été utilisées et une recherche par mots-clés effectuée pour compléter la codification.

Le **taux de refus** est défini comme le taux nombre d'enfants ayant refusé sur le nombre d'enfants éligibles.

Un ménage est considéré **à dépister** lorsqu'un des appels effectués pour ce cas a résulté en un code de réponse de dépistage.

Le **taux à dépister** est défini comme le nombre de cas à dépister sur le total de l'échantillon (éligible ou non).

Un ménage est considéré **non dépisté** lorsqu'il est non-répondant et que le code final de réponse est un code de dépistage.

Le **taux dépisté avec succès** est défini comme le nombre d'enfants dépistés sur le nombre d'enfants à dépister.

Le **taux de consentement parental** est le taux de non-réponse ayant trait à l'autorisation accordée par la PMR (ou le conjoint) d'administrer le questionnaire des enseignants et/ou du directeur de l'école.

Le **taux de consentement du conseil scolaire** est le taux de non-réponse ayant trait à l'autorisation accordée par le conseil scolaire d'administrer le questionnaire des enseignants et/ou du directeur de l'école.

La **couverture** est la mesure selon laquelle l'échantillon peut représenter adéquatement tous les éléments de la population cible.

10.3 Les consentements

Pour chaque enfant échantillonné, le consentement est demandé aux parents puis aux commissions scolaires pour administrer un questionnaire aux enseignants et aux directeurs d'école.

Le taux de consentement des parents est très élevé : 94,8 % au niveau national. Prises une par une, aucune variable ne présente des taux de consentement très différents par modalité du taux national.

Pour les enfants dont les parents ont donné leur consentement à contacter les écoles, 96 % d'entre eux sont dans des établissements scolaires pour lesquels la commission scolaire a donné son accord. L'accord des commissions scolaires est construit à partir des envois de questionnaires aux professeurs ou aux directeurs. Les taux de réponse des commissions scolaires sont légèrement différents d'une province à l'autre et par type d'école.

Province	Type d'école 1 : non-confessionnelle 2 : confessionnelle	Taux de réponse pondéré
NL	1	98,7 %
PE	1	100%
NE	1	91,4 %
NE	2	100%
NB	1	98 %
QC	1	97,6 %
QC	2	89,1 %
ON	1	93,8 %
ON	2	94,6 %
MB	1	95,5 %
MB	2	100%
SK	1	97,3 %
SK	2	92,9 %
AB	1	94,6 %
AB	2	96,7 %
CB	1	92,6 %
CB	2	86,4 %
		95,9 %

10.4 La non réponse des enseignants et des directeurs

L'unité d'observation est **répondante** si l'enseignant **OU** le directeur d'école a répondu au questionnaire.

Voici un résumé des réponses des enseignants et des directeurs d'école :

	Réponses	Enseignants	Directeurs
6 582 enfants dont les enseignants ou le directeur ont reçu un questionnaire	5 196 réponses	2 940 enseignants uniques	2 605 réponses
			335 non réponses
		1 497 enseignants multiples	1 359 réponses
		Pas de réponse enseignant	138 non réponses
	1 386 non réponses		759 réponses

Le taux de réponse varie d'une province à l'autre, en fonction de la taille du secteur de résidence et selon le type d'école (80 % de réponse pour les écoles non confessionnelles et 68 % pour les écoles confessionnelles).

Province	Taux de réponse des enseignants ou directeurs
NL	88,8 %
PE	92 %
NE	84,3 %
NB	87,1 %
QC	78,9 %
ON	78 %
MB	77,1 %
SK	76,6 %
AB	70 %
CB	70,3 %
Canada	79 %

Par secteur de résidence	Taux de réponse des enseignants ou directeurs
région rurale	82 %
région urb < 30000 hab.	82,1 %
région urb 30000 à 99999	72,5 %
région urb 100000 à 499999	80,3 %
région urb + de 500000	72,4 %
Non déclaré	81,6 %
Canada	79 %

Dans les cas où l'enseignant n'a pas répondu mais que le directeur a répondu, ou vice versa, l'enregistrement relatif à l'enfant est considéré comme une réponse. Les données de la partie non répondante sont alors imputées afin que l'enregistrement ait des données complètes (voir chapitre 11).

11.0 Imputation

Comme on l'a précisé au chapitre précédent, un répondant est défini comme étant un enfant ou un jeune pour lequel au moins l'enseignant ou le directeur a répondu à l'enquête. L'imputation traite la non-réponse totale si une de ses deux parties est manquante. La non réponse partielle dans l'ensemble des données relatives à l'enfant n'est pas traitée par l'imputation. L'imputation est la procédure qui consiste à remplacer les valeurs manquantes par des valeurs plausibles.

Pour les détails complets de l'imputation, voir le guide de l'utilisateur du fichier principal.

11.1 Imputation des réponses des directeurs d'école

Sur les 5 196 questionnaires considérés comme des répondants pour la partie éducation, 4 437 (soit 85 %) ont des informations relatives aux enseignants. Sur ces 4 437, 473 questionnaires (soit 11 %) ne contiennent pas d'informations du directeur de l'école. Ce sont ces 473 questionnaires qui sont sujets à l'imputation par donneur.

Pour associer des donneurs aux receveurs, on essaie de voir si les réponses des enseignants pour lesquelles une réponse du directeur est aussi disponible sont semblables aux réponses des enseignants sans données du directeur.

Les classes d'imputation seront au minimum créées en fonction de la province, le type d'école et la taille du secteur de résidence. Avec cette stratégie, il a peu de « cases » pour lesquelles le nombre de réponses de directeur à imputer est très élevé. Certains questionnaires ont un secteur de résidence non identifié. Dans ces cas là, on utilise le secteur de résidence le plus fréquent dans la province et le type d'école.

Dans le processus d'imputation, les classes minimales seront formées du type d'école, de la province et de la taille du secteur de résidence. D'autres variables sont également utilisées : elles portent sur la perception des enseignants quant à l'organisation de l'école, le soutien, l'esprit d'équipe... mais aussi sur l'attention des enfants en classe. Ces variables sont différentes si le questionnaire simple ou multiple est posé à l'enseignant. De plus, on construit un indicateur de niveau de scolarité entre le primaire et le secondaire basé sur l'âge de l'enfant. Cet indicateur vaut un si l'enfant a plus de 11 ans, 0 sinon. Cette variable est utilisée dans le processus d'imputation.

Le processus d'imputation utilise la macro Impudon. A la première étape on demande d'attribuer aux receveurs un donneur qui ait les mêmes caractéristiques pour les variables : type d'école, province, taille du secteur de résidence et enfant de plus de 11 ans ou non.

Dans la procédure on précise que l'on souhaite une similarité des variables de perception des enseignants entre les receveurs et les donneurs. Toutes les variables de perception ont un poids identique pour le calcul de la distance, à savoir qu'on ne privilégie pas une variable plutôt qu'une autre dans le calcul de la distance entre deux observations.

Pour chaque variable de similitude, on calcule un écart relatif entre la valeur du receveur et celle du donneur. Ensuite on construit une distance globale pour l'ensemble des variables de perception qui est la somme pondérée des **écarts relatifs**, donc une forme de distance totale basée sur les **champs de similitude**. Le donneur le plus proche du receveur au sens de cette distance sera choisi.

On demande par ailleurs qu'il y ait concordance sur l'âge de l'enfant entre le donneur et le receveur.

Cette première étape amène à trouver 425 donneurs pour 470 receveurs. Sur les 425 donneurs, 385 ont été utilisés une seule fois, 35 deux fois, et 5 trois fois.

On effectue une deuxième passe d'imputation pour trouver un donneur pour les 3 questionnaires restants à imputer. Pour cette passe, on ne demande plus de concordance au niveau de l'âge de l'enfant. Un seul et même donneur est utilisé pour ces trois enregistrements.

11.2 Imputation des réponses des enseignants

Parmi les 5 196 réponses au questionnaire de l'enfant, 3 964 questionnaires (2 605 + 1 359) ont des réponses de l'enseignant et du directeur et 759 questionnaires n'ont pas de réponse des enseignants mais une réponse des directeurs. L'idée est donc d'imputer la réponse des enseignants pour ces 759 questionnaires.

Les classes d'imputation sont construites par croisement de la **province**, du **type d'école**, du **secteur de résidence** et du **groupe d'âge (plus de 11 ans ou non)**. Là encore, le processus d'imputation utilise la macro Impudon.

Multiplicité des donneurs	Nombre de donneurs
1	574
2	67
3	9
4	5

Après le premier passage de la macro, on a pu associer un donneur à 755 des enseignants non répondants sur les 759. Reste à trouver un donneur pour 4 receveurs.

La seconde passe est faite en élargissant la contrainte liée à la taille du secteur de résidence : la concordance exacte entre la taille du secteur de résidence du donneur et du receveur n'est plus exigée. Une proximité entre la valeur de cette variable pour le donneur et pour le receveur suffit à associer un donneur au receveur. On trouve alors un donneur pour imputer 3 receveurs. Pour le dernier à imputer, celui-ci est le seul d'une école confessionnelle. On l'impute par un voisin d'une école de type non confessionnelle.

12.0 Pondération, traitement de la non-réponse et limites

La phase de pondération vise à produire des poids individuels pour chaque enfant répondant. L'échantillon des enfants répondants permet alors de représenter la population des enfants vivant dans une des 10 provinces canadiennes, âgés de 6 à 15 ans, fréquentant une école publique. Chaque enfant se voit ainsi attribuer une pondération finale qui tient compte de toutes les étapes du processus.

Les pondérations produites sont élaborées à partir du poids transversal de l'enfant au cycle 4 de l'ELNEJ. Seul **un poids transversal final est produit**.

12.1 Ajustement de la sous couverture

Pour ajuster la sous couverture, est calculée pour chaque enfant éligible une probabilité d'être sous échantillonné. Cette probabilité est le résultat d'une modélisation logistique en fonction de la province de résidence, la taille du secteur de résidence et l'âge de l'enfant.

Le modèle est donné par :

$p = \text{Probabilité (unité est sous échantillonnée)}$

and $\text{logit}(p) = 0,7979 + 0,6163 * (\text{Résider à NL}) + \dots + 0,3270 * (\text{Enfant de 15 ans})$.

Paramètre	Estimation	Erreur	Chi-Square	Pr > ChiSq
Constante	0,7979	0,0505	249,4939	<,0001
<i>Référence = Québec</i>				
NL	0,6163	0,1858	11,0068	0,0009
PE	1,0762	0,3847	7,8259	0,0052
NE	0,3972	0,1403	8,0075	0,0047
NB	0,6405	0,1619	15,6586	<,0001
ON	-0,5118	0,0556	84,7562	<,0001
MB	0,6752	0,1274	28,0678	<,0001
SK	0,5093	0,1293	15,5189	<,0001
AB	0,1876	0,0795	5,5697	0,0183
CB	0,2067	0,0767	7,2711	0,0070
<i>Référence = région urbaine de 30 000 à 99 000 hab</i>				
Région urbaine de 100 000 à 499 000 hab	0,2180	0,0557	15,3115	<,0001
Région urbaine non déclarée	0,4307	0,1992	4,6728	0,0306
<i>Référence : enfants entre 9 et 12 ans</i>				
AGE6	-0,3617	0,0737	24,1122	<,0001
AGE7	-0,6191	0,0695	79,4446	<,0001
AGE8	-0,4073	0,0696	34,2719	<,0001
AGE13	0,2310	0,0781	8,7442	0,0031
AGE14	0,2207	0,0765	8,3122	0,0039
AGE15	0,3270	0,0793	17,0115	<,0001

Le modèle montre un phénomène intéressant quant à la sous-couverture en fonction de l'âge de l'enfant. Plus l'âge de l'enfant est grand, plus la probabilité de l'unité à être échantillonnée est élevée.

Au niveau des provinces, on voit que l'Ontario a une sous couverture confirmée.

Le modèle associe une probabilité de sous échantillonnage à chaque enfant éligible. On établit alors 5 classes de regroupement. Une classe de regroupement est constituée des enfants pour lesquels la probabilité estimée est comprise entre le i ème quantile et le $i+1$ ème quantile. On prend ici les quantiles d'ordre 20 %, 40 %, 60 % et 80 %.

Sur le modèle estimé ci-dessus, les quantiles pour la probabilité estimée d'être sous échantillonnée sont : 0,595(20 %), 0,682(40 %), 0,734(60 %) et 0,8 (80 %). Étant donné que la probabilité de ne pas être sous échantillonnée est en moyenne plus faible que celle d'être sous échantillonnée, le groupe (p entre 0 à 20 %) contiendra proportionnellement plus d'enfants sous échantillonnés.

Nombre d'enfants Sous échantillonnés	P estimée entre				0-20 %	Ensemble
	80-100 %	60-80 %	40-60 %	20-40 %		
Non	370	456	550	753	1 009	3 138
Oui	1 845	1 591	1 416	1 250	1 136	7 238
Ensemble	2 215	2 047	1 966	2 003	2 145	10 376
Facteur d'ajustementmoyen	1,21	1,31	1,39	1,6	1,89	1,5

Pour chaque enfant sous échantillonné, sa pondération est ajustée au sein du groupe. Les facteurs d'ajustement sont en moyenne de 1,2 à 1,9.

12.2 Ajustement du consentement des parents

L'ensemble des données est réduit aux enfants sous-échantillonnés. Les données sont traitées avec Knowledge Seeker en vue d'élaborer des groupes de réponse dits homogènes. Les variables utilisées pour la classification sont :

- Le comportement de l'adulte face à la cigarette;
- Le secteur de résidence;
- La religion;
- Le statut de l'occupation du logement;
- Le type de logement;
- La fratrie;
- L'implication dans des activités sportives de l'adulte avec l'enfant;
- La santé de l'enfant;
- Le niveau de revenu suffisant;
- La province;
- Le sexe;
- L'activité de l'enfant avec ses amis;
- Né au Canada ou non;
- Adulte actif ou non;
- Dépression de l'adulte.

La variable d'intérêt est le consentement des parents.

Quarante-quatre (44) groupes de réponse homogènes sont créés. On impose un minimum de 30 observations par groupe avec un taux de réponse d'au moins 60 % en général. Nous avons accepté parfois des taux de réponse plus bas.

Le taux pondéré le plus élevé de non consentement est de 52,6 %. Ce groupe comprend 39 enfants de l'échantillon : 28 pour lesquels les parents ont donné leur consentement et 11 pour lesquels les parents ont refusé leur consentement. Ce sont des filles avec un frère ou une sœur qui habitent dans un secteur urbain de 500 000 hab. ou plus et dont l'adulte de référence ne fume pas, est de religion presbytérienne, baptiste ou témoins de Jéhovah.

Le facteur d'ajustement varie donc de 1 à 2,11.

12.3 Ajustement du consentement des commissions scolaires

Sur les 6865 enfants pour lesquels les parents ont donné leur consentement, 6582 enfants sont dans des établissements scolaires pour lesquels la commission scolaire a donné son accord.

L'accord des commissions scolaires est construit à partir des envois de questionnaires aux professeurs ou aux directeurs.

L'ajustement pour non réponse est fait ici par province et type d'école uniquement.

Les facteurs d'ajustement sont assez faibles vu le taux de réponse élevé.

Province	Type d'école 1 : non- confessionnelle 2 : confessionnelle	Nb de questionnaires pour lesquels les parents ont donné leur consentement	Taux de réponse pondéré	Facteur d'ajustement
NL	1	445	98,7 %	1,013
PE	1	263	100 %	1,000
NE	1	536	91,4 %	1,094
NE	2	1	100 %	1,000
NB	1	483	98 %	1,0204
QC	1	1 148	97,6 %	1,0242
QC	2	68	89,1 %	1,1198
ON	1	951	93,8 %	1,0640
ON	2	475	94,6 %	1,0563
MB	1	594	95,5 %	1,0467
MB	2	1	100 %	1,000
SK	1	483	97,3 %	1,027
SK	2	132	92,9 %	1,080
AB	1	507	94,6 %	1,057
AB	2	171	96,7 %	1,032
CB	1	581	92,6 %	1,0811
CB	2	25	86,4 %	1,146
		6 865	95,9 %	

12.4 Ajustement pour non-réponse des enseignants ou directeurs

Sur ces 6 582 questionnaires envoyés, 6 278 correspondent à un questionnaire envoyé à l'enseignant et 5 038 à un questionnaire envoyé au directeur d'école.

Etant donné la définition d'une unité répondante (la réponse de l'enseignant ou du directeur est enregistrée), 5 196 enfants sont qualifiés de répondants. Ceci correspond à un taux de réponse pondéré de 79 %.

L'ajustement de non réponse est basé sur une discrimination des unités répondantes versus les unités non répondantes. L'analyse est faite à partir des variables observées communément sur les deux sous populations des répondants et des non répondants.

Les variables discriminantes sont : la province, la taille du secteur de résidence, l'âge de l'enfant, le soutien social, la situation des parents (biologiques ou non), la fratrie de l'enfant, le statut d'occupation, la santé de l'enfant, celle de la personne de référence, l'adéquation revenus/besoins de la famille, le statut de travailleur de la dernière année.

Le premier niveau de discrimination de la fonction de réponse est la province. Les provinces maritimes sont regroupées ensemble. Manitoba et Saskatchewan sont ensemble, ainsi que l'Alberta et la Colombie britannique.

Pour les maritimes, le taux de réponse est plus faible si l'enfant a plus de 11 ans (questionnaire des enseignants multiples) et plus particulièrement si celui-ci a plus de 11 ans et si le score de dépression est élevé;

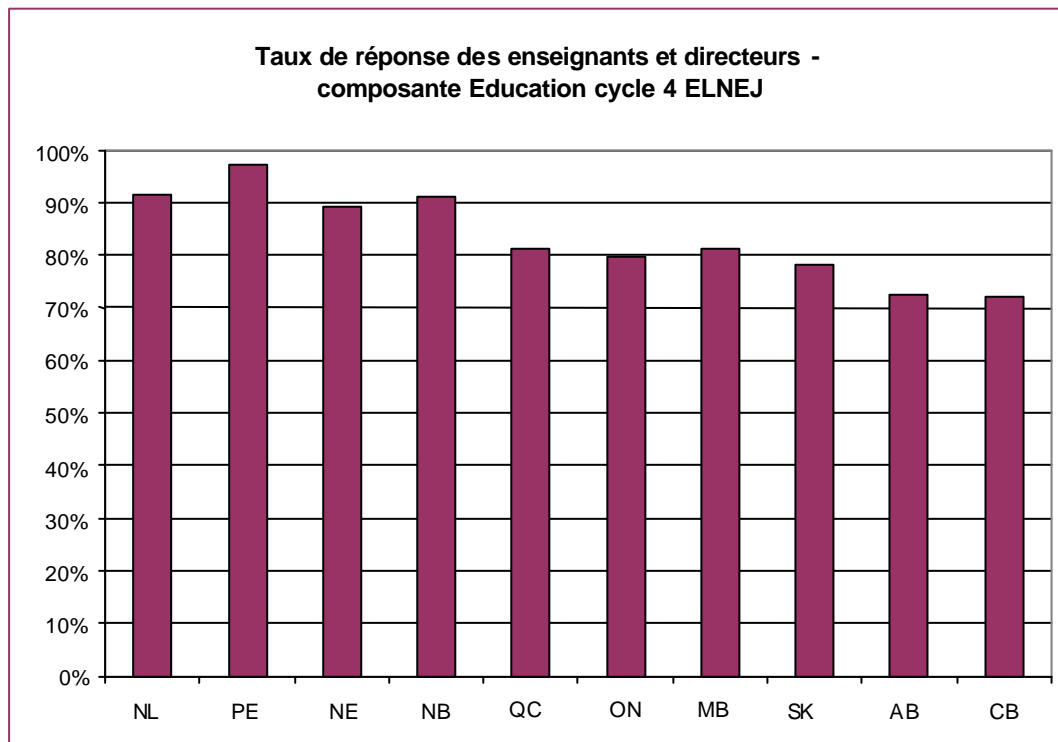
Au Québec, les variables de discrimination sont le fait d'aller à la bibliothèque publique, la situation des parents (biologique...), la fratrie, la taille du secteur de résidence, le comportement émotif de l'enfant; le taux de réponse le plus faible est pour les enfants qui vont moins d'une ou deux fois par an à la bibliothèque publique et n'ayant qu'un parent biologique sur les deux;

En Ontario, la taille du secteur de résidence vient ensuite discriminer le comportement de réponse. Dans les très grosses villes, le comportement émotif des enfants et la fratrie expliquent les comportements de réponse. Dans les autres villes, la situation matrimoniale de la personne de référence, s'il est né au Canada ou non, le fait qu'il ne réponde pas aux questions sur la dépression, son statut de travailleur caractérisent la réponse des enseignants.

Au Manitoba et Saskatchewan, pour les plus de 11 ans, le comportement émotif de l'enfant est important. Pour les moins de 11 ans, la pratique de sports ou jeux avec l'adulte discrimine la réponse.

En Alberta et Colombie britannique, les variables discriminantes sont : la taille du secteur de résidence, la santé de l'adulte, son statut de travailleur, s'il est marié ou non, le comportement émotif de l'enfant.

L'ajustement moyen est de 1,24. Il est plus élevé dans les provinces de l'ouest canadien et moins élevé dans les provinces maritimes.



12.5 Résumé des ajustements

En résumé, les quatre ajustements (sous couverture, consentement des commissions scolaires, consentement des parents, réponse des professeurs et directeurs) corrigent en moyenne les poids initiaux des enfants d'un facteur 1,99.

Cet ajustement total varie d'une province à l'autre et aussi d'un ajustement à l'autre : l'ajustement concernant la sous couverture et l'ajustement relatif à la non réponse des professeurs ou directeurs sont les plus importants.

Population	Taille de l'échantillon	Moyenne Ajustement			
		1 : sous couverture	2 : consentement commissions scolaires	3 : consentements parents	4 : réponse enseignants, directeurs
Ensemble	5 196	1,44	1,06	1,05	1,24
Ecole non confessionnelle	4 636	1,41	1,06	1,04	1,24
Ecole confessionnelle	560	1,63	1,06	1,06	1,30
NL	388	1,26	1,05	1,01	1,11
PE	242	1,22	1,06	1,00	1,11
NB	420	1,33	1,05	1,10	1,11
NE	413	1,25	1,06	1,02	1,11
QC	944	1,48	1,05	1,03	1,24
ON	1 054	1,76	1,05	1,06	1,28
MB	444	1,26	1,07	1,05	1,25
SK	449	1,30	1,06	1,04	1,25
AB	442	1,47	1,08	1,05	1,37
CB	400	1,44	1,08	1,09	1,47

12.6 Limitations

Tel que mentionné précédemment, le but principal de la repondération du fichier de l'éducation était de donner un certain répit aux analystes qui devaient faire leur analyse avec l'échantillon réduit. Ce sous-échantillonnage ne touchait que les données recueillies au cycle 4, et le but était donc d'améliorer la possibilité d'analyse au niveau transversal seulement. Grâce à des consultations, nous avons déterminé un certain nombre de variables de la composante principale qui seraient probablement utilisées dans la plus grande part de l'analyse. Certaines des variables ont été incluses dans un ajustement proportionnel itératif des poids afin de conserver un certain degré de cohérences entre les estimations ponctuelles établies à partir du sous-échantillon et de l'échantillon complet.

Avec l'inclusion de certaines variables de la composante principale, on espère que les utilisateurs seront en mesure de faire la plupart des analyses transversales sans se lier au fichier principal. Si un utilisateur a besoin du fichier principal, on doit tenir compte que les estimations peuvent être différentes puisqu'elles ne sont pas utilisées dans le calage, certaines différences dues au recouvrement du sous-ensemble ne seront pas reflétées dans le poids ajusté de la composante école. Pareillement, une analyse longitudinale va nécessiter un appariement au fichier principal afin d'obtenir les poids longitudinaux, l'information provenant de la composante école va apparier seulement à un sous-ensemble du fichier complet, aucun ajustement est fait pour compenser l'information manquante.

Les lecteurs doivent noter, quoique tous les efforts pour l'ajustement proportionnel des caractéristiques du sous-ensemble à l'échantillon complet ont été faits, on ne puisse tenir compte tout les biais qui peuvent existés.

13.0 Qualité des données et recouvrement de l'enquête principal

Les renseignements au sujet de la recouvrement de la composante sur l'éducation se trouvent au chapitre 10 du présent guide de l'utilisateur. Pour d'autres détails, voir le guide d'utilisateur du fichier principal.

14.0 Lignes directrices concernant la totalisation, l'analyse et la diffusion de données

Cette partie du document décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou diffusent des données tirées du fichier de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées d'arriver aux mêmes chiffres que Statistique Canada, tout en étant en mesure d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

14.1 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées du fichier de microdonnées de la composante Éducation de l'ELNEJ correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices suivantes en ce qui concerne l'arrondissement de ces estimations.

- a) Les estimations d'un tableau statistique qui ne sont pas des totaux partiels marginaux ou des totaux marginaux doivent être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Dans cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de un la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter de 1 le chiffre des centaines.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être arrondis à la centaine près à l'aide de la technique d'arrondissement classique.
- c) Les moyennes, proportions, taux et pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et/ou les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique décrite ci-dessus.
- e) Si, en raison de contraintes d'ordre technique ou de toute autre nature, une autre technique d'arrondissement que la technique classique est utilisée et que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.
- f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées par les utilisateurs. Les estimations non arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne le sont en réalité.

14.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Plan d'échantillonnage

Le plan d'échantillonnage utilisé pour la composante Education de l'ELNEJ n'est pas autopondéré. Aussi, pour la production d'estimations simples, y compris de tableaux statistiques ordinaires, l'utilisateur doit appliquer le poids de sondage approprié. Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir du fichier de microdonnées ne pourront alors être considérées comme représentatives de la population visée par l'enquête et elles ne correspondront pas à celles produites par Statistique Canada. De fait, le poids attribué à chaque enfant reflète le nombre d'enfants représentés par un répondant donné.

Avantage d'utiliser un poids ajusté ou ré-échelonné

Pour toute analyse portant sur une corrélation ou une autre statistique où une mesure de signification est requise, il est recommandé d'utiliser des poids « analytiques ». Il s'agit des poids initiaux ré-échelonnés en vue de produire des chiffres de l'échantillon plutôt que des chiffres de la population. Le poids analytique est obtenu en multipliant le poids démographique par la taille de l'échantillon et en divisant le produit par la population totale estimée. Le résultat est un poids moyen de 1 et une somme de poids égale à la taille de l'échantillon.

L'un des avantages d'utiliser le poids ajusté est d'éviter une surestimation de la signification pendant l'estimation. Bien que l'estimation produite à l'aide de cette technique sous-estime probablement la vraie mesure de l'erreur-type, elle fournit néanmoins un bon indicateur initial de la signification lorsqu'on fait une analyse exploratoire. Les utilisateurs auraient avantage à utiliser les méthodes recommandées (poids bootstrap) pour déterminer la signification de leurs résultats avant d'établir leur rapport final.

Mise en garde concernant l'utilisation du poids ajusté

L'inconvénient, en ce qui a trait au poids ajusté, est que le numérateur n'est pas pondéré jusqu'à la population cible. Les tables des coefficients de variance, décrites au chapitre 12 et présentées à l'annexe 3 du guide de l'utilisateur du fichier principal, ne sont plus utiles en tant que mesure de la qualité des données.

Les méthodologistes de l'enquête recommandent d'utiliser les poids bootstrap et le programme d'accompagnement pour calculer des estimations plus précises de la variance.

Il est à noter que les sous-domaines résultant de l'exclusion de la non-réponse partielle ne sont pas pris en compte dans ce poids ré-échelonné. Ce ré-échelonnage des poids ne redistribue pas le poids démographique des unités ayant des données manquantes, à moins qu'on ne suppose qu'il est réellement aléatoire.

Différences de progiciels

Les utilisateurs doivent également prendre note qu'en raison de la manière dont certains progiciels traitent la variable des poids, ceux-ci ne permettent peut-être pas d'obtenir des estimations correspondant exactement à celles de Statistique Canada.

14.3 Définitions des estimations de type nominal et des estimations quantitatives

Unité d'analyse

Il est à souligner que le fichier de la composante Éducation de l'ELNEJ a été établi de façon à ce que l'enfant constitue l'unité d'analyse. Le poids qui paraît sur chaque enregistrement (DEDUW01C) pour l'échantillon transversal) est un poids correspondant à un « enfant ». Il est donc impossible de calculer des estimations concernant les parents ou les familles à partir du fichier de microdonnées de la composante Education de l'ELNEJ.

14.3.1 Estimations de type nominal

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui appartiennent à une catégorie définie. On peut aussi utiliser l'expression « estimation d'un agrégat » pour parler d'une estimation du nombre de personnes qui possèdent une caractéristique donnée.

Exemples de questions de type nominal

Q : À votre avis, dans quelle mesure le ou les parents ou tuteurs s'intéressent-ils à l'éducation de cet(te) élève?

1. Très intéressés
2. Assez intéressés
3. Pas intéressés
4. Je ne sais pas / pas d'opinion

Q : Quelle influence ces conseils, associations ou comités exercent-ils sur les politiques ou pratiques de votre école?

1. Très grande influence
2. Grande influence
3. Un peu d'influence
4. Peu d'influence
5. Aucune influence

14.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux, de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur certains ou sur tous les membres de la population visée par l'enquête. De plus,

elles comprennent plus particulièrement des estimations de la forme $\frac{\hat{X}}{\hat{Y}}$ où

\hat{X} est une estimation de la quantité totale pour la population visée par l'enquête et \hat{Y} une estimation du nombre de personnes dans la population étudiée qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen d'élèves dans une classe. Le numérateur est une estimation du nombre total d'élèves et le dénominateur est le nombre de classes (obtenues par le nombre de professeurs répondants).

Exemples de questions quantitatives

Combien y a-t-il d'élèves dans la classe de cet(te) élève?

élèves

14.3.3 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre d'enfants possédant une caractéristique particulière à partir du fichier de microdonnées en additionnant les poids finaux de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. Ces estimations peuvent être transversales ou longitudinales.

On obtient les proportions et les rapports de la forme $\frac{\hat{X}}{\hat{Y}}$ de la façon suivante :

- (a) en additionnant les poids finaux des enregistrements possédant la caractéristique qui nous intéresse, pour le numérateur \hat{X} ;
- (b) en additionnant les poids finaux des enregistrements possédant la caractéristique qui nous intéresse, pour le dénominateur \hat{Y} ; puis
- (c) en divisant l'estimation du numérateur par celle du dénominateur.

14.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement et en additionnant cette quantité pour tous les enregistrements visés.

Par exemple, pour obtenir une estimation du nombre total d'élèves qui présentent un trouble de la parole, il faut : multiplier le nombre d'élèves de la classe ayant un trouble de la parole par le poids final; puis additionner ensuite cette valeur pour tous les enregistrements indiquant un au moins un élève avec un trouble de la parole dans la classe.

Dans la classe de cet(te) élève, combien d'élèves (incluant cet(te) élève) présentent l'un ou l'autre des problèmes chroniques suivants, que ce problème ait été diagnostiqué ou non par un professionnel? (Certains élèves peuvent faire partie de plus d'une catégorie.)

- a) Trouble de la parole, déficience auditive ou visuelle, mobilité réduite ou autre problème de santé qui nuit à son apprentissage?

élèves

On calcule le numérateur \hat{X} de la même façon qu'une estimation quantitative et le dénominateur \hat{Y} , de la même façon qu'une estimation de type nominal. Par exemple, pour estimer le nombre moyen d'élèves par classe ayant un trouble de la parole, il faut :

- a) estimer le nombre total d'élèves ayant un trouble de la parole de la façon décrite ci-dessus;

- b) estimer le nombre d'enfants de cette catégorie en additionnant les poids finaux de tous les enregistrements correspondant aux enfants dont au moins un élève de la classe a un trouble de la parole;
- c) Diviser l'estimation obtenue en a) par celle calculée en b).

14.4 Lignes directrices pour l'analyse statistique

Plan d'échantillonnage

L'ELNEJ est basée sur un plan d'échantillonnage complexe comportant une stratification et de multiples degrés de sélection ainsi que des probabilités inégales de sélection des répondants.

L'utilisation des données provenant d'enquêtes aussi complexes pose des problèmes aux analystes parce que le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance qui doivent être utilisées. Pour éviter tout biais des estimations et des analyses provenant de telles enquêtes, il faut utiliser des poids d'échantillonnage.

Variance estimée

Bien que de nombreuses méthodes d'analyse que l'on trouve dans les logiciels statistiques permettent d'utiliser des poids, la définition ou le sens de ces derniers diffère de ce qui devrait être utilisé dans une enquête par sondage, de sorte que si les estimations faites au moyen de ces logiciels sont exactes dans bien des cas, par contre les variances calculées sont peu fiables. Les variances d'estimations simples comme les totaux, les proportions et les rapports (pour les variables qualitatives) figurent dans les tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Transformer les poids

Dans le cas des autres techniques d'analyse (par exemple, la régression linéaire, la régression logistique et l'analyse de variance), il existe un moyen de rendre l'application des logiciels standard plus significative en incluant les probabilités inégales de sélection. La méthode transforme les poids de façon à ce que le poids moyen soit de 1.

Exemple de la transformation des poids

Par exemple, supposons qu'il faille procéder à une analyse portant sur tous les garçons. Les étapes de transformation des poids seraient les suivantes :

- sélectionner tous les enfants du fichier dont le SEXE (variable DMMCQ02) est masculin;
- calculer le poids MOYEN pour ces enfants en additionnant les poids finaux (variable DEDUW01C) de chaque enregistrement correspondant à un garçon, puis en divisant ce résultat par le nombre de garçons dans le fichier;
- pour chaque garçon, calculer le poids TRANSFORMÉ, qui correspond au poids final de l'enfant, divisé par le poids MOYEN;
- procéder à l'analyse portant sur les garçons en utilisant le poids TRANSFORMÉ.

Néanmoins, puisque l'on ne tient pas compte de la stratification et de la structure en grappes du plan d'échantillonnage avec cette méthode, les estimations de variance calculées représenteront probablement des sous-estimations des variances réelles.

Calcul d'estimations de variance

Le calcul d'estimations de variance vraiment significatives exige une connaissance détaillée du plan d'enquête. Ce niveau de détail ne peut figurer dans le fichier de microdonnées pour des raisons de confidentialité. Contre remboursement des frais, Statistique Canada peut calculer des variances qui tiennent compte de l'ensemble du plan d'échantillonnage de nombreuses statistiques. Les utilisateurs qui ont accès aux microdonnées peuvent produire de meilleures estimations de la variance pour certaines statistiques en utilisant les poids bootstrap (voir la partie 15.9).

14.5 Lignes directrices pour la diffusion en fonction du coefficient de variation

Lignes directrices pour la diffusion

Avant de diffuser ou de publier toute estimation tirée de l'ELNEJ, les utilisateurs doivent d'abord déterminer le niveau de qualité de l'estimation, à savoir acceptable, médiocre et inacceptable. Comme il en a été question au chapitre 10 dans le guide de l'utilisateur du fichier principal, des erreurs d'échantillonnage et des erreurs non dues à l'échantillonnage influent sur la qualité des données. Cependant, aux fins du présent document, le niveau de qualité d'une estimation est déterminé seulement en fonction de l'erreur d'échantillonnage illustrée par le coefficient de variation, comme il est indiqué au tableau ci-dessous.

D'abord, il faut déterminer le nombre d'enfants qui contribuent au calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée doit être considérée comme étant de qualité inacceptable.

En ce qui concerne les estimations pondérées fondées sur des échantillons composés de 30 enfants ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes directrices ci-dessous. Celles-ci devraient être appliquées aux estimations pondérées arrondies.

Toutes les estimations peuvent faire l'objet d'une diffusion. Toutefois, celles dont la qualité est médiocre ou inacceptable doivent être accompagnées d'une mise en garde à l'intention des utilisateurs subséquents.

14.6 Lignes directrices relatives au niveau de qualité

Niveau de qualité de l'estimation	Lignes directrices
1. Acceptable	Caractéristiques des estimations : taille d'échantillon de 30 ou plus et faibles coefficients de variation, compris entre 0,0 % et 16,5 %. Aucune mise en garde n'est requise.
2. Médiocre	Caractéristiques des estimations : taille d'échantillon de 30 ou plus et coefficients de variation élevés, compris entre 16,6 % et 33,3 %. On doit désigner ces estimations en utilisant la lettre M (ou un autre identificateur semblable). Elles doivent faire l'objet d'une mise en garde à l'intention des utilisateurs subséquents au sujet des niveaux d'erreur élevés associés aux estimations.
3. Inacceptable	Caractéristique des estimations : taille d'échantillon inférieure à 30, ou coefficients de variation très élevés, supérieurs à 33,3 %. Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Cependant, si l'utilisateur choisit de le faire, il doit alors désigner les estimations en utilisant la lettre I (ou un autre identificateur semblable) et les diffuser avec la mise en garde suivante : « Nous informons l'utilisateur que ... (préciser les 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 seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier également cette mise en garde. »

15.0 Calcul de la variance

Il a été mentionné précédemment que l'ELNEJ était une enquête probabiliste. Un échantillon a été sélectionné pour représenter la population cible. Une certaine variabilité est associée à la sélection de cet échantillon, qui serait différente si un autre échantillon avait été choisi. Cette variabilité est identifiée comme étant l'erreur échantillonnale au chapitre 10. À cette erreur, s'ajoutent des corrections pour tenir compte de la non-réponse. La mesure de cette variabilité est appelée la variance. Dans ce chapitre, nous expliquons l'importance de calculer la variance et surtout, nous présentons divers outils pour la calculer.

15.1 Importance du calcul de la variance

La variabilité ou variance d'une estimation est une bonne indication de la qualité d'une estimation. Une estimation avec une variance trop élevée est considérée comme non fiable. Afin de quantifier ce qui est une variance trop élevée, une mesure relative de la variabilité est utilisée, le coefficient de variation (c.v.). Le coefficient de variation est défini comme étant le ratio de la racine carrée de la variance sur l'estimation. La racine carrée de la variance est aussi connue sous le nom d'écart-type. L'utilisation du coefficient de variation plutôt que la variance directement nous permet de comparer sur une même échelle des estimations de magnitudes diverses. Ainsi, il est possible de quantifier la qualité de toute estimation avec le coefficient de variation.

De plus, pour différents tests statistiques tels que des hypothèses sur la différence entre 2 estimations, le calcul de la variance ou du coefficient de variation est requis afin de déterminer si la différence est statistiquement différente ou non. Le calcul de la variance ou du coefficient de variation est donc primordial.

15.2 Calcul de la variance et du coefficient de variation

La complexité du plan de sondage, les ajustements de non-réponse ne nous permettent pas de calculer une formule précise de variance. Une excellente approximation de la variance est l'utilisation de la méthode du bootstrap. Un ensemble de 1 000 poids bootstrap a été créé. Une approximation de la variance consiste à calculer l'estimation sur chacun de ces 1 000 ensembles de poids et de calculer la variance de ces 1 000 estimations.

Afin d'aider les utilisateurs à obtenir la variance et le coefficient de variation de leurs estimations, 2 outils, qui font tous usage des poids bootstrap, ont été développés. Ces outils sont :

- Des tables de variabilité d'échantillonnage approximative, basées sur un effet de plan représentatif de l'enquête, permettent d'obtenir des c.v. approximatifs pour des domaines d'âge ou de provinces.
- Des programmes pour le calcul de la variance en utilisant directement les poids bootstrap. Selon le type d'études et le niveau de précision requis, l'utilisation d'un ou plusieurs de ces outils est suggérée.

15.2.1 Tables de variabilité d'échantillonnage approximative

Afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une gamme étendue d'estimations de type nominal réalisées à partir de ce fichier de microdonnées, on a préparé un ensemble de « tables à consulter », appelées tables de variabilité d'échantillonnage approximative. Ces tables sont disponibles en format Word dans les Centres de données de recherche sous le nom de "TabVar_EDU_C4_F.doc". Elles 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.

15.2.2 Calcul des coefficients de variation

Les coefficients de variation (c.v.) dans ces tables sont obtenus au moyen de la formule de la variance utilisée pour l'échantillonnage aléatoire simple en y incorporant un facteur reflétant la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet du plan, on a commencé par calculer les effets du plan pour une gamme étendue de caractéristiques, puis on a choisi parmi les nombres ainsi obtenus une valeur modérée à employer dans les tables à consulter qui serait ensuite appliquée à l'ensemble des caractéristiques.

15.2.3 Allocation de l'échantillon

En ce qui concerne l'ELNEJ, l'allocation de l'échantillon prend en considération les points suivants :

- ✍ dans chaque province, la taille de l'échantillon du premier cycle devait autoriser la production d'estimations fiables pour tous les enfants longitudinaux de 0 à 11 ans (âge au cycle 1);
- ✍ il fallait également disposer d'un échantillon suffisamment important au premier cycle pour permettre la production d'estimations à l'échelle du Canada pour sept groupes d'âge ou cohortes : 0 à 11 mois, 1 an, 2 à 3 ans, 4 à 5 ans, 6 à 7 ans, 8 à 9 ans et 10 à 11 ans (âge au cycle 1);
- ✍ dans chaque province, la taille de l'échantillon du deuxième cycle devait autoriser la production d'estimations fiables pour tous les enfants de 0 à 11 ans (âge au cycle 1).

15.2.4 Effets du plan, des tailles de l'échantillon et des chiffres de population

Voici les tableaux des effets du plan, des tailles de l'échantillon et des chiffres de population par province tout d'abord, puis par groupe d'âge. Ces tableaux ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative. Les tableaux ne portent que sur l'échantillon **transversal** de la composante Éducation.

Province(s)	Effet de plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	2,1	388	67 135
Île-du-Prince-Édouard	1,8	242	19 082
Nouvelle-Écosse	1,8	420	112 090
Nouveau-Brunswick	2,2	413	90 501
Québec	2,6	944	789 907
Ontario	2,9	1 054	1 431 105
Manitoba	2,2	444	147 872
Saskatchewan	2,3	449	143 866
Alberta	2,1	442	382 874
Colombie-Britannique	2,2	400	440 982
Atlantiques	2,3	1 463	288 808
Autres	2,7	2 398	2 661 994
Prairies	2,9	1 335	674 612
Total CANADA	3,7	5 196	3 625 414

Age	Effet de plan	Taille de l'échantillon	Population
Enfants âgés de 6 à 11 ans	4,0	3 406	2 231 570
Enfants âgés de 12 à 15 ans	2,9	1 790	1 393 844
Total CANADA	3,7	5 196	3 625 414

Age	Effet de plan	Taille de l'échantillon	Population
Enfants âgés de 6 et 7 ans	2,7	1 322	698 748
Enfants âgés de 8 et 9 ans	3,5	1 048	766 069
Enfants âgés de 10 et 11 ans	3,5	1 036	766 753
Enfants âgés de 12 et 13 ans	2,9	915	688 396
Enfants âgés de 14 et 15 ans	2,9	875	705 447
Total CANADA	3,7	5 196	3 625 414

15.2.5 Nature approximative des tables de variabilité

Il convient de mentionner que tous les coefficients de variation sont approximatifs dans les tables de variabilité d'échantillonnage approximative et ne doivent donc pas être considérés comme des valeurs officielles. L'utilisation d'estimations de la variance réelle permettra probablement de produire des estimations dont les variances seront moins élevées. Par exemple, les estimations de la catégorie « inacceptable », selon les tables de variabilité d'échantillonnage approximative, pourraient se retrouver dans la catégorie « médiocre ».

N'oubliez pas - Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée doit être considérée comme étant « inacceptable », quelle que soit la valeur de son coefficient de variation. Cette règle s'applique parce que les formules utilisées pour estimer la variance ne tiennent pas dans le cas des échantillons de petite taille.

15.3 Comment utiliser les tables c.v. pour des estimations de type nominal

Les règles suivantes doivent permettre à l'utilisateur de calculer, à partir des tables de variabilité d'échantillonnage, les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population interviewée possédant une caractéristique donnée ainsi que des rapports et des écarts entre ces 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. Sur la table de variabilité d'échantillonnage correspondant à la région géographique ou au groupe d'âge approprié, repérez le nombre estimé dans la colonne à l'extrême gauche de la table (colonne intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) de gauche à droite jusqu'au premier nombre. Ce nombre constitue 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 ou d'un pourcentage estimé dépend à la fois de la grandeur de cette proportion ou de ce pourcentage et de la grandeur du total sur lequel est fondé cette proportion ou ce pourcentage. Les proportions ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est fondé sur un sous-ensemble de la population. Il faut noter que, dans les tables, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique du groupe d'âge auquel la table s'applique, le coefficient de variation de la proportion ou du pourcentage est égal à celui du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (par exemple, les personnes qui appartiennent à un sexe ou un groupe d'âge particulier d'une province), il faut se reporter à la proportion ou au pourcentage (en haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table, en bas). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriées.

Règle 3 : Estimations des différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est à peu près égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. L'erreur-type d'une différence $\hat{d} = (\hat{X}_1 - \hat{X}_2)$ est donc :

$$s_{\hat{d}} = \sqrt{(\hat{x}_1 a_1)^2 + (\hat{x}_2 a_2)^2}$$

où \hat{X}_1 représente l'estimation 1, \hat{X}_2 l'estimation 2, et a_1 et a_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 $s_{\hat{d}} / \hat{d}$.

Cette formule donne un résultat exact en ce qui a trait à la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

Règle 4 : Estimations de rapports

Si le numérateur n'est pas un sous-ensemble du dénominateur, l'écart-type du rapport entre les estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par R . L'erreur-type d'un rapport $\hat{R} = \hat{X}_1 / \hat{X}_2$ est donc :

$$s_{\hat{R}} = \hat{R} \sqrt{a_1^2 + a_2^2}$$

où a_1 et a_2 sont les coefficients de variation de \hat{X}_1 (le nombre de filles présentant des difficultés d'apprentissage) et \hat{X}_2 (le nombre de garçons présentant des difficultés d'apprentissage) respectivement. Le coefficient de variation de \hat{R} est donné par $s_{\hat{R}} / \hat{R}$. La formule tend à surestimer l'erreur si \hat{X}_1 and \hat{X}_2 sont corrélées positivement et elle tend à sousestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélées négativement.

Règle 5 : Estimations des différences entre des rapports

Dans ce cas, on combine les règles 3 et 4. On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le coefficient de variation de leur différence au moyen de la règle 3.

Avertissement au sujet des intervalles de confiance

Les lignes directrices sur la diffusion en ce qui concerne les estimations s'appliquent tout de même aux intervalles de confiance. Par exemple, si l'estimation est "minime", l'intervalle de confiance est aussi minime et devrait être accompagnée par un avertissement pour mettre en garde les utilisateurs subséquents des niveau élevés d'erreurs.

15.4 Exemples d'utilisation des tables de c.v. et du chiffrier pour des estimations nominales

Les exemples ci-dessous 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 que, d'après les données de la composante Éducation de l'ELNEJ, on estime que pour 92 342 enfants de 12 ans et plus ont redoublé une année. Comment l'utilisateur peut-il déterminer le c.v. de cette estimation?

1. Reportez -vous à la table de c.v. pour les enfants de 12 à 15 ans.
2. L'agrégat estimé (92 342) ne se trouve pas dans la colonne de gauche « Numérateur du pourcentage », il faut donc utiliser le nombre qui s'en rapproche le plus, c'est-à-dire 90 000.
3. Le coefficient de variation pour un agrégat estimé est la première inscription, autre que des astérisques, sur cette ligne, c'est-à-dire 14,4 %.
4. Le coefficient de variation approximatif du nombre d'enfants dont la personne référence a répondu qu'il avait un degré de réussite d'ensemble moyen est donc 14,4 %. Le résultat est « acceptable » et aucune mise en garde n'est nécessaire pour produire l'estimation, étant donné que le c.v. de l'estimation se situe dans la fourchette 0,0 % à 16,5 %.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons par exemple que d'après les données de la composante Éducation de l'ELNEJ, on estime que 32,2 % des 92 342 enfants de 12 ans et plus ayant redoublé une

année ont des troubles de la parole. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

1. Reportez -vous à la table de c.v. pour les enfants de 12 à 15 ans.
2. Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale, il faut utiliser à la fois le pourcentage (32,2 %) et la partie numérateur du pourcentage (29 734) pour déterminer le coefficient de variation.
3. Le numérateur, 29 734, ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage »), il faut donc utiliser le nombre qui s'en rapproche le plus, soit 30 000. De même, l'estimation du pourcentage ne figure pas parmi les entêtes de colonnes, il faut donc utiliser le nombre qui s'en rapproche le plus, à savoir 30,0 %.
4. Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 22,5 %, est le coefficient de variation à employer.
5. Le coefficient de variation approximatif est estimé à 22,5 %. Il faut donc faire une mise en garde lorsque l'on publie le résultat, soit que 32,2 % des enfants de 12 ans et plus ayant redoublé ont des troubles de la parole. Cette estimation est « médiocre », étant donné que le c.v. se situe dans la fourchette 16,5 % à 33,3 %.

Exemple 3 : Estimations des différences entre des agrégats ou des pourcentages

Supposons par exemple que d'après les données de la composante Éducation de l'ELNEJ, on estime que 60,5 % des enfants de 12 ans et plus ayant redoublés sont des garçons et que 39,5 % des enfants de 12 ans et plus ayant redoublé sont des filles. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations ?

1. À l'aide de la table de c.v. pour les enfants du groupe d'âge 12 à 15 ans utilisée de la même façon que dans l'exemple 2, vous établissez à 10,9 % le c.v. de l'estimation pour les filles et à 19,3 % le c.v. de l'estimation pour les garçons.
2. L'écart-type de l'estimation de la différence est donc donné par la règle 3, qui est,

$$s_{\hat{d}} = \sqrt{(\hat{x}_1 a_1)^2 + (\hat{x}_2 a_2)^2} \text{ soit } ((0,605 * 0,109)^2 + (0,395 * 0,193)^2)^{1/2} = 10,1 \%$$

Exemple 4 : Estimations de rapports

Supposons maintenant qu'un utilisateur désire comparer le nombre de filles de 12 à 15 ans ayant redoublé au nombre de garçons de 12 à 15 ans ayant redoublé. L'utilisateur est intéressé à comparer ces estimations sous la forme d'un rapport. Comment détermine-t-il le coefficient de variation de cette estimation?

1. Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation est le nombre de filles de 12 à 15 ans ayant redoublé et le dénominateur de l'estimation est le nombre de garçons de 12 à 15 ans ayant redoublé .
2. Reportez -vous à la table pour le groupe d'âge 12 à 15 ans.
3. Le numérateur de cette estimation de rapport est 36 475. Le chiffre qui se rapproche le plus de ce nombre est 35 000. On détermine le coefficient de variation pour cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit 24,3 %.
4. Le dénominateur de cette estimation de rapport est 55 867. Le chiffre qui se rapproche le plus de ce nombre est 55 000. On détermine le coefficient de variation de cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit 19,4 %.
5. L'écart-type de l'estimation du rapport est donc donné par la règle 4, qui est,

$$s_{\hat{R}} = \hat{R} \sqrt{a_1^2 + a_2^2} \text{ soit } (36\,475 / 55\,867) * (0,194^2 + 0,243^2)^{1/2} = 0,203$$

Le rapport filles de 12 à 15 ans ayant redoublé sur garçons de 12 à 15 ans ayant redoublé égale à 36 475 sur 55 867 soit 0,65.

Le coefficient de variation de cette estimation est égal à $0,203 / 0,65 = 31,2 \%$ c'est-à-dire « médiocre ». L'estimation peut donc être diffusée avec mise en garde puisque le c.v. se situe dans la fourchette 16,5 % à 33,3 %.

15.5 Comment utiliser les tables de c.v. pour calculer des intervalles

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance à 95 % peut être décrit de la façon suivante.

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation, alors, dans 95 % des cas, l'intervalle contiendra la valeur vraie de la population.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale 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 fois l'erreur-type et environ 99 chances sur 100 que l'écart soit inférieur à trois fois l'erreur-type. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance pour une estimation sont généralement exprimés sous forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme $(-k, +k)$, où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance pour une estimation à partir des tables de variabilité d'échantillonnage approximative, en commençant par trouver, dans la table appropriée, le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule ci-dessous pour obtenir l'intervalle de confiance IC correspondant :

$$IC_x = \left[\hat{X} - t\hat{X}s_{\hat{X}}; \hat{X} + t\hat{X}s_{\hat{X}} \right]$$

$t = 1$ si l'on désire un intervalle de confiance à 68 %

$t = 1,6$ si l'on désire un intervalle de confiance à 90 %

$t = 2$ si l'on désire un intervalle de confiance à 95 %

$t = 3$ si l'on désire un intervalle de confiance à 99 %

Une note à propos des lignes directrices pour la diffusion

Nota - Les lignes directrices qui s'appliquent à la diffusion des estimations s'applique aussi aux intervalles de confiance. Par exemple, si l'estimation est « médiocre » lui aussi. Il doit être accompagné d'une mise en garde pour avertir les utilisateurs subséquents des hauts niveaux d'erreurs.

Exemple d'utilisation de tables de c.v. pour obtenir des intervalles de confiance

Voici comment on calculerait un intervalle de confiance (IC) à 95 % pour la proportion estimée d'enfants de 12 à 15 ans ayant redoublés.

Estimation de $X = 4,4 \%$, $t = 2$.

L'estimation de l'alpha de $X = 14,8 \%$ est le coefficient de variation de cette estimation tel qu'il est déterminé à partir des tables.

$ICx = \{0,044 - (2)(0,044)(0,148); 0,044 + (2)(0,044)(0,148)\}$

$ICx = \{0,044 - 0,013; 0,044 + 0,013\}$

$ICx = \{0,031; 0,057\}$

On peut dire, avec une probabilité de 95 %, qu'entre 3,1 % et 5,7 % des enfants de 12 ans et plus ont redoublé.

15.6 Comment utiliser les tables de c.v. pour effectuer un test t

Tests d'hypothèses

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, un niveau de signification étant la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Soit \hat{X}_1 et \hat{X}_2 des estimations basées sur un échantillon pour deux caractéristiques qui nous intéressent. Supposons que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $S_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{S_{\hat{d}}}$ est compris entre -2 et 2, alors on ne peut tirer de conclusion à propos

de la différence entre les caractéristiques au niveau de 0,05. C'est-à-dire que les caractéristiques sont significatives.

Exemple d'utilisation des tables de c.v. pour effectuer un test t

Supposons que nous désirons tester, au niveau de signification de 5 %, l'hypothèse qu'il n'y a pas de différence entre la population de filles des enfants de 12 à 15 ans ayant redoublé et cette même proportion chez les garçons. Dans l'exemple 3 nous avons trouvé que l'erreur-type de la différence entre ces deux estimations était égal à 10,1 %. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{S_{\hat{d}}} = \frac{0.605 - 0.395}{0.101} = 2.08$$

Puisque $t = 2,08$ et se situe en dehors de l'intervalle $\{-2 ; 2\}$ on peut conclure avec un niveau de signification de 0,05 d'une différence entre la proportion de garçons parmi les enfants de 12 à 15 ans ayant redoublé et la proportion de filles parmi les enfants de 12 à 15 ans ayant redoublé.

15.7 Coefficients de variation d'estimations quantitatives

Estimations Quantitatives

Pour calculer l'erreur d'échantillonnage d'estimations quantitatives, il faudrait élaborer des tables spéciales, ce qui n'a pas été fait puisque la majorité des variables de la composante Éducation de l'ELNEJ sont surtout de type nominal.

Néanmoins, le coefficient de variation d'un total quantitatif est généralement supérieur à celui de l'estimation de type nominal correspondante (c'est-à-dire l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation de type nominal correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation du nombre d'élèves de 12 à 15 ans ayant redoublé serait plus élevé que le coefficient de variation de la proportion correspondante d'enfants ayant redoublé parmi les enfants de 12 à 15 ans. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, ce sera aussi le cas pour le coefficient de variation de l'estimation quantitative correspondante.

Pseudo-réplication

On peut calculer, au besoin, les coefficients de variation d'estimations de ce genre pour une estimation particulière au moyen d'une technique appelée pseudo-réplication, qui consiste à diviser les enregistrements des fichiers de microdonnées en sous-groupes (ou sous échantillons) et à calculer la variabilité de l'estimation d'un sous-échantillon à l'autre. Les utilisateurs désireux de calculer le coefficient de variation d'estimations quantitatives peuvent demander conseil à Statistique Canada en ce qui concerne la manière de répartir les enregistrements en sous-échantillons appropriés et les formules à utiliser pour ces calculs.

15.8 Seuils pour la diffusion des estimations relatives à la composante Éducation de l'ELNEJ

On trouvera dans les tableaux qui suivent les seuils de la composante Éducation de l'ELNEJ qui déterminent si les estimations sont d'une qualité « acceptable », « médiocre » ou « inacceptable ». Les utilisateurs sont encouragés à utiliser ces seuils lorsqu'ils publient des données extraites de la composante Éducation de l'ELNEJ. On trouvera tout d'abord un tableau indiquant les seuils aux niveaux provincial, régional et national. Le tableau suivant donne les seuils applicables pour les différents groupes d'âge.

Par exemple, une estimation de 5 000 concernant la Nouvelle-Écosse serait de qualité « médiocre ». Cela signifie qu'un avertissement devrait accompagner l'estimation afin de mettre les prochains utilisateurs en garde contre le taux élevé d'erreur associé à l'estimation.

Seuils pour la diffusion des estimations transversales par province et région

Province(s)	Acceptable, estimations d'au moins	Médiocre : estimations se situant entre	Inacceptable : Estimations d'au plus
Terre-Neuve-et-Labrador	11 000	3 500 – 12 000	3 500
Île-du-Prince-Édouard	4 500	1 500 – 4 500	1 500
Nouvelle-Écosse	15 500	4 500 – 15 500	4 500
Nouveau-Brunswick	15 000	4 500 – 15 000	4 500
Québec	73 500	19 500 – 73 500	19 500
Ontario	132 000	35 000 – 132 000	35 000
Manitoba	23 500	6 500 – 23 500	6 500
Saskatchewan	23 000	6 500 – 23 000	6 500
Alberta	58 000	16 500 – 58 000	16 500
Colombie-Britannique	74 500	21 000 – 74 500	21 000
Atlantiques	16 000	4 500 – 16 000	4 500
Autres	107 500	27 500 – 107 500	27 500
Prairies	50 000	13 000 – 50 000	13 000
Canada	91 500	23 000 – 91 500	23 000

Seuils pour la diffusion des estimations transversales par âge

Age	Acceptable, estimations d'au moins	Médiocre : estimations se situant entre	Inacceptable : Estimations d'au plus
Enfants âgés de 6 à 11 ans	92 000	23 500 – 92 000	23 500
Enfants âgés de 12 à 15 ans	79 000	20 500 – 79 000	20 500
Enfants âgés de 6 et 7 ans	49 500	13 000 – 49 500	13 000
Enfants âgés de 8 et 9 ans	83 000	22 500 – 83 000	22 500
Enfants âgés de 10 et 11 ans	84 500	23 000 – 84 500	23 000
Enfants âgés de 12 et 13 ans	71 000	19 000 – 71 000	19 000
Enfants âgés de 14 et 15 ans	76 000	20 500 – 76 000	20 500

15.9 Programmes SAS et SPSS pour calculer la variance, en utilisant les poids bootstrap

Des macros ont été développées, en SAS et en SPSS, pour calculer la variance avec les poids Bootstrap. Afin d'obtenir des estimations précises de la variance, cette méthode est la meilleure des trois méthodes proposées. Elle a pour avantage de calculer la variance pour la variable d'intérêt et ce, quel que soit le domaine. En plus de tenir compte du plan de sondage, cette variance tient également compte des caractéristiques spécifiques de la variable d'intérêt pour laquelle la variance doit être calculée. De plus, contrairement aux autres méthodes, nous ne sommes pas limités à certains domaines pré-identifiés.

Cette méthode présente beaucoup d'avantages mais exige un peu plus d'effort de la part du chercheur qui l'utilise. Le calcul de la variance avec l'utilisation de ces macros exige plus de temps que les deux autres méthodes. De plus, l'utilisateur doit évidemment se familiariser avec ces macros avant de les soumettre. Toutefois, ces macros ont été conçues de façon à ce qu'elles soient faciles à utiliser.

L'utilisateur doit évidemment avoir accès aux macros, aux fichiers de données et aux fichiers de poids Bootstrap. Cet accès est possible dans les Centres de Données de Recherche (CDR). De plus, de la documentation détaillée existe dans les CDR pour utiliser ces macros SAS ou SPSS pour le calcul de la variance.

Malgré le temps requis plus long pour faire le calcul, cette méthode est celle qui est recommandée pour calculer la variance d'estimations qui doivent être publiées dans des articles analytiques. Cette méthode permet d'obtenir une variance qui est beaucoup plus précise qu'avec les autres méthodes proposées. Le fichier contenant les poids Bootstrap a 1 000 variables (BSW1 à BSW1000) correspondant aux 1 000 poids bootstrap et des variables d'identification de l'enfant.

16.0 Direct Assessment

La recherche sur la petite enfance joue un rôle de premier plan dans l'élaboration des politiques relatives aux jeunes enfants. L'ajout d'outils d'évaluation du développement des jeunes enfants à l'ELNEJ permettra de mieux comprendre les processus du développement au cours de la petite enfance et fournira des données pertinentes qui serviront de fondement aux orientations stratégiques visant cette population.

Les outils d'évaluation retenus aux fins du cycle 4 de l'ELNEJ (collecte des données en 2000-2001) ont été choisis à partir d'une recension détaillée des écrits, de l'élaboration d'un cadre de recherche sur le développement et l'apprentissage de l'enfant, de consultations auprès de nombreux experts au Canada et à l'étranger, de l'examen des documents traitant des différents outils possibles et de l'essai sur le terrain des outils les plus prometteurs. Ces outils ont également été examinés à la lumière de divers critères, notamment leur fiabilité et leur validité, la couverture des domaines visés par le cadre de recherche, la capacité de l'outil de distinguer le développement normal et les retards de développement, la facilité d'utilisation par des intervieweurs profanes et la disponibilité de l'outil en anglais ou en français (ou la facilité de traduction en français ou en anglais). La décision finale a été largement conditionnée par l'opinion des principaux experts qui conseillent l'équipe de l'ELNEJ depuis un certain temps.

Dû à la nature unique de ce fichier, certaines variables du fichier principal sont incluses afin d'encourager l'analyse. Les variables suivantes ont été incluses pour chaque répondant où l'information s'applique : les scores EVIP-R (cycles 1 à 4), les scores en mathématique (cycles 1 à 4), les scores en rédaction (cycles 2 et 3). Ces mesures sont décrites en détails dans le guide de l'utilisateur du fichier principal.

17.0 Enjeux analytiques

Le but de ce chapitre est de donner aux utilisateurs des données de l'ELNEJ un aperçu des différents enjeux analytiques qu'ils doivent considérer lors de tout projet d'analyse avec les données de l'ELNEJ. Quelques points soulevés dans ce chapitre l'ont déjà été avec plus de détails dans d'autres chapitres de ce guide. Toutefois, l'objectif ici est de réunir, dans un seul chapitre, tous les enjeux importants ayant un impact potentiel sur l'analyse de données.

17.1 Impact de l'utilisation d'un plan de sondage complexe sur l'analyse

Tel que décrit dans le chapitre sur l'échantillon (chapitre 5) et celui sur la pondération (chapitre 12), les enfants dans l'échantillon de l'ELNEJ ont été sélectionnés à partir d'un plan de sondage complexe, afin de répondre aux besoins des clients, tout en tenant compte de certaines contraintes opérationnelles. Afin de faire un bon usage des données et de pouvoir produire des estimations valides, les poids de sondage, longitudinaux ou transversaux selon le type d'analyse, doivent être utilisés.

17.2 L'unité d'analyse

Pour l'ELNEJ, l'unité d'analyse est toujours l'enfant. Certains éléments d'information ont été recueillis au niveau du ménage, mais il n'est pas possible de faire d'estimations au niveau du ménage. Toutes les estimations doivent être faites au niveau de l'enfant.

17.3 Type d'analyse : transversale, longitudinale ou répétée

Étant donné la réduction de l'échantillon de la composante de l'éducation de l'enquête, les analystes devront composer avec les effets d'un plus petit échantillon sur le type d'analyse qui peut être effectuée. Il est clair qu'avec la structure actuelle de ce fichier, nous avons favorisé l'analyse transversale de ces données. Mais ceci étant dit, les analystes ne sont pas limités à ce type d'analyse seulement.

Avec l'ELNEJ, les utilisateurs ont la possibilité de faire différents types d'analyse : analyse longitudinale, analyse transversale ou même considérer les données de l'ELNEJ comme venant d'une enquête répétée. Pour chacun de ces types d'analyse, différents enjeux doivent être considérés.

17.3.1 Analyse longitudinale

Pour toute analyse longitudinale, la population représentée par les poids longitudinaux est toujours la population au moment de la sélection initiale de l'enfant. Maintenant rendu au cycle 4, l'analyste doit faire face aux problèmes d'érosion de l'échantillon, de représentativité dû à la non-réponse totale et de non-réponse partielle. De plus, comme les enfants sélectionnés au cycle 1 ont répondu à l'enquête à 4 reprises maintenant, un biais d'accoutumance peut se développer. Tous ces problèmes sont définis plus loin dans ce chapitre.

De plus, pour la première fois au cycle 4, deux ensembles de poids longitudinaux sont disponibles. Le premier ensemble comprend les enfants longitudinaux qui ont répondu au cycle 4, sans égard à leur réponse aux cycles précédents. Le deuxième ensemble de poids longitudinaux, calculés pour les enfants introduits au cycle 1, comprend les enfants qui ont répondu à tous les cycles.

17.3.2 Analyse transversale

Pour toute analyse transversale, la population représentée par les poids transversaux est toujours la population du cycle actuel, c'est-à-dire le cycle 4. En plus des problèmes ci-haut mentionnés pour l'analyse longitudinale, un problème additionnel doit être pris en considération pour l'analyse transversale : la représentativité de la population actuelle. Tel que décrit dans le chapitre 5 du guide de l'utilisateur du fichier principal, l'échantillon transversal est composé d'enfants sélectionnés à différents moments dans le temps. Par exemple, les enfants de 6 à 17 ans en 2000 ont été choisis en 1994. Depuis, l'échantillon n'a pas été renouvelé. Donc, les immigrants depuis 1994 n'ont pas la chance de faire partie de cet échantillon. L'impact de cette situation est détaillé dans le chapitre sur la qualité des données. De plus, le sous-échantillonnage effectué pour la composante de l'éducation (décrit au chapitre 5 de ce guide) aura un effet sur la précision des estimations, en particulier pour les sous-domaines, car la taille de l'échantillon restant a été réduite.

17.3.3 Analyse répétée

Une analyse répétée est une analyse pour laquelle une même population cible est étudiée à différents moments dans le temps. Par exemple, avec l'ELNEJ, il est possible d'étudier les enfants d'un an en 1994-1995, en 1996-1997, en 1998-1999 et en 2000-2001.

Pour la composante de l'éducation, ce type d'analyse sera compliqué par le fait qu'il n'y a pas eu de repondération aux cycles précédents. Le traitement inégal de la non-réponse aura des répercussions sur toute analyse dans ce cas. Les utilisateurs qui veulent effectuer cette analyse délaisseraient probablement le poids ajusté et se tourneraient vers le poids de la composante principale en rattachant le sous-ensemble à l'échantillon complet. Selon la population cible utilisée, certains des enjeux présentés pour l'analyse longitudinale ou l'analyse transversale peuvent se présenter. D'autres enjeux sont également à considérer. Tout d'abord, pour un même groupe d'âge, le nombre d'enfants dans l'échantillon peut varier beaucoup d'un cycle à l'autre. Ceci est particulièrement vrai pour les enfants de 1 an et de 5 ans.

De plus, comme la composante de l'éducation ne s'applique qu'à la partie longitudinale de l'échantillon au cycle 4, les analystes doivent se rendre compte que les échantillons sont dépendants d'un cycle à l'autre. Les enfants pour qui les données de la composante de l'éducation sont recueillies faisaient partie de l'échantillon au cycle précédent. Même dans le cas de cohortes d'âge distinctes, bien des répondants peuvent en fait être les frères ou sœurs de répondants du cycle suivant.

Un certain nombre d'autres problèmes sont examinés dans le guide de l'utilisateur du fichier principal.

17.4 Non-réponse totale et partielle

Comme toute enquête, l'ELNEJ est confrontée à de la non-réponse. Il existe deux grandes catégories de non-réponse, la non-réponse totale et la non-réponse partielle. Par définition, un enfant répondant est un enfant pour lequel une composante adulte ou sa composante enfant ou jeune a été complétée. Un ménage répondant sans composante adulte complète pourrait avoir 1 enfant répondant et 1 enfant non-répondant.

17.4.1 Non-réponse totale

La non-réponse totale est l'absence de données pour un enfant sélectionné, à cause de facteurs tels que le refus de participer ou l'impossibilité de dépister l'enfant. La non-réponse totale est prise en considération et corrigée lors de la pondération. D'ailleurs, le chapitre sur la pondération décrit bien le processus. Plus la non-réponse totale est importante, plus il devient difficile d'ajuster le poids des unités répondantes pour tenir compte de cette non-réponse. De plus, du point de vue longitudinal, avec le premier ensemble de poids longitudinaux décrits ci-haut, il est possible qu'un enfant répondant au cycle 4 soit un non-répondant à un cycle précédent. L'analyste doit donc en tenir compte dans son étude.

17.4.2 Non-réponse partielle

Au cycle 4, pour qu'un enfant soit déclaré répondant, la composante enfant/jeune ou la composante de la Personne qui Connait Mieux (PCM) l'enfant doit être complétée. Ceci signifie donc qu'il est possible de retrouver de la non-réponse à certaines questions ou même à certaines composantes complètes et que l'enfant soit tout de même un répondant. Pour un enfant donné, les sources d'information sont nombreuses. Selon l'âge de l'enfant, l'information peut venir d'une ou de plusieurs des sources suivantes : de lui-même, d'un de ses parents, du professeur, du directeur d'école. Conséquemment, il est possible qu'une de ces sources d'information n'ait pas fourni l'information désirée mais que l'enfant soit quand même considéré répondant.

Seules les variables de revenu sont imputées donc, toutes les autres variables sont laissées manquantes (identifiées par Ne sait pas ou Non déclaré). L'analyste doit donc tenir compte de la non-réponse en insistant sur les limitations des données due à cette non-réponse partielle dans son analyse.

17.5 Traitement des données

Le contenu de l'ELNEJ est très riche. Afin de pouvoir rendre disponibles les données dans un temps raisonnable, il est impossible de vérifier en profondeur chacune des variables. Conséquemment, lors du traitement des données, on se concentre sur des variables-clés de l'enquête. Il est ainsi possible que des erreurs se glissent dans les données. Toutefois, des efforts sont déployés afin de réduire au maximum le nombre d'erreurs.

17.6 Calcul de la variance avec certains logiciels

Les logiciels SAS ou SPSS utilisés couramment par les analystes peuvent calculer correctement les estimations en utilisant les poids de sondage. À l'exception de quelques procédures en SAS, ces logiciels ne sont pas programmés pour tenir compte des effets de plan de sondage lors du calcul de la variance. En effet, ces logiciels ne peuvent pas tenir compte de la complexité de l'ELNEJ (plan de sondage complexe, ajustement de non-réponse et post-stratification pour tenir compte des estimations démographiques) pour calculer la variance. Il en résulte habituellement une sous-estimation de celle-ci, sous-estimation qui peut être quelquefois importante. L'utilisateur est donc encouragé à utiliser les poids Bootstrap pour l'estimation de la variance.

Certains logiciels sont capables d'utiliser les poids Bootstrap. SUDAAN et WesVar calculent des estimations de variance basées sur le plan d'échantillonnage au moyen de la méthode de rééchantillonnage à répliques équilibrées (RRE). La création de poids RRE diffère en théorie de celle des poids Bootstrap mais l'estimateur de variance est le même, alors les poids Bootstrap de l'ELNEJ peuvent être utilisés dans ces logiciels.

Il existe une version autonome de SUDAAN, ainsi qu'une version qui s'intègre avec SAS. La dernière accorde la flexibilité de faire appel aux procédures de SUDAAN à l'intérieur d'un programme SAS. À l'aide de ses neuf procédures, on est capable de calculer des estimations de moyennes, ratios et totaux, d'effectuer des tests d'indépendance et d'homogénéité, et d'ajuster des modèles de régression linéaire, log-linéaire, logistique et de survie. SUDAAN peut faire la lecture de fichiers SAS, SPSS et d'autres formats communs.

Le logiciel WesVar présente une approche "point and click" à l'analyse, ce qui le rend très facile à apprendre. Les types d'analyse sont plus limités que dans SUDAAN, mais WesVar offre quand même le calcul des estimations de moyennes, ratios et totaux, des tests d'indépendance et d'homogénéité, et la modélisation linéaire et logistique. La lecture de fichiers SAS, SPSS et d'autres formats communs est possible avec WesVar, mais le logiciel requiert l'étape supplémentaire de créer un fichier de données WesVar pour effectuer l'analyse.

Les poids Bootstrap peuvent être utilisés avec d'autres logiciels qui offrent un environnement de programmation et les techniques d'analyse désirées. Des macros SAS et SPSS ont été conçues par l'ELNEJ afin d'utiliser les poids bootstrap pour arriver à des estimations de variance basées sur le plan d'échantillonnage.

18.0 Annexe I - Clichés d'enregistrement

Cette annexe décrit les fichiers de données. On y indique pour chacune des variables son emplacement dans le fichier, son nom, sa longueur et si elle est numérique ou alphabétique.

Exemple :

Position du début de la variable	Nom de la variable	Longueur de la variable	Caractéristiques de la variable
@01213	DEDUW01C	12.4	Valeur numérique comportant quatre décimales et d'une longueur maximale de 12 caractères en incluant le symbole de la décimale.
@00840	DMMCQ02	\$1.	Le symbole \$ indique une valeur alphabétique d'une longueur maximale de 1 caractère.

18.1 Fichier de l'éducation

@00001	DEDUTQ01	1.	@00037	DEDUSQ35	1.
@00002	DEDUTQ02	1.	@00038	DEDUSQ36	1.
@00003	DEDUTQ03	1.	@00039	DEDUSQ37	1.
@00004	DEDUTQ04	1.	@00040	DEDUSQ38	1.
@00005	DEDUTQ05	1.	@00041	DEDUSQ39	1.
@00006	DEDUTQ06	1.	@00042	DEDUSQ40	1.
@00007	DEDUTQ07	1.	@00043	DEDUSQ41	1.
@00008	DEDUTQ08	1.	@00044	DEDUSQ42	1.
@00009	DEDUTQ09	1.	@00045	DEDUTQ43	1.
@00010	DEDUTQ10	1.	@00046	DEDUTQ44	1.
@00011	DEDUTQ11	1.	@00047	DEDUSQ45	2.
@00012	DEDUMQ13	1.	@00049	DEDUSQ46	2.
@00013	DEDUTQ12	1.	@00051	DEDUSQ47	2.
@00014	DEDUSQ14	2.	@00053	DEDUSQ48	2.
@00016	DEDUMQ15	2.	@00055	DEDUSQ49	2.
@00018	DEDUSQ16	1.	@00057	DEDUSQ50	2.
@00019	DEDUSQ17	1.	@00059	DEDUSQ51	2.
@00020	DEDUSQ18	1.	@00061	DEDUSQ52	2.
@00021	DEDUSQ19	1.	@00063	DEDUTQ54	1.
@00022	DEDUSQ20	1.	@00064	DEDUTQ55	1.
@00023	DEDUSQ21	1.	@00065	DEDUTQ56	1.
@00024	DEDUSQ22	1.	@00066	DEDUTQ57	1.
@00025	DEDUSQ23	1.	@00067	DEDUTQ 58	1.
@00026	DEDUSQ24	1.	@00068	DEDUTQ59	1.
@00027	DEDUSQ25	1.	@00069	DEDUTQ60	1.
@00028	DEDUSQ26	1.	@00070	DEDUTQ53	1.
@00029	DEDUSQ27	1.	@00071	DEDUTQ61	1.
@00030	DEDUSQ28	1.	@00072	DEDUTQ62	1.
@00031	DEDUSQ29	1.	@00073	DEDUTQ63	1.
@00032	DEDUSQ30	1.	@00074	DEDUTQ64	1.
@00033	DEDUSQ31	1.	@00075	DEDUTQ65	1.
@00034	DEDUSQ32	1.	@00076	DEDUTQ66	1.
@00035	DEDUSQ33	1.	@00077	DEDUTQ67	1.
@00036	DEDUSQ34	1.	@00078	DEDUTQ68	1.

@00079	DEDUTQ69	1.	@00174	DEDUS129	2.
@00080	DEDUSQ70	1.	@00176	DEDUS130	2.
@00081	DEDUTQ71	1.	@00178	DEDUS131	2.
@00082	DEDUMQ72	2.	@00180	DEDUT132	2.
@00084	DEDUTQ73	2.	@00182	DEDUT133	2.
@00086	DEDUTQ74	1.	@00184	DEDUT134	2.
@00087	DEDUTQ75	1.	@00186	DEDUT135	2.
@00088	DEDUTQ76	1.	@00188	DEDUT136	2.
@00089	DEDUTQ77	2.	@00190	DEDUT137	2.
@00091	DEDUMQ78	1.	@00192	DEDUT138	2.
@00092	DEDUMQ79	1.	@00194	DEDUT139	2.
@00093	DEDUMQ80	1.	@00196	DEDUT140	2.
@00094	DEDUMQ81	1.	@00198	DEDUS141	1.
@00095	DEDUMQ82	1.	@00199	DEDUS142	1.
@00096	DEDUMQ83	1.	@00200	DEDUS143	1.
@00097	DEDUMQ84	2.	@00201	DEDUS144	1.
@00099	DEDUTQ86	1.	@00202	DEDUS145	1.
@00100	DEDUSQ87	1.	@00203	DEDUS146	1.
@00101	DEDUSQ88	1.	@00204	DEDUS147	1.
@00102	DEDUSQ89	1.	@00205	DEDUS148	1.
@00103	DEDUSQ90	1.	@00206	DEDUS149	1.
@00104	DEDUSQ91	1.	@00207	DEDUS150	1.
@00105	DEDUSQ92	1.	@00208	DEDUS151	1.
@00106	DEDUSQ93	1.	@00209	DEDUS152	1.
@00107	DEDUSQ94	1.	@00210	DEDUS153	1.
@00108	DEDUSQ95	2.	@00211	DEDUS154	1.
@00110	DEDUSQ96	2.	@00212	DEDUS155	1.
@00112	DEDUSQ97	2.	@00213	DEDUS156	1.
@00114	DEDUSQ98	2.	@00214	DEDUS157	1.
@00116	DEDUSQ99	2.	@00215	DEDUS158	1.
@00118	DEDUT100	2.	@00216	DEDUS159	1.
@00120	DEDUM101	2.	@00217	DEDUS160	1.
@00122	DEDUM102	1.	@00218	DEDUS161	1.
@00123	DEDUM103	1.	@00219	DEDUS162	1.
@00124	DEDUM104	2.	@00220	DEDUS163	1.
@00126	DEDUM105	2.	@00221	DEDUS164	1.
@00128	DEDUM106	2.	@00222	DEDUS165	1.
@00130	DEDUM107	2.	@00223	DEDUS166	1.
@00132	DEDUM108	2.	@00224	DEDUS167	1.
@00134	DEDUM109	2.	@00225	DEDUS168	1.
@00136	DEDUM110	2.	@00226	DEDUS169	1.
@00138	DEDUM111	2.	@00227	DEDUS170	1.
@00140	DEDUS112	2.	@00228	DEDUS171	1.
@00142	DEDUS113	2.	@00229	DEDUS172	1.
@00144	DEDUS114	2.	@00230	DEDUS173	1.
@00146	DEDUT115	2.	@00231	DEDUS174	1.
@00148	DEDUT116	2.	@00232	DEDUS175	1.
@00150	DEDUT117	2.	@00233	DEDUS176	1.
@00152	DEDUT118	2.	@00234	DEDUS177	1.
@00154	DEDUT119	2.	@00235	DEDUS178	1.
@00156	DEDUT120	2.	@00236	DEDUS179	1.
@00158	DEDUT121	2.	@00237	DEDUS180	1.
@00160	DEDUT122	2.	@00238	DEDUS181	1.
@00162	DEDUS123	2.	@00239	DEDUS182	1.
@00164	DEDUS124	2.	@00240	DEDUS183	1.
@00166	DEDUT125	2.	@00241	DEDUT184	2.
@00168	DEDUT126	2.	@00243	DEDUT185	2.
@00170	DEDUS127	2.	@00245	DEDUT187	2.
@00172	DEDUT128	2.	@00247	DEDUT188	2.

@00249	DEDUT189	1.	@00343	DEDUM249	2.
@00250	DEDUM190	1.	@00345	DEDUM250	2.
@00251	DEDUM191	1.	@00347	DEDUS251	1.
@00252	DEDUM192	1.	@00348	DEDUS2 52	2.
@00253	DEDUM193	1.	@00350	DEDUS253	2.
@00254	DEDUM194	1.	@00352	DEDUS254	2.
@00255	DEDUM195	1.	@00354	DEDUS255	2.
@00256	DEDUM196	1.	@00356	DEDUS256	2.
@00257	DEDUM197	2.	@00358	DEDUS257	2.
@00259	DEDUT198	1.	@00360	DEDUS258	1.
@00260	DEDUS200	1.	@00361	DEDUS259	2.
@00261	DEDUS201	1.	@00363	DEDUS260	2.
@00262	DEDUS202	1.	@00365	DEDUS261	2.
@00263	DEDUS203	1.	@00367	DEDUS262	2.
@00264	DEDUS204	1.	@00369	DEDUS263	2.
@00265	DEDUT205	1.	@00371	DEDUS264	2.
@00266	DEDUT206	1.	@00373	DEDUS265	1.
@00267	DEDUT207	1.	@00374	DEDUS266	2.
@00268	DEDUT208	1.	@00376	DEDUS267	2.
@00269	DEDUT209	1.	@00378	DEDUS268	2.
@00270	DEDUT210	1.	@00380	DEDUS269	2.
@00271	DEDUT211	1.	@00382	DEDUS270	2.
@00272	DEDUM212	1.	@00384	DEDUS271	2.
@00273	DEDUM213	1.	@00386	DEDUT272	2.
@00274	DEDUM214	1.	@00388	DEDUS273	2.
@00275	DEDUM215	1.	@00390	DEDUT274	2.
@00276	DEDUM216	1.	@00392	DEDUT275	2.
@00277	DEDUM217	1.	@00394	DEDUT276	2.
@00278	DEDUT218	2.	@00396	DEDUT277	2.
@00280	DEDUT219	2.	@00398	DEDUT278	2.
@00282	DEDUT220	2.	@00400	DEDUT279	2.
@00284	DEDUT221	2.	@00402	DEDUT280	2.
@00286	DEDUS222	2.	@00404	DEDUT281	2.
@00288	DEDUS223	2.	@00406	DEDUT282	2.
@00290	DEDUT224	2.	@00408	DEDUT283	2.
@00292	DEDUT225	2.	@00410	DEDUT2 84	2.
@00294	DEDUT226	2.	@00412	DEDUT285	2.
@00296	DEDUS227	3.	@00414	DEDUT286	2.
@00299	DEDUS228	3.	@00416	DEDUT287	2.
@00302	DEDUS229	3.	@00418	DEDUT288	2.
@00305	DEDUS230	2.	@00420	DEDUT289	2.
@00307	DEDUS231	2.	@00422	DEDUT290	2.
@00309	DEDUT232	2.	@00424	DEDUT291	2.
@00311	DEDUT233	2.	@00426	DEDUT292	2.
@00313	DEDUT234	2.	@00428	DEDUT293	2.
@00315	DEDUT235	2.	@00430	DEDUT294	2.
@00317	DEDUT236	2.	@00432	DEDUT295	2.
@00319	DEDUT237	2.	@00434	DEDUT296	2.
@00321	DEDUT238	2.	@00436	DEDUT297	1.
@00323	DEDUT239	2.	@00437	DEDUT298	2.
@00325	DEDUT240	2.	@00439	DEDUT299	3.
@00327	DEDUT241	2.	@00442	DEDUT300	3.
@00329	DEDUT242	2.	@00445	DEDUT301	3.
@00331	DEDUT243	2.	@00448	DEDUT302	2.
@00333	DEDUT244	2.	@00450	DEDUT303	2.
@00335	DEDUM245	2.	@00452	DEDUT304	2.
@00337	DEDUM246	2.	@00454	DEDUM305	2.
@00339	DEDUM247	2.	@00456	DEDUS306	1.
@00341	DEDUM248	2.	@00457	DEDUS307	1.

@00458	DEDUS308	1.	@00517	DEDUP370	1.
@00459	DEDUS309	1.	@00518	DEDUP371	1.
@00460	DEDUS310	1.	@00519	DEDUP372	1.
@00461	DEDUS311	1.	@00520	DEDUP373	1.
@00462	DEDUS312	1.	@00521	DEDUP374	1.
@00463	DEDUS313	1.	@00522	DEDUP375	1.
@00464	DEDUS314	1.	@00523	DEDUP376	1.
@00465	DEDUS315	1.	@00524	DEDUP377	1.
@00466	DEDUS316	1.	@00525	DEDUP378	1.
@00467	DEDUS317	1.	@00526	DEDUP379	1.
@00468	DEDUT318	1.	@00527	DEDUP380	1.
@00469	DEDUT319	1.	@00528	DEDUP381	1.
@00470	DEDUT320	1.	@00529	DEDUP382	1.
@00471	DEDUT321	1.	@00530	DEDUP383	1.
@00472	DEDUT322	1.	@00531	DEDUP384	4.
@00473	DEDUT323	1.	@00535	DEDUP385	3.
@00474	DEDUT324	1.	@00538	DEDUP386	3.
@00475	DEDUT325	1.	@00541	DEDUP387	3.
@00476	DEDUT326	1.	@00544	DEDUP388	3.
@00477	DEDUT327	1.	@00547	DEDUP389	3.
@00478	DEDUT328	1.	@00550	DEDUP390	3.
@00479	DEDUT329	1.	@00553	DEDUP391	2.
@00480	DEDUT330	1.	@00555	DEDUP392	2.
@00481	DEDUT331	1.	@00557	DEDUP393	2.
@00482	DEDUT332	1.	@00559	DEDUP394	2.
@00483	DEDUT333	1.	@00561	DEDUP395	2.
@00484	DEDUT334	1.	@00563	DEDUP396	2.
@00485	DEDUT335	1.	@00565	DEDUP397	2.
@00486	DEDUT337	1.	@00567	DEDUP398	2.
@00487	DEDUT338	1.	@00569	DEDUP399	2.
@00488	DEDUT339	1.	@00571	DEDUP400	2.
@00489	DEDUT340	1.	@00573	DEDUP401	2.
@00490	DEDUT342	1.	@00575	DEDUP402	2.
@00491	DEDUP344	1.	@00577	DEDUP403	2.
@00492	DEDUP345	1.	@00579	DEDUP404	2.
@00493	DEDUP346	1.	@00581	DEDUP405	2.
@00494	DEDUP347	1.	@00583	DEDUP406	3.
@00495	DEDUP348	1.	@00586	DEDUP407	3.
@00496	DEDUP349	1.	@00589	DEDUP408	3.
@00497	DEDUP350	1.	@00592	DEDUP410	3.
@00498	DEDUP351	1.	@00595	DEDUP411	3.
@00499	DEDUP352	1.	@00598	DEDUP412	3.
@00500	DEDUP353	1.	@00601	DEDUP413	3.
@00501	DEDUP354	1.	@00604	DEDUP414	3.
@00502	DEDUP355	1.	@00607	DEDUP415	3.
@00503	DEDUP356	1.	@00610	DEDUP416	3.
@00504	DEDUP357	1.	@00613	DEDUP417	3.
@00505	DEDUP358	1.	@00616	DEDUP418	3.
@00506	DEDUP359	1.	@00619	DEDUP419	3.
@00507	DEDUP360	1.	@00622	DEDUP420	3.
@00508	DEDUP361	1.	@00625	DEDUP421	3.
@00509	DEDUP362	1.	@00628	DEDUP422	3.
@00510	DEDUP363	1.	@00631	DEDUP423	2.
@00511	DEDUP364	1.	@00633	DEDUP424	2.
@00512	DEDUP365	1.	@00635	DEDUP425	2.
@00513	DEDUP366	1.	@00637	DEDUP426	2.
@00514	DEDUP367	1.	@00639	DEDUP427	2.
@00515	DEDUP368	1.	@00641	DEDUP428	2.
@00516	DEDUP369	1.	@00643	DEDUP429	2.

@00645	DEDUP430	2.	@00747	DEDUP496	2.
@00647	DEDUP431	2.	@00749	DEDUP497	2.
@00649	DEDUP432	2.	@00751	DEDUP498	2.
@00651	DEDUP433	2.	@00753	DEDUP499	2.
@00653	DEDUP435	2.	@00755	DEDUP500	2.
@00655	DEDUP436	2.	@00757	DEDUP501	2.
@00657	DEDUP438	3.	@00759	DEDUP502	1.
@00660	DEDUP439	3.	@00760	DEDUP503	2.
@00663	DEDUP441	3.	@00762	DEDUP504	3.
@00666	DEDUP442	3.	@00765	DEDUP505	3.
@00669	DEDUP444	3.	@00768	DEDUP506	3.
@00672	DEDUP445	3.	@00771	DEDUP507	3.
@00675	DEDUP447	3.	@00774	DEDUP508	3.
@00678	DEDUP448	3.	@00777	DEDUP509	3.
@00681	DEDUP450	3.	@00780	DEDUP510	1.
@00684	DEDUP451	3.	@00781	DEDUP511	1.
@00687	DEDUP453	1.	@00782	DEDUP512	1.
@00688	DEDUP454	1.	@00783	DEDUP513	1.
@00689	DEDUP455	1.	@00784	DEDUP514	1.
@00690	DEDUP456	1.	@00785	DEDUP515	1.
@00691	DEDUP457	1.	@00786	DEDUP516	1.
@00692	DEDUP458	1.	@00787	DEDUP517	1.
@00693	DEDUP459	1.	@00788	DEDUP518	1.
@00694	DEDUP460	1.	@00789	DEDUP519	1.
@00695	DEDUP461	1.	@00790	DEDUP520	1.
@00696	DEDUP462	1.	@00791	DEPCS15	2.
@00697	DEDUP463	1.	@00793	DETSS17	2.
@00698	DEDUP464	1.	@00795	DETSS26	2.
@00699	DEDUP465	1.	@00797	DETSS28A	2.
@00700	DEDUP466	1.	@00799	DETSS28C	2.
@00701	DEDUP467	1.	@00801	DETSS50A	2.
@00702	DEDUP468	1.	@00803	DETSS50B	2.
@00703	DEDUP469	1.	@00805	DETSS51	2.
@00704	DEDUP470	1.	@00807	DGEHD03	2.
@00705	DEDUP471	1.	@00809	DMMPQ01	3.
@00706	DEDUP472	1.	@00812	DMMPQ02	\$1.
@00707	DEDUP473	1.	@00813	DMMPQ03A	4.
@00708	DEDUP474	1.	@00817	DMMPQ03B	2.
@00709	DEDUP475	1.	@00819	DMMPQ03C	2.
@00710	DEDUP476	1.	@00821	DMMPQ04	2.
@00711	DEDUP477	1.	@00823	DMMSQ01	3.
@00712	DEDUP478	1.	@00826	DMMSQ02	\$1.
@00713	DEDUP479	1.	@00827	DMMSQ03A	4.
@00714	DEDUP480	1.	@00831	DMMSQ03B	2.
@00715	DEDUP481	1.	@00833	DMMSQ03C	2.
@00716	DEDUP482	1.	@00835	DMMSQ04	2.
@00717	DEDUP483	1.	@00837	DMMCQ01	3.
@00718	DEDUP484	1.	@00840	DMMCQ02	\$1.
@00719	DEDUP485	3.	@00841	DMMCQ03A	4.
@00722	DEDUP486	3.	@00845	DMMCQ03B	2.
@00725	DEDUP487	3.	@00847	DMMCQ03C	2.
@00728	DEDUP488	3.	@00849	DDMCD03	2.
@00731	DEDUP489	3.	@00851	DDMCD06	2.
@00734	DEDUP490	3.	@00853	DDMCD06B	2.
@00737	DEDUP491	2.	@00855	DDMCD15	1.
@00739	DEDUP492	2.	@00856	DDMCD16	1.
@00741	DEDUP493	2.	@00857	DDMCD18	2.
@00743	DEDUP494	2.	@00859	DDMCD19	2.
@00745	DEDUP495	2.	@00861	DEDPdQ4b	2.

@00863	DEDSdQ4b	2.	@00998	DEDCdQ7F	2.
@00865	DEDPD01	2.	@01000	DEDCdQ9A	1.
@00867	DESD01	2.	@01001	DEDCdQ9B	2.
@00869	DEDHcQ8A	1.	@01003	DEDCQ10	2.
@00870	DEDHcQ8B	1.	@01005	DEDCd11A	2.
@00871	DEDHcQ8C	1.	@01007	DEDCQ14A	2.
@00872	DEDHcQ8D	1.	@01009	DEDCb14A	2.
@00873	DEDHcQ8E	1.	@01011	DEDCQ14B	2.
@00874	DLFPcD5A	4.	@01013	DEDCQ14C	2.
@00878	DLFPcD6A	\$4.	@01015	DEDCb14C	2.
@00882	DLFPcD7A	2.	@01017	DEDCQ14D	2.
@00884	DLFPcD8A	2.	@01019	DEDCb14H	2.
@00886	DLFPD25	1.	@01021	DEDCd15C	1.
@00887	DLFPD34	2.	@01022	DEDCd15D	2.
@00889	DLFPbD38	2.	@01024	DEDCQ17	2.
@00891	DLFPD51	1.	@01026	DEDCQ18A	1.
@00892	DLFScD5A	4.	@01027	DEDCQ18B	2.
@00896	DLFScD6A	\$4.	@01029	DEDCc18C	2.
@00900	DLFScD7A	2.	@01031	DEDCc18D	2.
@00902	DLFScD8A	2.	@01033	DHLCcD2A	4. 2
@00904	DLFSD25	1.	@01037	DHLCbD4C	5. 3
@00905	DLFSD34	2.	@01042	DHLCbD4D	1.
@00907	DLFSbD38	2.	@01043	DHLCbD45	1.
@00909	DLFSD51	1.	@01044	DBECdS06	2.
@00910	DLFHD49B	2.	@01046	DBECdS07	2.
@00912	DLFHD50	2.	@01048	DBECdS08	2.
@00914	DINHQ03	7.	@01050	DBECdS09	2.
@00921	DINPD02	2.	@01052	DBECS10	2.
@00923	DINPCD04	7.	@01054	DBECdS11	2.
@00930	DINScD02	2.	@01056	DSLcdQ7	2.
@00932	DINScD04	7.	@01058	DPRCS03	2.
@00939	DINHd01B	2.	@01060	DPRCS04	2.
@00941	DCHPD01	1.	@01062	DPRCS05	2.
@00942	DRSPdD01	1.	@01064	DPRCS06	2.
@00943	DCHSD01	1.	@01066	DPRCbS09	2.
@00944	DRSSdD01	1.	@01068	DSDCQ1	2.
@00945	DDPPS01	2.	@01070	DSDCQ3	4.
@00947	DFNHS01	2.	@01074	DSDCD01	3.
@00949	DSFHS5	2.	@01077	DSDCD02	3.
@00951	DSFHS6	2.	@01080	DSDCD03	2.
@00953	DSPHS01	2.	@01082	DSDCD04	2.
@00955	DSDPQ1	2.	@01084	DSDCD05	2.
@00957	DSDPQ3	4.	@01086	DPPCS01	3.
@00961	DSDSQ1	2.	@01089	DPPCS02	3.
@00963	DSDSQ3	4.	@01092	DMACS01	2.
@00967	DSDPD01	3.	@01094	DMACS02	3.
@00970	DSDPD02	3.	@01097	DMACS03	3.
@00973	DSDPD03	2.	@01100	DGEHbD04	2.
@00975	DSDPD04	2.	@01102	DDMCD08	2.
@00977	DSDPD05	2.	@01104	DLFPQ02	1.
@00979	DSDSD01	3.	@01105	DINHdQ05	1.
@00982	DSDSD02	3.	@01106	DHLPQ01	2.
@00985	DSDSD03	2.	@01108	DHLPQ02	1.
@00987	DSDSD04	2.	@01109	DSDPQ8	2.
@00989	DSDSD05	2.	@01111	DHHHQ01	1.
@00991	DEDCdQ0	2.	@01112	DHHHQ06	2.
@00993	DEDCcQ0A	2.	@01114	DHLCQ01	2.
@00995	DEDCQ06	1.	@01116	DHLCQ45I	1.
@00996	DEDCD03	2.	@01117	DHLCQ52A	1.

@01118	DRLCQ01	2.	@01152	BMACS02	3.
@01120	DLTCdQ14	2.	@01155	BRECS01	2.
@01122	DPRCQ03	2.	@01157	BRECS02	3.
@01124	DPRCQ07	2.	@01160	APPCS01	3.
@01126	DPRCQ11	2.	@01163	APPCS02	3.
@01128	CPPCS01	3.	@01166	AMACS01	2.
@01131	CPPCS02	3.	@01168	AMACS02	3.
@01134	CMACS01	2.	@01171	FLAG_TE	1.
@01136	CMACS02	3.	@01172	FLAG_PR	1.
@01139	CRECS01	2.	@01173	PERSRUK	\$14.
@01141	CRECS02	3.	@01187	FIELDRUK	\$12.
@01144	BPPCS01	3.	@01199	DSCH_ID	8.
@01147	BPPCS02	3.	@01207	CHILDIID	6.
@01150	BMACS01	2.	@01213	DEDUW01C	12. 4;

18.2 Fichier secondaire

@00001	DAGCdQ1	1.	@00041	DAGCd387	1.
@00002	DAGCdQ2	1.	@00042	DAGCd388	1.
@00003	DAGCdQ3	1.	@00043	DAGCd389	1.
@00004	DAGCdQ4	1.	@00044	DAGCd390	1.
@00005	DAGCdQ5	1.	@00045	DAGCd393	1.
@00006	DAGCdQ6	1.	@00046	DAGCd394	1.
@00007	DAGCdQ33	1.	@00047	DAGCd430	1.
@00008	DAGCdQ34	1.	@00048	DAGCd432	1.
@00009	DAGCdQ35	1.	@00049	DAGCdQ7	1.
@00010	DAGCdQ36	1.	@00050	DAGCdQ8	1.
@00011	DAGCdQ37	1.	@00051	DAGCdQ9	1.
@00012	DAGCdQ38	1.	@00052	DAGCdQ10	1.
@00013	DAGCdQ39	1.	@00053	DAGCdQ11	1.
@00014	DAGCdQ77	1.	@00054	DAGCdQ12	1.
@00015	DAGCdQ79	1.	@00055	DAGCdQ40	1.
@00016	DAGCdQ80	1.	@00056	DAGCdQ41	1.
@00017	DAGCdQ81	1.	@00057	DAGCdQ42	1.
@00018	DAGCdQ82	1.	@00058	DAGCdQ43	1.
@00019	DAGCdQ83	1.	@00059	DAGCdQ44	1.
@00020	DAGCdQ84	1.	@00060	DAGCdQ45	1.
@00021	DAGCd140	1.	@00061	DAGCdQ46	1.
@00022	DAGCd141	1.	@00062	DAGCdQ47	1.
@00023	DAGCd142	1.	@00063	DAGCdQ89	1.
@00024	DAGCd143	1.	@00064	DAGCdQ91	1.
@00025	DAGCd144	1.	@00065	DAGCdQ92	1.
@00026	DAGCd145	1.	@00066	DAGCdQ93	1.
@00027	DAGCd146	1.	@00067	DAGCdQ94	1.
@00028	DAGCd202	1.	@00068	DAGCdQ95	1.
@00029	DAGCd204	1.	@00069	DAGCdQ96	1.
@00030	DAGCd205	1.	@00070	DAGCd150	1.
@00031	DAGCd206	1.	@00071	DAGCd151	1.
@00032	DAGCd207	1.	@00072	DAGCd152	1.
@00033	DAGCd261	1.	@00073	DAGCd153	1.
@00034	DAGCd262	1.	@00074	DAGCd154	1.
@00035	DAGCd263	1.	@00075	DAGCd155	1.
@00036	DAGCd310	1.	@00076	DAGCd156	1.
@00037	DAGCd311	1.	@00077	DAGCd157	1.
@00038	DAGCd312	1.	@00078	DAGCd158	1.
@00039	DAGCd352	1.	@00079	DAGCd159	1.
@00040	DAGCd353	1.	@00080	DAGCd160	1.

@00081	DAGCd212	1.	@00140	DAGCd282	1.
@00082	DAGCd213	1.	@00141	DAGCd283	1.
@00083	DAGCd215	1.	@00142	DAGCd284	1.
@00084	DAGCd216	1.	@00143	DAGCd302	1.
@00085	DAGCd217	1.	@00144	DAGCd327	1.
@00086	DAGCd218	1.	@00145	DAGCd328	1.
@00087	DAGCd272	1.	@00146	DAGCd368	1.
@00088	DAGCd273	1.	@00147	DAGCd369	1.
@00089	DAGCd274	1.	@00148	DAGCd403	1.
@00090	DAGCd318	1.	@00149	DAGCd404	1.
@00091	DAGCd320	1.	@00150	DAGCd405	1.
@00092	DAGCd321	1.	@00151	DAGCd406	1.
@00093	DAGCd359	1.	@00152	DAGCd407	1.
@00094	DAGCd360	1.	@00153	DAGCd408	1.
@00095	DAGCd361	1.	@00154	DAGCd409	1.
@00096	DAGCd397	1.	@00155	DAGCd410	1.
@00097	DAGCd398	1.	@00156	DAGCd442	1.
@00098	DAGCd399	1.	@00157	DAGCd443	1.
@00099	DAGCd400	1.	@00158	DAGCd444	1.
@00100	DAGCd401	1.	@00159	DAGCdQ19	1.
@00101	DAGCd437	1.	@00160	DAGCdQ20	1.
@00102	DAGCd438	1.	@00161	DAGCdQ21	1.
@00103	DAGCdQ13	1.	@00162	DAGCdQ22	1.
@00104	DAGCdQ14	1.	@00163	DAGCdQ23	1.
@00105	DAGCdQ15	1.	@00164	DAGCdQ24	1.
@00106	DAGCdQ16	1.	@00165	DAGCdQ56	1.
@00107	DAGCdQ17	1.	@00166	DAGCdQ57	1.
@00108	DAGCdQ18	1.	@00167	DAGCdQ58	1.
@00109	DAGCdQ48	1.	@00168	DAGCdQ59	1.
@00110	DAGCdQ49	1.	@00169	DAGCdQ60	1.
@00111	DAGCdQ50	1.	@00170	DAGCdQ61	1.
@00112	DAGCdQ51	1.	@00171	DAGCdQ62	1.
@00113	DAGCdQ52	1.	@00172	DAGCdQ63	1.
@00114	DAGCdQ53	1.	@00173	DAGCd113	1.
@00115	DAGCdQ54	1.	@00174	DAGCd114	1.
@00116	DAGCdQ55	1.	@00175	DAGCd115	1.
@00117	DAGCd101	1.	@00176	DAGCd116	1.
@00118	DAGCd102	1.	@00177	DAGCd117	1.
@00119	DAGCd103	1.	@00178	DAGCd118	1.
@00120	DAGCd104	1.	@00179	DAGCd119	1.
@00121	DAGCd105	1.	@00180	DAGCd120	1.
@00122	DAGCd106	1.	@00181	DAGCd174	1.
@00123	DAGCd107	1.	@00182	DAGCd175	1.
@00124	DAGCd108	1.	@00183	DAGCd176	1.
@00125	DAGCd166	1.	@00184	DAGCd177	1.
@00126	DAGCd167	1.	@00185	DAGCd178	1.
@00127	DAGCd168	1.	@00186	DAGCd179	1.
@00128	DAGCd169	1.	@00187	DAGCd180	1.
@00129	DAGCd170	1.	@00188	DAGCd181	1.
@00130	DAGCd171	1.	@00189	DAGCd182	1.
@00131	DAGCd224	1.	@00190	DAGCd183	1.
@00132	DAGCd226	1.	@00191	DAGCd235	1.
@00133	DAGCd227	1.	@00192	DAGCd236	1.
@00134	DAGCd228	1.	@00193	DAGCd237	1.
@00135	DAGCd229	1.	@00194	DAGCd238	1.
@00136	DAGCd230	1.	@00195	DAGCd239	1.
@00137	DAGCd231	1.	@00196	DAGCd240	1.
@00138	DAGCd232	1.	@00197	DAGCd241	1.
@00139	DAGCd277	1.	@00198	DAGCd285	1.

@00199	DAGCd287	1.	@00236	DAGCd131	1.
@00200	DAGCd288	1.	@00237	DAGCd132	1.
@00201	DAGCd289	1.	@00238	DAGCd187	1.
@00202	DAGCd290	1.	@00239	DAGCd188	1.
@00203	DAGCd291	1.	@00240	DAGCd190	1.
@00204	DAGCd292	1.	@00241	DAGCd191	1.
@00205	DAGCd333	1.	@00242	DAGCd192	1.
@00206	DAGCd334	1.	@00243	DAGCd193	1.
@00207	DAGCd336	1.	@00244	DAGCd194	1.
@00208	DAGCd337	1.	@00245	DAGCd195	1.
@00209	DAGCd376	1.	@00246	DAGCd196	1.
@00210	DAGCd377	1.	@00247	DAGCd197	1.
@00211	DAGCd413	1.	@00248	DAGCd249	1.
@00212	DAGCd414	1.	@00249	DAGCd250	1.
@00213	DAGCd416	1.	@00250	DAGCd251	1.
@00214	DAGCd417	1.	@00251	DAGCd252	1.
@00215	DAGCd418	1.	@00252	DAGCd298	1.
@00216	DAGCd448	1.	@00253	DAGCd299	1.
@00217	DAGCd450	1.	@00254	DAGCd300	1.
@00218	DAGCdQ25	1.	@00255	DAGCd301	1.
@00219	DAGCdQ26	1.	@00256	DAGCd341	1.
@00220	DAGCdQ27	1.	@00257	DAGCd342	1.
@00221	DAGCdQ28	1.	@00258	DAGCd343	1.
@00222	DAGCdQ29	1.	@00259	DAGCd344	1.
@00223	DAGCdQ30	1.	@00260	DAGCd384	1.
@00224	DAGCdQ65	1.	@00261	DAGCd385	1.
@00225	DAGCdQ66	1.	@00262	DAGCd420	1.
@00226	DAGCdQ67	1.	@00263	DAGCd422	1.
@00227	DAGCdQ68	1.	@00264	DAGCd423	1.
@00228	DAGCdQ69	1.	@00265	DAGCd424	1.
@00229	DAGCdQ70	1.	@00266	DAGCd425	1.
@00230	DAGCdQ71	1.	@00267	FIELDruk	\$12.
@00231	DAGCd126	1.	@00279	PERSruk	\$14.
@00232	DAGCd127	1.	@00293	CHILDID	6.
@00233	DAGCd128	1.	@00299	DWTCWO1C	12. 4
@00234	DAGCd129	1.	@00311	DWTCWO1L	12. 4
@00235	DAGCd130	1.	@00323	DWTCWO1L	12. 4;