

Ce document a été corrigé le 1er avril 2015.
Veuillez consulter l'errata joint à ce fichier
en cliquant [ici](#).

Enquête montréalaise sur l'expérience préscolaire des enfants à la maternelle (EMEP) 2012

RAPPORT MÉTHODOLOGIQUE

Cette publication a été réalisée et produite par l'Institut de la statistique du Québec.

Pour tout renseignement concernant le contenu de cette publication, s'adresser à :

Direction des enquêtes longitudinales et sociales
Institut de la statistique du Québec
1200, avenue McGill College, bureau 500
Montréal (Québec) H3B 4J8
Téléphone : 514 873-4749
Télécopieur : 514 864-9919

Ou

Téléphone : 1 800 463-4090
(aucuns frais d'appel au Canada et aux États-Unis)

Site Web : www.stat.gouv.qc.ca

Citation suggérée :

THIBODEAU K. et L. GINGRAS (2013). *Enquête montréalaise sur le parcours préscolaire des enfants de maternelle. Rapport méthodologique*, Québec, Institut de la statistique du Québec, 24 p.

Nicole Descroisselles, révision linguistique
Institut de la statistique du Québec

© Gouvernement du Québec, Institut de la statistique du Québec, 2013.

Toute reproduction autre qu'à des fins de consultation personnelle est interdite sans l'autorisation du gouvernement du Québec.

Mai 2013

Table des matières

| | |
|--|----|
| Introduction..... | 5 |
| 1. Objectifs..... | 6 |
| 2. Population cible et base de sondage | 7 |
| 3. Plan d'échantillonnage et stratification | 8 |
| 4. Taille et répartition de l'échantillon..... | 9 |
| 5. Collecte..... | 12 |
| 5.1 Le prétest..... | 12 |
| 5.2 La stratégie de collecte..... | 12 |
| 6. Taux d'admissibilité et taux de réponse des deux phases | 13 |
| 7. Traitement des données | 15 |
| 7.1 La validation | 15 |
| 7.2 La pondération | 15 |
| 8. Précision des estimations..... | 17 |
| 9. Portée et limites de l'enquête..... | 21 |
| Références bibliographiques..... | 23 |
| ANNEXE 1 : Calcul de la variance d'échantillonnage..... | 24 |

Liste des tableaux :

| | |
|--|----|
| Tableau 1 : Taille des échantillons des deux phases d'échantillonnage de l'EMEP selon les variables de stratification, Montréal, 2012..... | 11 |
| Tableau 2 : Répartition des élèves de maternelle échantillonnés de la première phase du tirage de l'échantillon de l'EMEP, 2012..... | 13 |
| Tableau 3 : Répartition des élèves de maternelle échantillonnés de la deuxième phase du tirage de l'échantillon de l'EMEP, 2012..... | 14 |
| Tableau 4 : Taux de réponse de la première phase du tirage de l'échantillon de l'EMEP, 2012..... | 14 |
| Tableau 5 : Taux de réponse de la deuxième phase du tirage de l'échantillon de l'EMEP, 2012..... | 14 |

Introduction

L'*Enquête montréalaise sur le parcours préscolaire des enfants de maternelle* (EMEP) a eu lieu au printemps 2012. Elle contient une multitude de renseignements recueillis auprès des parents d'enfants qui résidaient à Montréal et fréquentaient la maternelle 5 ans en 2011-2012. Toutefois, sa grande richesse tient au fait que l'on y trouve également les données tirées de l'EQDEM¹ pour ces enfants. La banque de données ainsi constituée documente de façon exceptionnelle le développement et les expériences préscolaires des enfants montréalais à la maternelle.

Il importe de bien connaître les éléments méthodologiques d'une enquête afin de faire une utilisation adéquate des données. En effet, cette connaissance est essentielle car plusieurs éléments ont une incidence directe sur les estimations produites : plan de sondage de l'enquête, sous-populations visées par les indicateurs, taille et répartition de l'échantillon sélectionné, procédure de collecte en général, validation et codification des données, stratégie de pondération, etc.

Ce rapport présente la méthodologie générale de l'EMEP. La première section présente les objectifs de cette enquête, mais aussi ceux de l'*Enquête québécoise sur le développement des enfants à la maternelle* à laquelle elle est appariée. Viennent ensuite les sections traitant des éléments liés au plan d'échantillonnage utilisé pour l'enquête : population visée, base de sondage, plan d'échantillonnage, taille et répartition de l'échantillon. Le processus de collecte des données et les résultats obtenus (admissibilité et taux de réponse) sont ensuite décrits. Les sections suivantes portent sur le traitement des données : validation et pondération des données recueillies. Finalement, la précision des données recueillies, les méthodes d'estimation de la variance et la portée et les limites de l'enquête sont abordées.

1. L'*Enquête québécoise sur le développement des enfants à la maternelle* a également été réalisée au printemps 2012 par l'Institut de la statistique du Québec.

1. Objectifs

L'Institut de la statistique du Québec (ISQ) a été mandaté par la Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal (DSP-Montréal) pour réaliser, en 2012, l'EMEP. Les données de cette enquête sont appariées aux données montréalaises de l'*Enquête québécoise sur le développement des enfants à la maternelle 2012* (EQDEM), aussi réalisée par l'Institut.

Les objectifs spécifiques de l'EMEP sont les suivants :

- Documenter le parcours préscolaire des enfants montréalais fréquentant une maternelle sur l'île de Montréal. Le parcours préscolaire comprend l'ensemble des services éducatifs et de loisirs utilisés par l'enfant : services de garde, prématernelle, activités artistiques ou physiques, ateliers de stimulation, programme de littératie, etc.
- Identifier les principaux facteurs favorables et défavorables (obstacles, inadéquation entre besoins des parents et services offerts, raisons personnelles, etc.) pour l'utilisation des services à la petite enfance et les raisons de non-fréquentation des services de garde.
- Documenter les caractéristiques socioéconomiques et d'emploi des parents.
- Documenter les conditions dans lesquelles vivent les familles et le soutien social dont elles bénéficient.
- Documenter les ressources et le niveau de sécurité du quartier.
- Examiner plus précisément la situation des enfants défavorisés.
- Permettre l'arrimage des données de l'enquête avec celles de l'EQDEM afin de mettre en lien l'expérience préscolaire des enfants et leurs conditions de vie avec différents aspects de leur développement.

Les objectifs de l'EQDEM sont de :

- Recueillir les données permettant de dresser un portrait de l'état de développement de tous les élèves québécois de maternelle selon les cinq dimensions mesurées par l'*Instrument de mesure du développement de la petite enfance* (IMDPE), à savoir :
 - La santé physique et le bien-être;

- Les compétences sociales;
- La maturité affective;
- Le développement cognitif et langagier;
- Les habiletés de communication et les connaissances générales.
- Décrire quelques caractéristiques de l'enfant et, brièvement, ses antécédents quant aux services éducatifs préscolaires utilisés (fréquentation d'un service de garde et de la maternelle 4 ans).
- Obtenir des mesures du développement à l'échelle régionale.

L'échantillon de l'EMEP a été sélectionné parmi les répondants de l'EQDEM satisfaisant aux deux critères suivants : la résidence et la fréquentation d'un établissement sur l'île de Montréal.

2. Population cible et base de sondage

La population visée par l'enquête est composée de tous les enfants fréquentant la maternelle 5 ans sur le territoire de l'île de Montréal en 2011-2012, résidant sur l'île de Montréal et ayant participé à l'EQDEM. Les enfants reconnus comme étant des élèves handicapés ou des élèves en difficulté d'adaptation ou d'apprentissage (EHDAA) sont exclus.

La base de sondage utilisée pour sélectionner l'échantillon de l'enquête provient de l'entrepôt de données du ministère de l'Éducation, du Loisir et du Sport (MELS) qui est alimenté en continu par le système Charlemagne. Ce fichier présente plusieurs avantages; il permet notamment d'exclure la majeure partie des personnes non visées par l'enquête tout en assurant une excellente couverture des individus ciblés. Le code postal de la résidence de l'élève ainsi que la région associée à l'école qu'il fréquente ont été utilisés pour déterminer l'admissibilité. Les fichiers du MELS sont mis à jour régulièrement, ce qui garantit une bonne qualité de la base de sondage. Cependant, comme tout fichier administratif utilisé comme base de sondage, l'entrepôt de données du MELS comporte quelques faiblesses.

Malgré la mise à jour régulière des fichiers du MELS, les coordonnées de l'élève sont parfois inexactes ou même inexistantes, compliquant du même coup le contact et la collecte auprès de la population échantillonnée².

Par ailleurs, la base de sondage peut engendrer une sous-couverture puisque certaines personnes admissibles à l'enquête ne sont pas inscrites dans le fichier du MELS. De même, la base de sondage pourrait compter des personnes non admissibles³, par exemple, au moment du tirage de l'échantillon, celles n'ayant pas encore informé le MELS d'un déménagement. Il est toutefois difficile de quantifier de manière précise la couverture de la base de sondage étant donné que la taille de la population visée ne peut être établie de façon exacte. On estime, malgré tout, que la sous-couverture est faible.

Ainsi, la population échantillonnée (population parmi laquelle on peut tirer un échantillon) pourrait différer quelque peu de la population visée. On peut toutefois faire l'hypothèse que la population non couverte par l'enquête ressemble à celle qui l'est quant à ses caractéristiques; dans ce cas, l'erreur de couverture aurait un impact négligeable sur les estimations produites. On peut donc conclure que les résultats peuvent être inférés à l'ensemble de la population visée par l'enquête.

3. Plan d'échantillonnage et stratification

Étant donné que le tirage de l'échantillon de l'EMEP a dû s'effectuer avant même que la collecte de l'EQDEM soit terminée, nous avons prévu un échantillonnage à deux phases. La première phase consistait à tirer un échantillon, parmi les élèves fréquentant l'école et résidant à Montréal, dans le fichier du MELS en date du mois de décembre 2011. La seconde phase se résumait à tirer un échantillon parmi les répondants de l'EQDEM du premier échantillon.

2. À titre indicatif, il y a eu environ 5 % de numéros de téléphone manquants et environ 3 % d'adresses manquantes.

3. Il peut y avoir un certain délai dans la mise à jour de l'information; entre-temps, un enfant peut avoir changé d'école ou de lieu de résidence sans en avoir informé le MELS.

Le tirage de l'échantillon s'est fait selon une stratification de la base de sondage visant à assurer un certain contrôle de la composition de l'échantillon et à respecter les objectifs de l'enquête. Les strates sont formées du croisement entre l'indice de défavorisation matérielle⁴, la langue d'enseignement de l'école et le statut privé/public de l'école. Les strates de défavorisation ont été construites en fonction de l'environnement socioéconomique dans lequel vivent les enfants, sur la base du code postal du lieu de résidence de l'enfant. Ceux vivant dans les milieux défavorisés définis selon le quintile inférieur (le 5^e) de l'indice de défavorisation matérielle (IDM) ont été suréchantillonnés. Le groupe des autres enfants se classe dans les quintiles 1 à 4 de l'IDM. Cet indicateur de défavorisation comportant deux catégories (« défavorisés » et « autres ») a été attribué à chaque enfant. Pour les deux autres variables de stratification, on a tenu compte de la répartition de la population des enfants selon le statut (public ou privé) et la langue d'enseignement de l'école.

Précisons que, tout au long de la collecte, une mise à jour du statut de répondant EQDEM de chaque élève s'effectuait et permettait de mieux suivre les taux de réponse de l'EQDEM selon les différentes strates.

4. Taille et répartition de l'échantillon

La taille initiale de l'échantillon a été fixée à 2 000 unités en supposant un taux de réponse de 60 % pour la collecte de l'EMEP. Le nombre total de répondants nécessaire pour atteindre les objectifs mentionnés précédemment a été fixé à 1 200. Un tel échantillon permet d'obtenir une marge d'erreur d'au plus 3 %, avec des coefficients de variation acceptables (moins de 15 %) pour toutes les proportions supérieures à 5 %. En faisant des hypothèses quant aux taux de réponse et d'admissibilité, une taille d'échantillon initiale a pu être déterminée pour chacune des strates de l'enquête. Pour obtenir une bonne précision à l'intérieur de la catégorie des « défavorisés », un suréchantillonnage a été requis; celui-ci a été établi de façon à viser 40 % de

4. Voir PAMPALON et autres, 2004.

répondants de la catégorie des « défavorisés » permettant d'obtenir de l'information sur 480 enfants.

Comme nous l'avons déjà mentionné, l'échantillonnage s'est fait en deux phases. La première consistait à tirer un échantillon de 4 000 élèves parmi ceux fréquentant l'école et résidant à Montréal dans le fichier du MELS en date du mois de décembre 2011. Cet échantillon a été tiré avant que l'EQDEM soit terminée. La seconde phase se résumait à tirer un échantillon parmi les répondants de l'EQDEM du premier échantillon (une fois ceux-ci connus) afin de le réajuster selon la stratification prévue.

On a procédé à un tirage systématique du nombre de personnes déterminé par le plan de stratification pour atteindre les objectifs de précision du plan de sondage, tout en tenant compte du fait que le groupe des « défavorisés » affichait un taux de réponse plus faible que le groupe des « autres » lors du prétest de l'EMEP. Le tableau 1 présente les tailles d'échantillon des deux phases. Étant donné qu'il n'y avait pas d'objectif à atteindre au sujet de la langue d'enseignement et qu'il y avait un nombre insuffisant d'unités dans un groupe (groupe « défavorisés » – école privée – langue d'enseignement anglaise), les deux langues ont été regroupées dans une seule strate (groupe « défavorisés » – école privée – les deux langues). Le découpage par langue avait pour but de suivre les taux de réponse lors de la collecte et d'ajuster en conséquence pour la deuxième phase s'il y avait lieu de le faire.

Tableau 1

Taille des échantillons des deux phases d'échantillonnage de l'EMEP selon les variables de stratification, Montréal, 2012

| Strate | 1 ^{re} phase | 2 ^e phase |
|---|-----------------------|----------------------|
| 1 : groupe « autres »/école privée/français comme langue d'enseignement | 427 | 157 |
| 2 : groupe « autres »/école privée/anglais comme langue d'enseignement | 136 | 44 |
| 3 : groupe « autres »/école publique/français comme langue d'enseignement | 1465 | 744 |
| 4 : groupe « autres »/école publique/anglais comme langue d'enseignement | 372 | 198 |
| 5 : groupe « défavorisés »/école privée/anglais et français comme langue d'enseignement | 99 | 41 |
| 6 : groupe « défavorisés »/école publique/français comme langue d'enseignement | 1 355 | 737 |
| 7 : groupe « défavorisés »/école publique/anglais comme langue d'enseignement | 146 | 79 |
| Total | 4 000 | 2 000 |

On peut évaluer l'effet du plan de sondage de l'EMEP sur les différentes estimations faites à partir de cette enquête, c'est-à-dire estimer l'incidence de la non-proportionnalité de la répartition sur la précision des résultats. L'effet de plan illustre donc l'efficacité d'une taille d'échantillon⁵. Ainsi, dans la présente enquête, lorsque les estimations sont produites séparément pour les deux groupes de l'indice de défavorisation matérielle, l'effet de plan moyen⁶ est estimé à 1,06 pour celui des « autres » et à 1,18 pour celui des « défavorisés ». Pour l'ensemble de la population visée, l'effet de plan moyen est d'environ 1,17, soit un effet de plan assez faible.

5. Par exemple, un effet de plan égal à 2 signifie que la même précision d'un résultat pourrait être obtenue avec un échantillon aléatoire simple deux fois moins grand.

6. Puisque l'effet de plan varie selon la variable d'analyse, pour calculer un effet de plan moyen, il faut procéder à plusieurs estimations et calculer la moyenne des effets de plan obtenus pour chacune des variables.

5. Collecte

5.1 Le prétest

Au début du printemps 2012, soit du 6 au 20 mars, un prétest a été effectué auprès de 168 familles résidant sur l'île de Montréal et ayant un enfant en première année dans une école, privée ou publique, de l'île de Montréal. Le choix des élèves de première année plutôt que ceux de la maternelle avait pour but d'éviter toute nuisance à l'EQDEM qui était alors commencée.

Les objectifs de ce prétest étaient, notamment, de vérifier la compréhension du questionnaire en français et en anglais, d'évaluer la durée du questionnaire et d'estimer les taux de réponse et de refus de l'enquête.

Quelques recommandations ont été émises à la suite de ce prétest. Il a été notamment suggéré de raccourcir ou d'éclaircir certaines questions, et d'envisager une réduction du nombre de questions afin de diminuer le fardeau des répondants. Retenons que, dans l'ensemble, les principales notions du questionnaire ont été bien comprises par les répondants.

5.2 La stratégie de collecte

La collecte des données de l'EMEP s'est déroulée du 25 avril au 2 juillet 2012. Elle a été réalisée au moyen d'entrevues téléphoniques assistées par ordinateur (ITAO) à partir du logiciel Command Center (CC3). La durée moyenne de l'entrevue a été d'un peu plus de 26 minutes. La stratégie de collecte retenue comportait l'expédition d'une lettre de présentation de l'enquête à la personne sélectionnée; cette lettre l'informait aussi qu'elle serait contactée par téléphone pour participer.

La stratégie de collecte prévoyait aussi, en cas d'absence ou de refus, plusieurs tentatives permettant de joindre les personnes ou de les convaincre de participer à l'EMEP⁷. Quant aux personnes pour lesquelles le MELS n'avait pas de numéros de téléphone (environ 5 % des personnes échantillonnées), des recherches manuelles ont été effectuées afin de trouver des

7. À l'exception des cas de refus ferme qui n'étaient pas recontactés.

numéros valides; des lettres ont été adressées à ces personnes en leur demandant de contacter l'Institut.

6. Taux d'admissibilité et taux de réponse des deux phases

Tous les EHDAAs sont exclus de l'EMEP, y compris ceux qui sont en classe régulière. Toutefois, le statut de ces élèves n'a été connu qu'après la collecte de données. Ils ont donc été retranchés de l'échantillon après coup. Puisque l'échantillonnage a été fait en deux phases, les taux d'admissibilité et de réponse ont été calculés pour chacune d'elles.

Lors de la première phase du tirage de l'échantillon, 4 000 élèves ont été sélectionnés parmi 18 140 unités visées. De ces 4 000 élèves, 94 ont été déclarés inadmissibles : 43 EHDAAs, 6 élèves n'habitant pas Montréal ou n'y fréquentant pas une école et 45 élèves ayant quitté l'école et n'ayant pas été retracés. Cela conduit à un taux d'admissibilité de 97,7 % pour cette phase. Toutefois, parmi les élèves admissibles, 2 983 ont participé à l'EQDEM, duquel sont sélectionnés les participants de la seconde phase d'échantillonnage.

Tableau 2

Répartition des élèves de maternelle échantillonnés de la première phase du tirage de l'échantillon de l'EMEP, 2012

| | n |
|-------------------------------------|--------------|
| Répondants de l'EQDEM : | |
| Défavorisés | 1 208 |
| Autres | 1 775 |
| Total | 2 983 |
| Non-répondants de l'EQDEM | 923 |
| Total des élèves admissibles | 3 906 |
| Élèves inadmissibles | 94 |
| Échantillon de départ | 4 000 |

Lors de la deuxième phase, on a sélectionné 2 000 élèves parmi les 2 983 répondants de l'EQDEM. Une fois l'admissibilité de la deuxième phase traitée, l'échantillon a été ramené à

1 985 élèves dû à des cas d'inadmissibilité qui se sont ajoutés en cours de route. Parmi ces 1 985 unités préalablement sélectionnées, 35 étaient des EHDA (cas inadmissibles), ce qui porte le taux d'admissibilité à 98,2 %.

Tableau 3

Répartition des élèves de maternelle échantillonnés de la deuxième phase du tirage de l'échantillon de l'EMEP, 2012

| | n |
|-------------------------------------|--------------------------|
| Répondants de l'EMEP | 1184 |
| Non-répondants | 766 |
| Total des élèves admissibles | 1950 |
| Élèves inadmissibles | 50 |
| Total | 2 000¹ |

1. Les unités échantillonnées parmi les 2 983 répondants de l'EQDEM.

L'objectif d'un taux de réponse de 60 % a été atteint et les taux des deux groupes (« autres » et « défavorisés ») sont similaires (tableaux 4 et 5).

Tableau 4

Taux de réponse de la première phase du tirage de l'échantillon de l'EMEP, 2012

| | n | Taux de réponse pondéré (%) |
|--------------------------------|--------------|-----------------------------|
| Répondants de l'EQDEM : | | |
| Défavorisés | 1 208 | 77,8 |
| Autres | 1 775 | 74,1 |
| Total | 2 983 | 74,9 |

Tableau 5

Taux de réponse de la deuxième phase du tirage de l'échantillon de l'EMEP, 2012

| | n | Taux de réponse pondéré (%) |
|-------------------------------|--------------|-----------------------------|
| Répondants de l'EMEP : | | |
| Défavorisés | 695 | 61,4 |
| Autres | 489 | 61,4 |
| Total | 1 184 | 61,4 |

7. Traitement des données

7.1 La validation

Le recours aux entrevues téléphoniques assistées par ordinateur (ITAO) dans l'EMEP permet une saisie automatique par l'intervieweur des réponses fournies par les répondants. En effet, le questionnaire est entièrement programmé. De plus, les questions sont en très grande majorité autocodées, c'est-à-dire que l'intervieweur n'a qu'à cocher le choix du répondant; la saisie s'en trouve donc réduite au minimum. Par ailleurs, le questionnaire informatisé comprend plusieurs validations de base qui s'effectuent pendant l'entrevue, notamment la vérification du respect des intervalles de réponse pour les questions fermées et celle de l'adéquation des sauts de section pour les blocs de questions ne s'adressant qu'à un sous-groupe de répondants.

Les données ont également fait l'objet d'une validation *a posteriori* où on a principalement vérifié la cohérence des réponses d'un même répondant, ce qui a permis de relever quelques erreurs.

Enfin, comme cette enquête vise à produire des estimations selon deux groupes de défavorisation matérielle, on a aussi procédé à une validation minutieuse des codes postaux fournis par les répondants afin de s'assurer que chaque répondant était classé dans le bon groupe.

7.2 La pondération

La pondération consiste à attribuer à chaque répondant une valeur, c'est-à-dire un poids, qui correspond au nombre de personnes qu'il « représente » dans la population. La pondération permet de rapporter les données des répondants à la population visée et de faire ainsi des inférences adéquates à cette population, bien que cette dernière n'ait pas été observée dans sa totalité. C'est également un des éléments à considérer pour estimer correctement la précision

des données. À titre d'exemple, la répartition de l'échantillon selon les strates « défavorisés » et « autres » ne représente pas bien celle de la population de la région de Montréal. La pondération permet de corriger cette distorsion créée par le plan de sondage. De plus, il est connu que, dans ce type d'enquête, la probabilité de répondre varie selon certaines caractéristiques sociodémographiques. Il est donc nécessaire de tenir compte de ces éléments, qui devraient, autant que possible, être intégrés à la pondération utilisée pour la production de statistiques tirées de l'EMEP.

La stratégie de pondération qui a été développée par l'Institut pour cette enquête tient compte, entre autres, de la probabilité de sélectionner une unité dans la base de sondage, des élèves inadmissibles, de l'ajustement de la non-réponse totale à l'EQDEM et à l'EMEP de même que d'un ajustement aux totaux de population en tenant compte des caractéristiques suivantes : le groupe de défavorisation matérielle, le sexe et l'indice de vulnérabilité produit dans l'EQDEM⁸.

La première étape consiste à calculer l'inverse de la probabilité de sélection de la première phase d'échantillonnage de chaque élève. Ensuite, un ajustement doit être fait pour l'admissibilité à l'enquête (selon les critères de la population visée).

L'étape suivante est l'ajustement pour la non-réponse totale à l'EQDEM. Il s'agit donc de modéliser le fait d'être répondant ou non à cette enquête et de créer des groupes de pondération correspondant à des regroupements d'élèves ayant des caractéristiques et des proportions semblables⁹. Le poids des répondants est ajusté par l'inverse du taux de réponse dans chacun des groupes.

La méthode du score a été utilisée pour déterminer ces groupes. Cette méthode consiste à faire une régression logistique pour modéliser une variable dichotomique indiquant la présence ou l'absence de réponse selon diverses variables. Les valeurs prédites de cette régression logistique

8. Cet indice permet d'identifier les élèves ayant au moins une vulnérabilité quant aux cinq dimensions étudiées dans l'EQDEM. La vulnérabilité est définie selon un critère empirique en considérant les enfants vulnérables comme étant ceux dont le score pour une dimension donnée se situe sur ou sous le seuil correspondant au 10^e centile de l'ensemble des enfants à la maternelle au Québec en 2011-2012.

9. Les variables retenues dans le premier modèle sont : la commission scolaire, la proportion des EHDAAs dans l'école et le statut allophone de l'élève.

sont ensuite regroupées au moyen de méthodes de classification afin de constituer les groupes de pondération¹⁰.

Un ajustement a été appliqué pour tenir compte de la seconde sélection (échantillon de 2 000 unités). Puis, on a procédé à un ajustement pour la non-réponse totale à l'EMEP sur le même principe que celui effectué dans l'EQDEM pour la non-réponse totale¹¹. Mentionnons qu'un faible taux de réponse a été observé dans le groupe des élèves provenant du secteur privé avec l'anglais comme langue d'enseignement et l'anglais comme langue maternelle, soit 30,8 %. Il faut donc être prudent dans l'interprétation des résultats car les données associées à ce groupe sont de moins bonne qualité.

Enfin, la dernière étape de la pondération est la poststratification. Cet ajustement permet de s'assurer que la somme des poids de toutes les personnes répondantes de l'enquête correspond bien à la population visée par l'enquête. Il s'agit de découper la population en groupes ayant des caractéristiques semblables. La population de référence utilisée est un extrait de la base de sondage de l'EQDEM ajusté¹² pour tenir compte de la définition de la population visée de l'EMEP. La poststratification a été réalisée par l'application d'un facteur pour obtenir une somme de poids égale aux effectifs pondérés de la population visée (base de sondage de l'EQDEM des élèves fréquentant une école à Montréal et résidant à Montréal) par l'indice de défavorisation matérielle en quartiles, le sexe et l'indice de vulnérabilité de l'EQDEM.

8. Précision des estimations

Dans toute enquête statistique, les estimations produites sont entachées d'erreurs. Il y a, d'une part, des erreurs liées à l'échantillonnage, c'est-à-dire celles qui sont attribuables au fait que l'on interroge seulement une partie de la population visée, et d'autre part, celles qu'entraînent

10. Pour plus de détails sur la méthode du score, voir HAZIZA et BEAUMONT, 2007 et ELTINGE et YANSANEH, 1997.

11. Les variables retenues dans le second modèle sont : la commission scolaire, la langue d'enseignement de l'école, la langue maternelle et l'estimation de l'indice de vulnérabilité de l'EQDEM.

12. Pour plus de détails à ce sujet, consulter le rapport de l'EQDEM, qui sera publié à l'automne 2014 et disponible sur le site de l'Institut : <http://www.stat.gouv.qc.ca>.

la non-réponse ou la sous-couverture affectant la base de sondage, les erreurs de mesure associées à des questions difficiles à interpréter et les erreurs liées à la saisie ou au traitement des données. Nous présentons ici les erreurs liées à l'échantillonnage.

Lorsque l'on infère les résultats obtenus à partir d'un échantillon à la population visée, on ne peut parler de valeurs exactes, mais bien d'une estimation de ces valeurs exactes. Il s'avère donc nécessaire d'estimer la précision de chaque estimation de proportion issue de l'échantillon. La complexité du plan d'échantillonnage de l'EMEP influence la précision des estimations.

Puisque l'impact n'est pas négligeable, il doit être pris en compte dans l'estimation de la précision. Certains logiciels d'analyse statistique, tels que SPSS et SAS, incluent des procédures qui permettent d'obtenir des estimations sans biais de la précision pour certains plans de sondage complexes¹³, comme celui utilisé dans l'EMEP. La variance peut être estimée à l'aide d'une méthode de linéarisation, aussi appelée approximation par séries de Taylor. Pour utiliser cette méthode, les indicateurs de strates et d'unités primaires d'échantillonnage (UPE) sont nécessaires. Dans le cas de l'EMEP, il n'y a qu'un seul degré d'échantillonnage car l'UPE est équivalente à l'individu répondant. Seul l'indicateur de strate doit être spécifié.

D'autres logiciels, par exemple STATA (version 12) ou SUDAAN peuvent aussi être utilisés aux mêmes fins. Ces logiciels ont l'avantage de permettre l'utilisation de plusieurs méthodes d'estimation de la variance qui tiennent compte de la complexité du plan de sondage : méthode de linéarisation et méthode de ré-échantillonnage (ex. : méthode du *bootstrap*)¹⁴. Enfin, le logiciel WESVAR (version 4.2) permet d'utiliser la méthode du *bootstrap*.

Pour utiliser la méthode du *bootstrap*, il faut recourir aux ensembles de poids *bootstrap* fournis avec le fichier de microdonnées de l'enquête, puisque ce sont ceux-là mêmes qui reflètent la façon dont l'échantillon a été sélectionné. Par ailleurs, contrairement à la méthode de linéarisation, la méthode du *bootstrap* permet de tenir compte de tous les ajustements

13. Dans la version 9 de SAS, ces procédures sont incluses dans le module de base, alors que pour SPSS (version 21), il faut se procurer le module complémentaire *Complex samples* pour y avoir accès.

14. Le lecteur trouvera une description détaillée des méthodes de rééchantillonnage dans les ouvrages de LOHR (1999), RUST et RAO (1996) et WOLTER (1985).

apportés à la pondération. En ce sens, la première doit être considérée comme une méthode plus approximative pour l'estimation de la variance tandis que la seconde est plus précise.

Selon le type de fichier utilisé, on recourra à des méthodes d'estimation spécifiques. Dans le cas du fichier maître de l'EMEP, disponible au Centre d'accès aux données de recherche de l'Institut de la statistique du Québec (CADRISQ), les poids *bootstrap* sont inclus. Lorsqu'on utilise le fichier masqué (FMII), on doit plutôt recourir à la variable de stratification qui est incluse dans le fichier.

Brièvement, la méthode du *bootstrap* fonctionne de la manière suivante. Pour obtenir une estimation de la variance associée à l'estimation d'une proportion (statistique calculée à partir du poids de sondage), il suffit de recalculer cette même estimation de la proportion un grand nombre de fois en utilisant chacun des ensembles de poids *bootstrap* fournis avec les données de l'enquête. Cette opération donne une distribution de valeurs possibles pour l'estimateur de la proportion. La variabilité observée entre les différentes proportions estimées permet d'estimer la variance ou d'autres mesures de précision. Dans l'EMEP, 2 000 ensembles de poids *bootstrap* ont été produits.

En pratique, voici les principales étapes de l'estimation de la variance associée à une estimation à l'aide de la méthode du *bootstrap* :

- a) L'estimation ponctuelle du paramètre d'intérêt (proportion, moyenne, etc.) est d'abord calculée en utilisant le poids de sondage inclus dans le fichier de données;
- b) La même estimation ponctuelle est calculée en utilisant cette fois chacun des 2 000 ensembles de poids *bootstrap* contenus dans le fichier. Deux mille estimations *bootstrap* du paramètre d'intérêt sont ainsi obtenues;
- c) Finalement, la variance (selon la formule utilisée pour un plan aléatoire simple) de ces 2 000 estimations *bootstrap* est calculée. Cette variance correspond à l'estimation de la variance de l'estimation¹⁵ du paramètre d'intérêt calculée en a).

Pour estimer la précision, on utilise la marge d'erreur et le coefficient de variation.

15. Pour plus de détails sur le calcul de la variance de l'estimation, voir l'annexe.

La marge d'erreur est une mesure de la précision statistique d'une estimation. À un niveau de confiance de 95 %, elle se calcule ainsi :

$$me = 1,96 \times \sqrt{\text{variance de l'estimation}}$$

À partir de cette marge d'erreur, on peut calculer un intervalle de confiance (IC) à 95 % qui est associé à l'estimation :

$$IC = \text{estimation} \pm me$$

Cet intervalle indique l'étendue des valeurs possibles que peut prendre la variable étudiée dans la population observée.

Lorsque l'approximation par la loi normale n'est pas possible, c'est-à-dire dans les cas où la proportion est trop petite, compte tenu de la taille efficace associée, l'intervalle de confiance basé sur la loi normale doit être remplacé par un intervalle de confiance basé sur la loi binomiale.

Le coefficient de variation (CV) est une autre mesure permettant de mesurer la précision relative d'une estimation. Le CV s'exprime comme suit :

$$CV = \frac{\sqrt{\text{variance de l'estimation}}}{\text{estimation}}.$$

Plus le CV est élevé moins l'estimation est précise, tandis qu'à l'inverse, plus le CV est petit meilleure est la précision. Lorsqu'une estimation obtient un CV inférieur à 15 %, la précision est jugée bonne. Une estimation dont le CV se situe entre 15 % et 25 % a une précision passable. Enfin, lorsque le CV est supérieur à 25 %, l'estimation est de faible précision; elle n'est fournie qu'à titre indicatif et devrait être utilisée avec circonspection. Les résultats publiés à partir de l'enquête devraient toujours être accompagnés de cette mesure de précision.

Pour vérifier l'atteinte des objectifs de précision, les effets de plan auxquels sont soumises les différentes estimations produites à partir de cette enquête ont été examinés. Le fait que ces effets de plan sont en général très près des hypothèses de départ (voir section 4) et que les nombres de répondants visés par le plan de sondage dans les groupes de défavorisation ont été atteints presque partout, montre que l'enquête a atteint ses objectifs de précision fixés. Ainsi, la majorité des estimations produites ont une précision semblable à celle qui était prévue dans le plan de sondage.

9. Portée et limites de l'enquête

La principale force de cette enquête est la possibilité d'intégrer les données de l'EQDEM sur le développement des enfants au moment où ils sont à la maternelle avec celles de l'EMEP sur leur parcours avant la maternelle. Par contre, le fait d'avoir deux phases et donc de la non-réponse à chacune des phases engendre une non-réponse totale plus importante. C'est l'information de la première phase qui a fourni une qualité acceptable de données en permettant de bien modéliser la non-réponse de la seconde phase, la non-réponse associée à l'EMEP.

Par ailleurs, pour les utilisateurs des bases de données de l'EQDEM et de l'EMEP, il est important de mentionner que les estimations pour une même variable peuvent différer légèrement. Deux raisons sont à l'origine de ces différences.

Soulignons d'abord le fait que la population visée de l'EMEP est définie par la région de résidence et la région de fréquentation scolaire, tandis que les résultats de l'EQDEM sont produits selon la région de résidence uniquement. De plus, le fait que l'échantillon de l'EMEP est un sous-échantillon de l'EQDEM entraîne des différences dans les estimations même si la pondération a permis de minimiser ces différences. Précisons toutefois que, dans la plupart des cas, les estimations diffèrent très peu entre les deux enquêtes. Également, pour quelques variables de contrôle telles que l'indicateur de défavorisation (basé sur le quartile) et l'indicateur de vulnérabilité de l'EQDEM (le fait d'avoir au moins une vulnérabilité), les estimations sont exactement les mêmes. De plus, d'autres variables telles que la langue

d'enseignement et le réseau d'enseignement de l'école ont été considérées à toutes les étapes de la pondération pour ainsi s'assurer d'obtenir des estimations de qualité.

L'EMEP est une enquête transversale. Elle permet de déceler des associations entre deux variables et des différences entre des groupes de la population (par exemple celui des « défavorisés » et celui des « autres »). Par contre, ce genre d'enquête ne permet pas d'établir des liens de causalité entre les caractéristiques étudiées, ou entre des variables tirées de l'EQDEM et de l'EMEP.

Néanmoins, cette enquête, dont les résultats sont inférables aux enfants vivant et fréquentant la maternelle sur l'île de Montréal, fournit des données de qualité.

Références bibliographiques

ELTINGE, J. L., et I. S. YANSANEH (1997). "Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey", *Techniques d'enquête*, vol. 23, n° 1, p. 33-40.

HAZIZA, D., et J.-F. BEAUMONT (2007). "On the construction of imputation classes in surveys", *International Statistical Review*, vol. 75, n° 1, p. 25-43.

INSTITUT DE LA STATISTIQUE DU QUÉBEC en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC et le MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX DU QUÉBEC (2010). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé de la population 2008*, Québec, Gouvernement du Québec, 117 p.

LOHR, S. (1999). *Sampling: Design and Analysis*, 1st Ed., Boston, Duxbury Press, 596p.

PAMPALON, R., D. HAMEL et G. RAYMOND (2004). *Indice de défavorisation pour l'étude de la santé et du bien-être au Québec – mise à jour 2001*, Institut national de santé publique du Québec, 11 p. [En ligne]. [<http://www.inspq.qc.ca>] (Consulté le 4 décembre 2012).

PICA, L. A., I. TRAORÉ, F. BERNÈCHE, P. LAPRISE, L. CAZALE, H. CAMIRAND, M. BERTHELOT, N. PLANTE et autres (2012). *L'Enquête québécoise sur la santé des jeunes du secondaire 2010-2011. Le visage des jeunes d'aujourd'hui : leur santé physique et leurs habitudes de vie*, Québec, Institut de la statistique du Québec, Tome 1, 258 p.

RUST, K. F., et J. N. K. RAO (1996). "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research*, vol. 5, n° 3, september, p. 283-310.

STATISTIQUE CANADA (2003). *Méthodes et pratiques d'enquête*, Ottawa, n° 12-587-XPF au catalogue, 434p.

WOLTER, K. M. (1985). *Introduction to Variance Estimation*, New York, Springer-Verlag, 428 p.

ANNEXE 1 : Calcul de la variance d'échantillonnage

La variance d'échantillonnage est la mesure qui est appliquée le plus souvent pour quantifier l'erreur d'échantillonnage (Statistique Canada, 2003). La méthode d'auto-amorçage (*bootstrap*) a été retenue pour estimer cette variance. On s'en sert donc pour mesurer la précision. Cette méthode de rééchantillonnage consiste à sélectionner des sous-échantillons à partir de l'échantillon principal et à produire des estimations pour chacun de ces sous-échantillons. En mesurant la dispersion entre ces différentes estimations, à l'aide de la formule suivante, on obtient une estimation de la variance d'échantillonnage :

$$\hat{\text{Var}}(\hat{B}) = \frac{\sum_{k=1}^n (\hat{B}_k - \hat{B})^2}{n-1}$$

où :

\hat{B} est l'estimation d'une certaine caractéristique B obtenue à partir de l'échantillon principal pour laquelle on désire estimer la variance d'échantillonnage;

n est le nombre de sous-échantillons produits, soit 2 000;

\hat{B}_k est l'estimation d'une certaine caractéristique B obtenue à partir du k^{e} sous-échantillon.

La précision de cet estimateur de la variance d'échantillonnage s'accroît avec le nombre de sous-échantillons considérés. Afin d'obtenir, pour chaque sous-échantillon, une estimation \hat{B}_k pouvant être inférée à la population, comme c'est le cas pour l'estimation \hat{B} , il faut procéder à la pondération de chaque sous-échantillon. On crée ainsi autant de pondérations que de sous-échantillons; ces pondérations sont appelées « poids d'auto-amorçage ». La création de chaque poids d'auto-amorçage se calque sur la façon dont les poids d'échantillonnage ont été obtenus (Rust et Rao, 1996).

Errata

Enquête montréalaise sur l'expérience préscolaire des enfants à la maternelle (EMEP) 2012 – Rapport méthodologique

Document corrigé le 1^{er} avril 2015

1. À la page 14, changer le tableau 5 par celui-ci:

Tableau 5

Taux de réponse de la deuxième phase du tirage de l'échantillon de l'EMEP, 2012

| | n | Taux de réponse pondéré (%) |
|-------------------------------|--------------|-----------------------------|
| Répondants de l'EMEP : | | |
| Défavorisés | 491 | 58,9 |
| Autres | 693 | 62,1 |
| Total | 1 184 | 61,4 |