

Étude longitudinale sur le développement des enfants du Québec (ÉLDEQ 2003-2011)

Pondération des données concernant les tâches cognitives administrées aux enfants au volet 2003

Nathalie Plante et Robert Courtemanche
Direction de la méthodologie, de la démographie et des enquêtes spéciales
Institut de la statistique du Québec
30 mars 2005

1. La non-réponse aux tâches cognitives : état de la situation

1.1 Introduction

Parmi les enfants dont la famille a participé à l'enquête, plusieurs n'ont pas obtenu de résultat à tous les jeux, parce qu'ils ont refusé de participer ou pour d'autres raisons. S'il s'avérait que les enfants pour lesquels les résultats à un jeu sont manquants diffèrent des autres enfants, les estimations produites à partir des données disponibles pourraient être biaisées pour l'ensemble de la population visée. Une analyse de la non-réponse aux jeux permet d'évaluer si les enfants répondants diffèrent de façon importante des enfants non-répondants. Le cas échéant, on tente de minimiser les biais potentiels à l'aide d'une pondération appropriée des données des répondants.

Le présent texte dresse un bilan de l'analyse de la non-réponse effectuée pour les données concernant les tâches cognitives du volet 2003. À la suite d'une brève présentation portant sur la non-réponse totale, nous examinons, à la section 1, la non-réponse partielle pour chacun des jeux. La section 2 porte sur la stratégie de pondération retenue, notamment le taux de réponse obtenu ainsi que la démarche et les résultats de la modélisation de la non-réponse pour l'ensemble des jeux. Finalement, la section 3 renseigne l'utilisateur sur le contenu du fichier de pondération des jeux du volet 2003 ainsi que sur certains aspects à considérer pour les analyses statistiques. Aux seules fins d'analyse, l'utilisateur peut s'en tenir à la lecture des sections 1 et 3 tout en sélectionnant les sous-sections se rapportant au jeu pour lequel les données sont analysées.

1.2 Non-réponse totale

Parmi les 1 759 répondants au questionnaire informatisé rempli par l'intervieweur (QIRI) du volet 2003, 997 enfants ont participé aux tâches cognitives de ce volet, en tout ou en partie¹. On compte par ailleurs 15 enfants ayant participé aux tâches cognitives sans pour autant avoir été répondants au QIRI. Règle générale, les enfants qui ont participé à un jeu ont aussi participé à l'ensemble des jeux, aussi la non-réponse partielle est-elle faible pour chacun des jeux. Nous considérons ici qu'un

¹ La proportion non-pondérée de répondants aux tâches cognitives, parmi les répondants au QIRI, est de 57 % pour le volet 2003. Une proportion pondérée ainsi qu'un taux de réponse pondéré relatif à l'échantillon initial sera présenté à la section 2.3 en tenant compte de la stratégie de pondération adoptée.

enfant est répondant global s'il a participé à au moins une des cinq tâches proposées et que le questionnaire a été dans ce cas bien rempli par l'intervieweur. Il est important de noter que la non-réponse totale aux jeux du volet 2003 est élevée, d'autant plus que les répondants possèdent des caractéristiques différentes des non-répondants et que l'on observe des liens significatifs entre certaines de ces caractéristiques et les mesures de développement cognitif.

1.3 Non-réponse partielle

Parmi les répondants globaux, les non-répondants partiels à une tâche donnée sont ceux qui n'ont pas complété cette tâche (sauf s'ils ont atteint le critère d'arrêt du jeu s'il y a lieu) ou dont les données recueillies sont invalides. Au volet 2003, certains enfants, bien que répondants à un jeu selon cette définition, ne se sont vu attribuer aucun score. Dans ce cas, l'absence de score semble liée aux réponses fournies par l'enfant et n'est pas due à une information incomplète sur le déroulement de la tâche. A fortiori, puisque les résultats obtenus par ces enfants diffèrent de ceux obtenus par les autres enfants, il ne serait pas adéquat de simplement considérer que les enfants répondants les représentent. Ainsi, autant que faire se peut, **les scores manquants devraient être imputés et ces enfants considérés répondants**. À tout le moins devrait-on étudier en quoi les enfants qui ont complété la tâche mais n'ont pas de score diffèrent des enfants auxquels un score a été attribué afin de nuancer les résultats des analyses.

1.3.1 Jeu de blocs (WPSI-R)

Au jeu de blocs, l'enfant doit reproduire un modèle. L'intervieweur devait interrompre le jeu après trois erreurs consécutives. Ainsi, parmi les 1 012 (997 + 15) enfants ayant participé aux tâches cognitives, 1 001 ont participé au jeu de blocs, dont sept enfants avec arrêt prématuré (cinq enfants n'ayant obtenu que des succès pour les modèles réalisés) et quatre enfants codés « refus partiel » où le jeu a été interrompu après trois erreurs consécutives. Ces quatre enfants sont répondants et un score devrait être calculé pour chacun d'eux. Si l'on considérait non-répondants les sept enfants avec arrêt prématuré, on obtiendrait un taux de non-réponse partielle de 1,8 %. Autrement, si l'on décidait de leur imputer des valeurs aux essais manquants, ce taux serait de 1,1 %.

Hormis les cas décrits, les enfants sont non-répondants partiels soit parce qu'ils ont refusé de faire la tâche en tout ou en partie, soit parce que la mère a aidé l'enfant, invalidant ainsi les données recueillies. Aucun score n'a été calculé jusqu'à maintenant pour l'ensemble des répondants à cette tâche.

1.3.2 Jeu de mots (ÉVIP)

Au jeu de mots, l'enfant doit repérer les images correspondant à une série de mots. On compte 1 003 répondants au jeu de mots, dont 972 enfants auxquels un score a été attribué. Le score à ce jeu donne le nombre d'échecs entre une base et un plafond, la base étant la plus haute série de huit bonnes réponses consécutives et le plafond étant la plus basse série de huit réponses consécutives contenant six échecs. Or pour 31 répondants, les résultats obtenus sont tels qu'il n'a pas été possible de définir une base ou un plafond. Aucun score n'a par conséquent été attribué à ces enfants répondants.

Sur la base des observations recueillies, il serait approprié de leur imputer un score, même imparfait, puisque la tâche a été complétée, d'autant plus que l'absence de score est directement liée aux résultats obtenus. Considérant qu'un score est attribué à chacun des 1 003 enfants répondants, on obtiendrait un taux de non-réponse partielle de 0,9 %. Autrement, le taux de non-réponse partielle serait de 4,0 % et l'analyse nécessiterait un examen du biais introduit par l'absence de score pour ces enfants.

Lorsque connue, la non-réponse partielle est due au refus de continuer de l'enfant ou d'un problème survenu dans l'administration de la tâche.

1.3.3 Jeu de mémoire (VCR)

Au jeu de mémoire, l'enfant doit pointer certaines images sur une carte dans le même ordre que l'intervieweur les a d'abord lui-même pointées. Le jeu est interrompu si l'enfant fait au moins une erreur à deux cartes consécutives.

Mille trois enfants ont participé au jeu de mémoire et se sont vu attribuer un score. Parmi ceux-ci, on compte 70 enfants pour lesquels certaines cartes étaient vides avant que le critère d'arrêt du jeu n'ait été atteint. Les scores semblent avoir été calculés en considérant les cartes vides comme des échecs. Considérant ces scores valides, le taux de non-réponse partielle n'est que de 0,9 %. La non-réponse partielle au VCR est due au manque de temps ou au fait que l'intervieweur n'a pas bien rempli le questionnaire, rendant ainsi les données invalides.

1.3.4 Jeu des intersections (FIT)

Le FIT est un jeu qui consiste à trouver l'intersection commune d'un certain nombre de formes géométriques reproduites sur un ensemble de planches. On compte 991 enfants y ayant participé. Bien que certains intervieweurs aient noté des problèmes d'interférence et de langue, ces derniers ne semblent pas avoir affecté les résultats de ce jeu de façon significative. Aucun score n'a été calculé pour l'instant pour cette tâche. Le taux de non-réponse partielle est de 2,1 %.

La non-réponse partielle est due au refus de l'enfant ou à un problème survenu durant l'administration de la tâche.

1.3.5 Activité sur les nombres

L'enfant doit répondre à certaines questions portant sur les nombres. Il doit obtenir au moins trois bonnes réponses à cinq questions au niveau 0 pour accéder au niveau 1. Le niveau 1 est interrompu après trois erreurs consécutives s'il y a lieu.

Au jeu des nombres, on compte 990 enfants dont la tâche s'est bien déroulée. On compte de plus 11 enfants auxquels un score a été attribué, bien qu'il y ait eu un problème (partiel) durant l'administration de la tâche. Dans tous les cas, ces enfants ont réussi le niveau 0 mais n'ont pas complété le niveau 1. Par ailleurs, 28 des 990 enfants dont la tâche s'est bien déroulée et ayant atteint le critère d'arrêt du niveau 0 ne se sont vu attribuer aucun score. Tel qu'il a été fait au volet 2002, il serait pertinent de calculer un score pour ces enfants (possiblement en considérant qu'ils auraient eu des échecs au niveau 1).

Considérant qu'un score est calculé pour les 28 enfants ayant atteint le critère d'arrêt du niveau 0 et que les scores sont valides pour les 11 enfants qui ont fait le niveau 0 mais n'ont pas complété le niveau 1, on obtient un taux de non-réponse partielle de 1,1 %.

1.4 Conclusion

Considérant l'ampleur de la non-réponse totale et les faibles taux de non-réponse partielle obtenus à chacune des tâches, seule la non-réponse totale a été analysée en vue de créer une pondération unique tenant compte de la non-réponse à l'ensemble des jeux. Sachant que les enfants qui ont participé aux jeux possèdent des caractéristiques différentes des autres enfants, cette pondération permettra de minimiser les biais dus à la non-réponse dans les estimations qui seront produites.

Notons que la collecte des données du volet 2003 a été affectée par des problèmes d'ordre budgétaire et a, de ce fait, dû débuter à peu de jours d'avis. Pour le volet 2004, le contexte de planification étant meilleur, la proportion de répondants obtenue aux jeux, parmi les répondants au QIRI, est de l'ordre de 80 %, comparativement à près de 60 % au volet 2003.

2. Pondération

La pondération consiste à assigner à chaque enfant défini répondant un poids statistique qui correspond au nombre d'enfants qu'il représente au sein de la population visée. Dans le cadre des jeux, la pondération a pour but de minimiser les biais potentiels liés au fait que les enfants répondants puissent être différents des enfants non-répondants.

Vu que la presque totalité des enfants ayant participé aux tâches cognitives du volet 2003 étaient également répondants au QIRI de ce volet, la stratégie de pondération retenue est hiérarchique, tout comme pour les volets précédents. En effet, la pondération des tâches cognitives du volet 2003 consiste en un ajustement pour la non-réponse fait à partir des poids QIRI de ce volet (poids de base), ces poids étant eux-mêmes basés sur les poids QIRI du volet 2002, et ainsi de suite.

2.1 Problèmes chroniques liés au développement

En lien avec la pondération des jeux des volets 2001 et 2002, la question des enfants souffrant de problèmes chroniques liés au développement a été étudiée sur la base des troubles diagnostiqués déclarés par le parent à l'un ou l'autre des volets de 2001 à 2004. On constate que parmi les 1 759 répondants au QIRI, 11 enfants souffrent d'autisme ou d'incapacité mentale. Parmi ces derniers, cinq enfants étaient disponibles pour participer aux tâches cognitives au volet 2003; un seul de ces enfants a finalement participé mais il a obtenu des scores faibles; les autres enfants ont été incapables d'accomplir les tâches demandées. Par ailleurs, les 57 enfants souffrant de problèmes chroniques tels que le déficit d'attention, la dysphasie, le retard intellectuel et les troubles de développement, d'apprentissage ou émotif présentent un taux de réponse semblable à l'ensemble des enfants.

Les enfants souffrant de problèmes de santé chroniques diagnostiqués tels que l'autisme ou l'incapacité mentale ont été considérés répondants « inaptes » pour l'ensemble des jeux. Dans la suite du texte, le terme « inapte » sera utilisé pour désigner ces enfants.

Les enfants « inaptes » peuvent être considérés répondants aux fins de l'analyse; il est alors suggéré d'imputer les scores les plus faibles (théoriques ou rencontrés) aux essais pour lesquels les données sont manquantes. Autrement, ces enfants pourraient également être exclus de certaines analyses pour lesquelles l'inférence est faite à la sous-population des enfants qui ne présentent pas de telle incapacité chronique.

Puisque les enfants souffrant d'un autre type de trouble chronique présentent un taux de réponse semblable à l'ensemble des enfants, leur non-réponse a été modélisée de la même façon que l'ensemble; la variable indiquant la présence ou non de trouble chronique a toutefois été considérée comme variable potentielle pour la formation des classes de pondération. Cette dernière n'a finalement pas été retenue.

2.2 Enfants non-répondants au QIRI

Les 15 enfants qui ont participé aux tâches cognitives du volet 2003 sans pour autant être répondants au QIRI n'avaient a priori aucun poids QIRI pour ce volet. Il a donc été nécessaire de leur attribuer un tel poids, pour ensuite faire un ajustement pour la non-réponse aux tâches cognitives. Toutefois, puisque les poids QIRI du volet 2003 ont eux-mêmes comme poids de base les poids du volet 2002, l'attribution d'un poids aurait été complexe pour les quatre enfants qui n'ont participé à l'enquête ni au volet 2003, ni au volet 2002 (dont deux enfants n'ayant participé à l'enquête qu'à deux volets). Par conséquent, pour les tâches cognitives du volet 2003, un poids a été calculé uniquement pour 1 018 enfants, soit 997 enfants répondants au QIRI ainsi qu'aux tâches cognitives au volet 2003, 10 enfants répondants au QIRI au volet 2003 mais n'ayant pas participé aux jeux car « inaptes » (mais considérés « répondants » aux jeux) et 11 enfants répondants au QIRI au volet 2002 (mais non au volet 2003) et aux tâches cognitives au volet 2003.

Pour les 11 enfants qui n'ont participé qu'aux tâches cognitives au volet 2003, les poids QIRI de ce volet ont été imputés selon la définition des classes de pondération utilisée pour la création de la pondération QIRI 2003 initiale. C'est dire que le même ensemble de variables socioéconomiques a été retenu pour la formation de ces classes. En redéfinissant les 11 enfants comme « répondants » au QIRI du volet 2003, on calcule un nouveau poids QIRI 2003 pour les 1018 enfants « répondants » en divisant leur poids QIRI 2002 par la proportion pondérée de « répondants » au QIRI 2003 au sein de la classe à laquelle ils appartiennent.

2.3 Taux de réponse global

Pour le calcul du taux de réponse, on considère ici que les répondants sont les 1 018 enfants auxquels un poids sera attribué pour les tâches cognitives du volet 2003. Le dénominateur regroupe les 1759 enfants ayant participé à l'enquête au volet 2003, en plus des 11 enfants ayant participé uniquement aux tâches cognitives. Sur cette base,

la proportion pondérée de répondants aux jeux parmi les 1770 « répondants » au QIRI est de 55 %. Sachant que le taux de réponse au QIRI est de 62 % au volet 2003, on obtient un taux de réponse pondéré de **34 %** ($62 \% \times 55 \%$) pour les tâches cognitives, relativement à l'échantillon initial de l'enquête ($n=2817$).

Le fait que les enfants ayant participé aux tâches cognitives soient caractérisés fait en sorte que certains sous-groupes sont sous-représentés de façon très importante au sein de l'échantillon des répondants. En effet, avec un taux de réponse global de 34 %, cela signifie que le taux de réponse est encore plus faible pour les sous-groupes pour lesquels la participation aux tâches est moindre. Par ailleurs, à travers les années, l'érosion semble toujours se produire au sein des sous-groupes des familles plus défavorisées ou dont la mère est jeune, peu scolarisée ou immigrante. Cela fait en sorte que la proportion de répondants peut devenir très faible dans certains sous-groupes au cours des ans. Il est alors risqué de penser qu'un petit nombre de répondants représente adéquatement les non-répondants. La pondération est dans ce cas un outil qui minimise les risques de biais mais il est peu probable qu'elle permette à elle seule de corriger les biais dus à la non-réponse. Il n'en demeure pas moins que le contexte de l'ÉLDEQ est favorable au calcul d'une pondération puisqu'une quantité appréciable de variables y sont mesurées, tant pour les répondants que les non-répondants.

2.4 Analyse de la non-réponse

Puisque la presque totalité des répondants aux tâches cognitives du volet 2003 sont également répondants au QIRI pour ce volet, la stratégie de pondération consiste à faire un ajustement pour la non-réponse aux jeux à partir des poids du QIRI du volet 2003 (poids de base). Pour l'ensemble des jeux, plusieurs variables ont été considérées pour la modélisation; il s'agit principalement de variables de nature socioéconomique et comportementale.

2.4.1 Recours au service de garde

La proportion de répondants varie significativement selon le recours actuel ou non à un mode de garde. Les enfants qui avaient recours à un mode de garde au moment de l'enquête ont en effet participé en plus grande proportion que les autres. Se pourrait-il que ce phénomène puisse être en partie expliqué par le fait que la collecte du volet 2003 comportait un volet « garderies », incitant possiblement à un effort plus grand pour obtenir la participation des enfants qui fréquentent un milieu de garde? Quoi qu'il en soit, le recours à un mode de garde ne semble pas lié aux mesures de développement disponibles au moment présent et, bien que considérée pour la modélisation, cette variable n'a pas été retenue pour la formation des classes de pondération.

2.4.2 Régions sociosanitaires

La région sociosanitaire de l'enfant au volet 1998, variable disponible par le biais de la base de sondage, a été considérée pour la modélisation de la non-réponse aux jeux. On observe en effet une proportion de répondants plus élevée que la proportion globale pour certaines régions (Chaudière-Appalaches, 75 %; Saguenay-Lac-St-Jean, 71 %) et moins élevée pour d'autres (Abitibi-Témiscamingue, 28 %; Mauricie, 43 %).

Par ailleurs, les scores obtenus à l'ÉVIP et au VCR présentent certaines variations régionales. Ces variations peuvent vraisemblablement être expliquées par certains facteurs socioéconomiques ou autres. Le fait de retenir une telle variable pour la formation des classes de pondération finales ne paraissait pas un choix judicieux. Ainsi, dans les cas où la région est ressortie dans les analyses préliminaires visant la formation des classes de pondération, certaines vérifications ont été faites. D'une part, il a été vérifié que la relation avec les mesures de développement demeurait significative avec le regroupement des régions obtenu selon la proportion de répondants. D'autre part, il a été vérifié si une autre variable significativement liée à la non-réponse aux jeux pouvait se substituer à la région. Dans tous les cas, ces vérifications ont conduit à ne pas retenir la région pour la formation des classes de pondération finales.

2.4.3 Définition des classes de pondération

L'ajustement de la pondération pour la non-réponse est basé sur la création de classes de pondération. Comme aux volets précédents, la technique employée est la modélisation par segmentation fondée sur l'algorithme CHAID (« Chi-square automatic interaction detection ») mis au point par Kass (1980). Les classes de pondération sont créées sous forme d'arborescence; elles ne résultent donc pas nécessairement du croisement de toutes les variables considérées pour la modélisation. Pour une famille donnée, l'ajustement de la pondération consiste à diviser le poids de référence par la proportion pondérée d'enfants répondants observée au sein de la classe à laquelle il appartient.

Pour l'ensemble des jeux, les enfants classifiés « inaptes » sont exclus du groupe des enfants sur lesquels porte l'analyse de la non-réponse. Nous supposons que nous avons une liste exhaustive des enfants déclarés souffrant de problèmes tels que l'autisme ou l'incapacité mentale (parmi les répondants à l'enquête des volets de 2001 à 2004). Le poids QIRI de base du volet 2003 est attribué à ces enfants; ils ne représentent donc aucun enfant non-répondant aux jeux.

Pour les autres enfants, les variables considérées pour la modélisation proviennent généralement du QIRI du volet étudié. Quelques variables ont également été étudiées sous forme longitudinale, c'est-à-dire qu'elles ont été créées à partir d'une même mesure prise à différents volets (volet étudié et volets antérieurs). Ces variables sont le niveau de suffisance du revenu (revenu insuffisant à au moins un volet c. autres), le revenu du ménage (revenu inférieur à 10 000\$ à au moins un volet c. autres), le type de famille (monoparentale à au moins un volet c. autres), le travail de la mère au cours des douze derniers mois (mère n'a pas travaillé au cours des douze derniers mois pour plus d'un volet c. autres) et la présence du père biologique (père absent à au moins un volet c. autres).

Pour une variable donnée, les proportions de répondants ont été comparées entre les différentes catégories de cette variable à l'aide d'un test du khi-deux approximatif. Par la suite, les variables pour lesquelles le seuil observé du test était inférieur à 0,15 ont été considérées pour la modélisation de la non-réponse. Parmi ce groupe de variables, ont d'abord été considérées les variables qui présentaient une relation significative, au seuil 0,15, avec les mesures de développement disponibles (scores à

l'ÉVIP, au VCR et au jeu des nombres). Ce lien avait été étudié au préalable au sein du sous-groupe des enfants répondants. En effet, l'ajustement de la pondération pour la non-réponse n'est utile que dans la mesure où les caractéristiques retenues sont liées aux mesures d'intérêt. Finalement, à l'intérieur des classes de pondération ainsi formées, un examen des variables restantes liées à la non-réponse a été fait afin de compléter l'analyse et de définir des classes plus fines s'il y a lieu.

2.4.5 Résultats

Parmi l'ensemble des variables considérées pour l'ajustement pour la non-réponse au volet 2003, les variables suivantes ont été retenues :

- le plus haut niveau de scolarité de la mère ou conjointe du père (FEDMD01)
- le niveau de suffisance du revenu (variable longitudinale : revenu insuffisant à au moins un volet c. autres)
- le statut actuel de l'union (FDEFD4A)
- la première langue apprise par l'enfant (FSDED06)

Au sein des différentes classes d'ajustement de la pondération, la proportion de répondants varie de 34 % à 66 % (relativement à une proportion globale de 55 %). Les proportions les plus faibles sont observées dans les classes où le parent est sans conjoint et où la mère/conjointe est absente ou peu scolarisée.

3. Fichier de pondération et recommandations

3.1 Fichier de pondération

Le fichier SAS POIDS_JEUX601 contient la variable de pondération PEJEUX6 qui prend une valeur pour chacun des 1 018 enfants qui ont participé à au moins une tâche cognitive au volet 2003 ou ont été déclarés « inaptes ». Ces poids doivent être utilisés dans les analyses portant sur les jeux afin d'inférer les résultats à la population visée tout en minimisant les biais dans les estimations. Notons qu'il s'agit de poids échantillonaires, c'est-à-dire de poids qui ont été multipliés par une constante de sorte que la somme des poids soit égale à la taille de l'échantillon. Ces poids peuvent par conséquent être utilisés pour faire des tests approximatifs avec des logiciels qui ne tiennent pas compte du plan de sondage complexe dans l'estimation de la variance et les tests statistiques. Afin de pallier au caractère approximatif de tels tests, il est recommandé d'adopter une approche conservatrice en abaissant le seuil théorique des tests statistiques. Par exemple, si l'on souhaite faire des tests au seuil théorique de 0,05, on peut choisir de n'interpréter que les résultats significatifs au seuil 0,01.

Le fichier POIDS_JEUX601 compte également la variable INAPTE_E6 permettant d'identifier les enfants déclarés « inaptes », c'est-à-dire souffrant d'autisme ou d'incapacité mentale. Cette variable prend la valeur « 1 » pour les onze enfants « inaptes »; autrement, elle prend une valeur nulle. Selon le besoin, ces enfants peuvent être inclus ou non dans les analyses. Dans le cas où ces enfants sont inclus dans les analyses, on doit imputer les valeurs les plus faibles (score nul, par exemple) aux valeurs manquantes pour ces enfants.

Finalement, le fichier POIDS_JEUX601 contient deux variables additionnelles, soit IMPUT_EVIP_E6 et IMPUT_NBRES_E6 indiquant les enfants pour lesquels un score devrait être imputé préalablement à l'analyse, excluant les cas des enfants « inaptes ». Ces variables prennent la valeur « 1 » lorsque cette imputation est requise et « 0 » autrement.²

3.2 Recommandations aux fins de l'analyse

Pour l'analyse des données des tests cognitifs du volet 2003, la pondération demeure incontournable étant donné les différences importantes observées entre répondants et non-répondants. Il faut toutefois garder en tête que les statistiques calculées à l'aide des données des tâches cognitives du volet 2003 devront être considérées comme étant exploratoires et non définitives étant donné l'ampleur de la non-réponse.

3.2.1 Vagues de collecte et effet de l'âge

Les vagues de collecte préassignées à chacun des enfants n'ont en général pas été respectées pour les tâches cognitives du volet 2003, tant du point de vue du mois de collecte comme tel que de l'ordre où les enfants devaient être interviewés. Par ailleurs, la répartition du mois de naissance de l'enfant selon la date d'administration des jeux n'est pas uniforme. En effet, les enfants ayant participé aux jeux en mai 2003 ou avant ont tendance à être nés plus tôt que l'ensemble des enfants, alors que les enfants y ayant pris part en septembre ou après sont proportionnellement plus nombreux à être nés plus tard que l'ensemble.

Puisque l'âge de l'enfant semble en général lié aux mesures de développement cognitif (Plante, N., Courtemanche, R. et Tremblay, M.-È., 2004), en l'absence d'une assignation aléatoire de vagues de collecte, il n'y a pas nécessairement indépendance entre l'âge et les variables explicatives d'un modèle. **Il est dans ce cas requis de tenir compte de l'âge de l'enfant lors de la modélisation pour vérifier s'il existe un effet confondant sur la relation entre la mesure de développement cognitif étudiée et l'une des variables explicatives, ou encore une interaction entre l'âge et une variable explicative.**

Un effet confondant indiquerait que la relation entre le score obtenu à un jeu et la variable explicative en cause est due, en tout ou en partie, à la variation de la distribution de l'âge selon cette variable explicative. Par exemple, on pourrait observer un effet confondant de l'âge sur la relation entre le score à l'ÉVIP et le statut socioéconomique si les enfants enquêtés en début de période de collecte étaient plus défavorisés. En effet, une telle situation mènerait vraisemblablement à une surestimation de l'écart de développement cognitif entre les enfants défavorisés et les autres puisque les enfants enquêtés plus tôt sont en moyenne plus jeunes. C'est dire

² Parmi les 31 enfants ayant un score manquant, il y a un enfant qui est non répondant au QIRI des volets 2002 et 2003 et de ce fait, aucun poids n'a été créé pour celui-ci. On compte donc 30 enfants pour lesquels un score doit être imputé à l'ÉVIP. Pour le jeu des nombres, il y a 28 enfants pour lesquels une telle imputation est requise.

que l'« effet » du statut socioéconomique sur le développement cognitif³ serait surévalué dû à la présence d'une plus grande proportion d'enfants plus jeunes parmi les plus défavorisés. Le fait d'inclure l'âge de l'enfant dans le modèle permet d'ajuster les paramètres associés aux autres variables explicatives pour tenir compte de cette répartition non uniforme de l'âge selon le statut socioéconomique.

L'absence d'indépendance entre l'âge et une variable explicative pourrait également se traduire en interaction entre ces deux variables. Si les effectifs le permettent, il serait intéressant de tester la présence d'interaction de l'âge avec les variables explicatives importantes.

Toutefois, notons que même si le simple fait d'avoir une date d'entrevue plus tardive fait en sorte que l'âge de l'enfant soit plus élevé lors de l'administration des jeux, les écarts d'âge observés entre ceux qui ont fait l'entrevue en début de période de collecte et ceux qui l'ont fait vers la fin ne semblent pas exagérément élevés.

3.2.2 Non-réponse partielle

Les poids associés aux différents jeux ont été ajustés pour tenir compte de la non-réponse totale aux jeux. Ces poids ne tiennent toutefois pas compte de la non-réponse partielle aux jeux, ni de la non-réponse totale aux instruments autoadministrés. Ainsi, pour un jeu donné, il faut être attentif aux résultats des analyses pour lesquelles le nombre d'unités d'analyse diffère de façon importante du nombre d'enfants pour lesquels un poids est disponible pour ce jeu. S'il s'avérait qu'une part importante de la non-réponse n'ait pas été prise en compte dans la pondération, les biais dans les estimations pourraient être non négligeables. Par exemple, on compte 1 018 enfants ayant un poids aux jeux du volet 2003. Parmi ceux-ci, on compte 875 enfants dont la mère a complété le questionnaire autoadministré de la mère (QAAM). Ainsi, dans une analyse impliquant à la fois des données des jeux et du QAAM du volet 2003, si l'on utilisait les poids PEJEUX6, on compterait 14 % de non-réponse partielle pour laquelle aucun ajustement des poids n'a été fait (en plus de la non-réponse partielle au jeu proprement dit), ce qui n'est pas négligeable. Dans un tel cas, il est recommandé de faire une analyse du biais en étudiant, par exemple, le sens attendu de ce dernier.⁴

³ Les données ayant été recueillies dans un contexte observationnel, on ne peut, à proprement parler, établir de relations de cause à effet puisque l'on n'a pas contrôlé ni mesuré tous les facteurs pouvant influencer les mesures de développement cognitif. Il demeure intéressant, dans le contexte d'un modèle statistique, d'inclure une variable confondante de manière à tenir compte de son « effet » sur la relation entre les autres variables.

⁴ Il s'agit, par exemple, de vérifier s'il existe une relation significative entre une variable caractérisant la non-réponse partielle (répondant ou non-répondant) et certaines caractéristiques des enfants ou de leurs familles qui sont liées à la mesure de développement cognitif étudiée. Par exemple, s'il s'avérait que les non-répondants partiels sont proportionnellement moins nombreux parmi les enfants présentant des caractéristiques favorables à une mesure élevée de développement, on obtiendrait un biais positif dans l'estimation de la proportion d'enfants ayant une valeur élevée pour cette mesure.

4. Références bibliographiques

KASS, G. V. (1980). « An exploratory technique for investigating large quantities of categorical data », *Applied Statistics*, 29, p.119-127.

PLANTE, N., COURTEMANCHE, R. et TREMBLAY, M-È. (2004). « Examen de l'effet de l'âge au moment de l'entrevue sur les mesures de l'ÉLDEQ », rapport non publié, Institut de la statistique du Québec.