

Étude longitudinale sur le développement des enfants du Québec (ÉLDEQ 1998-2002)

Pondération des données concernant les jeux administrés aux enfants aux volets 2001 et 2002

Nathalie Plante, Marie-Ève Tremblay et Robert Courtemanche
Direction de la méthodologie, de la démographie et des enquêtes spéciales
Institut de la statistique du Québec
27 avril 2004

1. Définition de la non-réponse

Les résultats aux jeux administrés aux enfants revêtent une grande importance puisque ce sont les seules mesures directes faites auprès de ces derniers. Ces résultats seront utilisés comme variables d'analyse dans les modèles visant à étudier le lien entre le développement de l'enfant et d'autres caractéristiques de celui-ci et de sa famille. Parmi les enfants dont les familles ont participé à l'enquête, un nombre non négligeable d'entre eux 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 jeux. À la section 1, la classification des enfants, selon qu'ils sont répondants ou non, est détaillée pour chacun des jeux considérés. La section 2 présente les taux de réponse, la démarche de modélisation de la non-réponse ainsi que les résultats obtenus pour chacun des jeux. Finalement, la section 3 renseigne l'utilisateur sur l'information contenue dans les fichiers de pondération des volets 2001 et 2002 ainsi que sur la façon d'utiliser les poids dans 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.

La non-réponse aux jeux a été analysée pour chacune des tâches suivantes¹:

- Jeu de mots - ÉVIP (volet 2001)
- Jeu de mémoire - VCR (volet 2001 et volet 2002)
- Tâche d'imitation de placement d'objets - IPO (volets 1999-2000-2001 simult.)
- Jeu du triangle caché - PEFT (volet 2002)
- Activité sur les nombres (volet 2002)

Parmi les enfants ayant participé à l'enquête, un certain nombre ont participé, en tout ou en partie, aux jeux visant à évaluer leur développement. Pour un jeu donné, par exemple, certains enfants ont complété la tâche, alors que d'autres ont initié la tâche sans la compléter, soit parce qu'elle était trop difficile, soit parce qu'ils se sont lassés.² Pour certains enfants, il est arrivé que la mère ou une autre personne interfère dans le déroulement de la tâche, invalidant ainsi les données recueillies. D'autres enfants ont par ailleurs refusé de participer ou étaient dans l'impossibilité de le faire parce qu'ils étaient malades, endormis ou absents, ou encore parce qu'ils semblaient incapables d'initier la tâche. Finalement, certains enfants étaient inaptes à réaliser la tâche demandée en raison de problèmes de santé chroniques graves diagnostiqués liés au développement.

Il s'agissait dans un premier temps de classer les enfants, pour chaque jeu, en répondants et non-répondants. Les répondants sont les enfants pour lesquels les données seront analysées. Afin d'inférer les résultats des analyses statistiques à la population visée, on doit associer à chacun de ces enfants un poids statistique qui indique le nombre d'enfants qu'il représente au sein de la population. Grosso modo, les enfants considérés répondants sont les suivants :

- enfants ayant complété la tâche, celle-ci s'étant déroulée sans problème
- enfants n'ayant pas complété la tâche en raison d'un problème de santé chronique lié au développement ou d'un problème de compréhension ou parce qu'ils jugeaient la tâche trop difficile à effectuer (on peut considérer dans ce cas que ces enfants auraient obtenu des scores faibles et par conséquent les inclure dans les analyses en imputant aux données manquantes les scores les plus faibles, théoriques ou rencontrés)

À l'opposé, les enfants non-répondants sont en général les suivants :

¹ Un jeu de blocs (WPPSI-R) a également été administré aux enfants des familles répondantes au volet 2001. Or plus de la moitié de ces enfants n'ont pas complété la tâche et il s'avère que ces derniers ont obtenu, en moyenne, des scores plus élevés que les enfants ayant complété la tâche. Étant donné le faible taux de réponse obtenu pour ce jeu ainsi que la complexité de la pondération envisagée, il a été décidé de ne pas créer de pondération spécifique à ce jeu. C'est donc dire que les résultats basés sur les données du jeu des blocs ne peuvent servir qu'à décrire l'échantillon des répondants et non la population visée.

² On considère qu'une tâche est complétée lorsque l'enfant a fait le jeu dans sa totalité ou qu'il a atteint un critère d'arrêt du jeu conditionnel aux réponses qu'il a fournies.

- enfants n'ayant pas initié ou n'ayant pas complété la tâche dû à une autre cause que la précédente, en général indépendante de lui-même ou momentanée (ex. : enfant absent, malade ou endormi, refus de l'enfant ou de la mère, refus de continuer parce que l'enfant s'est lassé ou sans raison apparente)
- enfants ayant fait la tâche en tout ou en partie mais ayant été dérangés ou influencés de façon importante durant le déroulement de la tâche

Pour la plupart des jeux, plusieurs indicateurs ont été développés afin de caractériser le statut de réponse d'un enfant à la tâche ainsi que le déroulement de celle-ci. Or de nombreuses incohérences entre ces différents indicateurs ont été notées. Par ailleurs, puisque les codes de réponse n'ont pas été standardisés d'une tâche à l'autre, la classification des enfants selon les quatre groupes ci-dessus définis a dû être faite sous certaines hypothèses. Les indicateurs retenus pour la classification des enfants selon leur participation aux jeux sont les variables décrivant l'état de passation de chacun des jeux. On suppose que l'information contenue dans ces variables est la plus fiable.³ La classification des enfants est également basée sur les données du Questionnaire informatisé rempli par l'intervieweuse (QIRI) ou du Questionnaire papier rempli par l'intervieweuse (QPRI) quant à l'identification des cas de problèmes de santé chroniques diagnostiqués (voir section 1.1.2).

1.1 Classification des enfants selon leur participation aux jeux

1.1.2 Enfants souffrant de problèmes de santé chroniques liés au développement

Sur la base des informations recueillies à l'aide du QIRI ou du QPRI des volets 2001, 2002 et 2003, les enfants souffrant de problèmes de santé chroniques diagnostiqués tels que l'autisme, l'incapacité mentale ou un trouble envahissant du développement ont été considérés répondants « inaptes » pour l'ensemble des jeux, qu'ils aient ou non participé à la tâche. Tout problème de ce type déclaré à au moins un des volets a été considéré pour classer ainsi les enfants. 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.

1.1.3 Jeu de mots (ÉVIP) - volet 2001

Pour le jeu de mots, aucune variable décrivant l'état de passation du jeu n'était disponible. Ainsi, la classification des enfants selon leur participation a été faite sur la base du score brut de l'ÉVIP (variable DEVES01). Hormis le cas des enfants classés « inaptes », ceux ayant un score sont les répondants à ce jeu. On fait l'hypothèse que

³ Pour le jeu de mots (ÉVIP), cet indicateur n'était pas disponible.

la tâche s'est bien déroulée dans tous les cas (aucun problème tel que maladie de l'enfant ou interférence de la mère, etc.). Tous les autres enfants sont considérés non-répondants.

1.1.4 Jeu de mémoire (VCR) - volets 2001 et 2002

Les répondants au jeu de mémoire comptent, en plus des enfants « inaptes » et des enfants ayant complété la tâche sans problème, quelques enfants ayant été déclarés incapables de faire la tâche, selon l'intervieweuse (code « incapable » selon la variable « état de passation » de la tâche). Contrairement aux enfants « inaptes » qui ont ainsi été classifiés en raison d'un problème de santé diagnostiqué rapporté par le répondant, certains enfants ont été classifiés « incapables » par l'intervieweuse, pour un jeu en particulier, lorsqu'ils ne semblaient pas être en mesure d'initier la tâche.

Notons que plusieurs des enfants classés « incapables » font partie du groupe des enfants classés « inaptes » d'après le QIRI. Ainsi, comme pour les répondants « inaptes », **les enfants « incapables » au jeu de mémoire sont considérés répondants avec imputation des résultats les plus faibles lorsque manquants.** On fait alors l'hypothèse que ces enfants n'avaient pas la capacité de faire le jeu et que leur non-participation n'est pas due à une cause momentanée ou indépendante d'eux. Tous les autres enfants sont considérés non-répondants.

1.1.5 Tâche d'imitation de placement d'objets (IPO) - volets 1999-2000-2001

Pour la tâche d'imitation de placement d'objets, la non-réponse est étudiée de façon longitudinale. Les répondants sont dans ce cas définis comme ceux qui sont répondants à chacun des volets 1999, 2000 et 2001.

Le groupe des enfants répondants est constitué de ceux ayant complété la tâche sans interférence importante⁴ ainsi que de ceux ayant été classés « inaptes » ou « incapables⁵ ». De plus, seuls les enfants pour lesquels tous les niveaux de difficulté ont été complétés (avec succès ou non) ont été considérés répondants puisque l'analyse de la réussite à ce jeu doit intégrer les résultats de chacun des essais. On suppose pour ces enfants qu'un abandon de la tâche n'était pas dû au fait que l'enfant avait eu de la difficulté à faire les premiers essais⁶.

1.1.6 Jeu du triangle caché (PEFT) - volet 2002

Comme pour le VCR et l'IPO, les enfants considérés répondants à ce jeu sont ceux ayant complété la tâche sans trop d'interférence ainsi que ceux classés « inaptes » ou « incapables ». Parmi les enfants ayant complété le jeu du triangle caché, un certain

⁴ Un nombre beaucoup plus grand de cas d'interférence importante ont été relevés au volet 2001, comparativement aux volets précédents. Cela pourrait être dû au fait que ces cas ont fait l'objet d'un code de réponse spécifique défini a priori au volet 2001, ce qui n'était pas le cas aux volets 1999 et 2000. On doit dans ce cas faire l'hypothèse qu'il n'y a pas eu d'interférence importante pour les enfants répondants des volets 1999 et 2000.

⁵ Enfants classés « incapables » d'après la variable d'état de passation du volet 2001.

⁶ Au volet 1999, l'enfant faisait deux essais pour chaque niveau de difficulté; on considère dans ce cas qu'un niveau est complété lorsqu'au moins un des deux essais a été complété pour ce niveau.

nombre ont subi un niveau élevé d'interférence lors du déroulement de la tâche et ont, de ce fait, été considérés comme non-répondants.

1.1.7 Activité sur les nombres - volet 2002

Les répondants à l'activité sur les nombres sont, en plus des enfants « inaptes » et « incapables », ceux qui ont obtenu un score final à cette tâche.

Il est important de préciser que l'activité sur les nombres comporte deux niveaux de difficulté et que la participation au second niveau est conditionnelle à l'obtention de trois bonnes réponses ou plus au premier niveau, lequel comporte cinq questions. Cela étant, on note que parmi les enfants ne possédant pas de score final à cette tâche, environ la moitié n'ont pas du tout participé alors qu'un peu plus du quart ont complété au moins le premier niveau. Pour ceux ayant complété le premier niveau seulement, les résultats obtenus pour les questions répondues sont comparables à ceux des enfants qui ont complété la tâche. Par contre, une forte proportion des enfants n'ayant pas complété le premier niveau ont cumulé trois échecs ou plus au premier niveau et auraient, de ce fait, été inadmissibles au second niveau. Ces enfants auraient par conséquent obtenu un faible score même s'ils avaient complété la tâche. L'information disponible sur le premier niveau pour les enfants qui ont participé au jeu, sans pour autant compléter l'ensemble de la tâche, peut être utilisée avantageusement pour la modélisation de la non-réponse puisqu'elle est directement liée à la mesure d'intérêt, soit le score final.

La non-réponse à l'activité sur les nombres a donc été analysée en deux étapes. En premier lieu, la non-réponse totale a été considérée, sans tenir compte de la non-réponse partielle. Les répondants comptent dans ce cas, en plus des enfants « inaptes » et « incapables », tous ceux ayant participé à la tâche, en tout ou en partie, à l'exception des enfants malades durant la tâche, aidés par leur mère, ou encore « allophones ». Les enfants pour lesquels il n'était pas possible de déterminer l'admissibilité au second niveau, puisqu'ils n'avaient pas complété le premier niveau ni obtenu trois échecs ou plus à ce niveau, ont été considérés non-répondants en raison d'une information insuffisante.

En second lieu, la non-réponse partielle a été considérée de manière à compléter l'analyse de la non-réponse. Ainsi, parmi les répondants totaux, le groupe des répondants est dans ce cas constitué des enfants ayant complété la tâche sans problème ainsi que de ceux classés comme étant « inaptes » ou « incapables ».

2. Analyse de la non-réponse

2.1 Taux de réponse

On compte respectivement 1 950 et 1 944 enfants ayant participé à l'enquête aux volets 2001 et 2002. Pour chacun de ces volets, le tableau I présente le nombre de répondants aux différents jeux, la proportion pondérée de répondants parmi les enfants ayant participé à l'enquête ainsi que le taux de réponse pondéré relativement à l'échantillon initial. Ce dernier peut être interprété comme une mesure de qualité des données. Sauf pour la tâche d'imitation de placement d'objets, il s'obtient par le

produit du taux de réponse pondéré transversal du QIRI du volet correspondant et de la proportion pondérée de répondants au jeu parmi les répondants au QIRI. Dans le cas de l'IPO, il est le résultat du produit du taux de réponse pondéré longitudinal au QIRI pour les volets 1999, 2000 et 2001 et de la proportion pondérée de répondants longitudinaux à cette tâche parmi les répondants longitudinaux au QIRI.

Tableau I – Nombre de répondants et taux de réponse pondéré par jeu

Jeu	Nombre de répondants	Prop. de rép. parmi rép. QIRI	Taux de réponse
Jeu de mots (ÉVIP) - volet 2001	1 805	90,7 %	61,9 %
Jeu de mémoire (VCR) - volet 2001	1 783	90,4 %	61,7 %
Jeu de mémoire (VCR) - volet 2002	1 851	94,4 %	64,4 %
Tâche d'imitation de placement d'objets (IPO) - volets 1999-2000-2001	1 602	81,3 %	55,4 %
Jeu du triangle caché (PEFT) - volet 2002	1 864	95,1 %	64,9 %
Activité sur les nombres - volet 2002	1 779	90,1 %	61,4 %

2.2 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. Par exemple, s'il s'avère que les enfants dont la mère ne parle ni le français ni l'anglais à la maison sont moins souvent répondants que les autres, les statistiques liées à cette variable seraient possiblement biaisées si elles n'étaient pas pondérées, comme ce pourrait être le cas pour les résultats au jeu de mots. Par ailleurs, si les enfants plus anxieux participent moins bien aux jeux que les autres et que l'anxiété peut être liée au développement cognitif des enfants, en l'absence d'une pondération appropriée, les estimations seraient entachées d'un biais dû à la non-réponse.

La pondération des données des jeux consiste à faire un ajustement additionnel à la pondération du QIRI du volet en question (pondération de référence). Cet ajustement est basé sur les résultats de la modélisation de la non-réponse spécifique à chacun des jeux. Pour l'ensemble des jeux, plusieurs variables ont été considérées pour la modélisation; elles sont principalement de nature socioéconomique et comportementale. Il est en effet vraisemblable de penser que les variables de comportement pourraient être liées au fait qu'un enfant participe ou non à une tâche. Par exemple, il se pourrait qu'un enfant plus timide soit moins enclin à suivre les directives données par l'intervieweuse. Les variables comportementales utilisées sont principalement des échelles de comportement, sauf dans le cas de la tâche d'imitation de placement d'objets où les questions originales du QIRI ont été utilisées pour le volet 2000. Notons que puisque les variables d'échelle de comportement comptent souvent bon nombre de catégories, un regroupement minimal de celles-ci a été fait de manière à obtenir un test du khi-deux valide pour la comparaison des taux de réponse entre les différentes catégories.

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), 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 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 0,20 ont été considérées pour la modélisation de la non-réponse. Dans le cas de la tâche d'imitation de placement d'objets, ce seuil a été abaissé à 0,10 étant donné le grand nombre de variables considérées pour l'ensemble des volets 1999, 2000 et 2001.

Comme pour l'instrument principal, la méthode de création de classes de pondération a été utilisée. Cette méthode est basée sur l'algorithme CHAID (« Chi-Square Automatic Interaction Detection ») mis au point par Kass (1980). Pour un enfant donné, l'ajustement de la pondération consiste à diviser le poids de référence du QIRI par le taux de réponse pondéré observé dans 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 souffrant de problèmes tels que l'autisme, l'incapacité mentale ou un trouble envahissant du développement (parmi les répondants à l'enquête aux volets 2001 ou 2002). Le poids QIRI de référence est attribué à ces enfants; ils ne représentent donc aucun enfant non-répondant aux jeux.

Par ailleurs, certains enfants identifiés comme souffrant de problèmes chroniques possiblement liés à leur développement (d'après le QIRI/QPRI des volets 2001, 2002 ou 2003) mais dont le problème de santé n'est pas l'un de ceux cités plus haut forment une classe de pondération à part. Les problèmes déclarés sont dans ce cas, pour la plupart, des troubles de développement ou d'apprentissage. Le choix de regrouper ces enfants en une même classe de pondération est lié au fait que l'on observe en général une non-réponse importante et des scores faibles au sein de ce sous-groupe. Ce faisant, on suppose que les répondants de ce sous-groupe ne représentent que les non-répondants du même sous-groupe.

Initialement, la langue parlée par l'enfant (parmi l'anglais ou le français) avait été considérée dans les modèles par le biais des variables originales DSDEQ6BA et DSDEQ6BB. Or les résultats étaient difficiles à interpréter lorsque la langue parlée à la maison par la mère intervenait dans la création des classes de pondération. Il a alors été décidé de dériver une nouvelle variable à partir des deux précédentes de manière à savoir si l'enfant parle le français seulement, l'anglais seulement, ces deux langues ou aucune de celles-ci.

Une variable à six modalités a également été créée pour décrire le mode de garde principal de l'enfant : aucun mode de garde, garde non régulière, garde en milieu familial à 5\$, garde en milieu familial privé, garderie à 5\$, garderie privée.

Pour quelques variables, certaines valeurs manquantes ont été imputées sur la base de l'information disponible aux différents volets. Cette imputation concerne peu d'enfants et n'a été faite qu'aux fins de la pondération.

2.3 Résultats

2.3.1 Jeu de mots (ÉVIP) - volet 2001

Puisque l'ÉVIP est un test de vocabulaire, la langue parlée par l'enfant paraît une variable primordiale pour modéliser la non-réponse à ce jeu. Parmi les variables considérées pour la modélisation, cette variable est effectivement celle qui est la plus fortement liée à la non-réponse, lorsque prise individuellement.

En plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement (autisme, incapacité mentale, trouble envahissant du développement, trouble de développement ou d'apprentissage), la modélisation a permis d'identifier les variables suivantes pour la création des classes de pondération :

- langues parlées par l'enfant (français et/ou anglais, variable créée à partir de DSDEQ6BA et DSDEQ6BB)
- langue parlée à la maison par la mère (DSDMD6A)
- échelle d'anxiété de séparation (DBEES04)
- principal mode de garde (dérivé à partir des variables DCREQ02A, DCREQ1A3, DCREQ1B2, DCREQ1C2 et DCRE1G1A)

2.3.2 Jeu de mémoire (VCR) - volet 2001

Pour le jeu de mémoire du volet 2001, en plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement, la modélisation a permis d'identifier les variables suivantes pour la création des classes de pondération :

- langue parlée à la maison par la mère (DSDMD6A)
- principal mode de garde (dérivé à partir des variables DCREQ02A, DCREQ1A3, DCREQ1B2, DCREQ1C2 et DCRE1G1A)
- échelle de troubles émotifs (DBEES02)

2.3.3 Jeu de mémoire (VCR) - volet 2002

Étant donné la proportion élevée de répondants au jeu de mémoire du volet 2002, un nombre minimal de classes de pondération ont été retenues. En plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement, les variables impliquées dans la formation de ces classes sont les suivantes :

- langues parlées par l'enfant (français et/ou anglais, variable créée à partir de ESDEQ6BA et ESDEQ6BB)
- langue parlée à la maison par la mère (ESDMD6A)
- échelle d'anxiété de séparation (EBEES04)

2.3.4 Tâche d'imitation de placement d'objets (IPO) - volets 1999-2000-2001

Dans l'ajustement pour la non-réponse longitudinale à la tâche d'imitation de placement d'objets pour les volets 1999 à 2001, les classes de pondération ont été formées en fonction des valeurs prises par les variables suivantes (en plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement) :

- le fait que l'enfant prend beaucoup de temps à s'habituer aux autres enfants (CBEEQ8R2)
- insuffisance de revenu à au moins un des volets entre 1998 et 2001 (variable longitudinale créée à partir de AINFD3A, BINFD3A, CINFD3A et DINFD3A)
- timidité de l'enfant avec les enfants qu'il ne connaît pas (CBEEQ8D1)
- langue parlée à la maison par la mère (DSDMD6A)
- recours à un mode de garde (BCREQ1A3)
- réaction de l'enfant face à une nouvelle personne (BTMEQ25)
- le fait que l'enfant ait prononcé une phrase partielle de trois mots ou plus (CMSEQ35)

2.3.5 Jeu du triangle caché (PEFT) - volet 2002

Étant donné la proportion de répondants élevée au jeu du triangle caché, un nombre minimal de classes de pondération ont été retenues. En plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement, les variables impliquées dans la formation de ces classes sont les suivantes :

- langues parlées par l'enfant (français et/ou anglais, variable créée à partir de ESDEQ6BA et ESDEQ6BB)
- principal mode de garde (dérivé à partir des variables ECREQ02A, ECREQ1A3, ECREQ1B2, ECREQ1C2 et ECRE1G1A)
- langue parlée à la maison par la mère (ESDMD6A)
- revenu inférieur à 10 000 \$ à au moins un des volets entre 1998 et 2002 (variable longitudinale créée à partir de AINFD03, BINFD03, CINFD03, DINFD03 et EINFD03)

La proportion pondérée de répondants au PEFT parmi les répondants au QIRI du volet 2002 varie de 70 % à 99 %, la proportion la plus faible étant observée chez les enfants qui ne parlent ni le français, ni l'anglais.

2.3.6 Activité sur les nombres - volet 2002

Étant donné la non-réponse partielle importante à l'activité sur les nombres, la pondération comporte deux étapes d'ajustement. Le poids de référence du QIRI du

volet 2002 est d'abord ajusté pour la non-réponse globale au jeu. Le poids des répondants est par la suite ajusté pour tenir compte de la non-réponse partielle. En définitive, seuls les répondants totaux ont un poids aux fins de l'analyse. Les répondants totaux sont les enfants qui ont obtenu un score final (en plus des enfants « inaptes » ou « incapables »).

À la première étape d'ajustement, des variables socioéconomiques et comportementales ont été utilisées, comme pour les autres jeux. Les variables retenues pour la création des classes de pondération sont les suivantes (en plus des variables permettant d'identifier les enfants souffrant de problèmes de santé chroniques liés au développement) :

- langues parlées par l'enfant (français et/ou anglais, variable créée à partir des variables ESDEQ6BA et ESDEQ6BB)
- principal mode de garde (dérivé à partir des variables ECREQ02A, ECREQ1A3, ECREQ1B2, ECREQ1C2 et ECRE1G1A)
- échelle d'anxiété (EBEES03)
- échelle d'anxiété de séparation (EBEES04)
- support social (ESUFT01)

À l'étape d'ajustement pour la non-réponse partielle, des variables décrivant la réussite au premier niveau ont été ajoutées. En premier lieu, deux classes de pondération ont été formées selon que l'enfant a réussi ou non le premier niveau, ce qui déterminait son admissibilité au second niveau. Cela fait en sorte que pour les enfants qui ont réussi le premier niveau et qui ont complété l'ensemble de la tâche, les poids sont ajustés à la hausse pour représenter les autres enfants qui ont réussi le premier niveau mais qui n'ont pas complété la tâche. À l'opposé, les enfants qui n'ont pas réussi le premier niveau mais l'ont complété verront leurs poids gonflés afin de représenter les enfants qui avaient échoué le premier niveau sans le terminer (trois erreurs ou plus). Au sein de chacune des deux classes ainsi formées, la non-réponse a été modélisée à l'aide des variables habituelles, en plus des variables décrivant le succès ou l'échec aux trois premières questions du premier niveau ainsi qu'à la question préliminaire. En effet, ces questions peuvent être considérées dans la modélisation puisqu'elles ont été répondues par la presque totalité des non-répondants partiels. Pour la classe des enfants ayant réussi le premier niveau, une variable cumulant le nombre de bonnes réponses obtenues au premier niveau (3 à 5) a également été ajoutée dans la modélisation.

Les enfants « inaptes » et « incapables » n'ont pas été considérés pour la modélisation puisque l'on impute dans ce cas les valeurs les plus faibles aux valeurs manquantes. Ces enfants ne comptent donc aucune non-réponse partielle pour laquelle un ajustement est requis. Pour les enfants « inaptes », le poids final demeure le poids de référence du QIRI. Pour les enfants « incapables », le poids final est le poids de référence du QIRI ajusté seulement pour la non-réponse totale à l'activité sur les nombres.

Pour la classe des enfants ayant réussi le premier niveau, les variables suivantes ont été retenues pour la formation des classes de pondération :

- langue parlée à la maison par la mère (ESDMD6A)

- nombre de bonnes réponses obtenues au premier niveau (variable créée à partir de EJNES01, EJNES02, EJNES03, EJNES04 et EJNES05)
- succès ou échec à la question préliminaire (EJNEQPR)
- temps passé par l'enfant à regarder la télévision durant la semaine (EACEQ4B1)
- langues parlées par l'enfant (français et/ou anglais, variable créée à partir de ESDEQ6BA et ESDEQ6BB)
- recours à un mode de garde (ECREQ1A3)

3. Fichiers de pondération, mises en garde et recommandations aux fins de l'analyse

Les variables de pondération pour les jeux sont incluses dans les fichiers SAS POIDS401 et POIDS501 se rapportant respectivement aux volets 2001 et 2002. Ceux-ci contiennent une variable de pondération spécifique à chacun des jeux (PEEVIP4, PEVCR4 et PEIPOL4 pour le volet 2001; PEVCR5, PETRIAN5 et PENBR5 pour le volet 2002). Ces poids doivent être utilisés dans les analyses portant sur ces 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 échantillonnaires, 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.

Les poids associés aux différents jeux ont été ajustés pour tenir compte de la non-réponse aux jeux. Ces poids ne tiennent toutefois pas compte de la non-réponse aux instruments autoadministrés, par exemple. 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. Tout comme en présence de non-réponse partielle, 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, parmi les 1 944 enfants répondants au QIRI au volet 2002, on compte 1 779 enfants ayant un poids pour l'activité sur les nombres du volet 2002. Parmi ceux-ci, on compte 1 729 enfants dont la mère a complété le QAAM. Ainsi, dans une analyse impliquant à la fois des données du jeu des nombres et du QAAM du volet 2002, si l'on utilisait les poids du jeu des nombres, on compterait moins de 5 % de la non-réponse pour laquelle aucun ajustement des poids n'a été fait, ce qui est acceptable. Toutefois, si l'on ajoutait également des données provenant des instruments autoadministrés des volets précédents, la situation pourrait devenir problématique. Dans un tel cas, il est recommandé de faire une analyse du biais en étudiant, par exemple, le sens attendu de ce dernier.

Chacun des fichiers compte également une variable permettant d'identifier les enfants « inaptes », c'est-à-dire souffrant de troubles chroniques déclarés au QIRI tels que l'autisme, l'incapacité mentale ou un trouble envahissant du développement

(INAPTE_E4 et INAPTE_E5). Ces variables prennent la valeur « 1 » pour les enfants « inaptes »; autrement, ils prennent 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.

On retrouve également dans les fichiers de pondération les variables indiquant l'état de passation des différents jeux (DVCRETAT, DIPOETAT, EVCRETAT, ETRIETAT et ENBRETAT). Aucune variable de ce type n'était disponible ni pour le jeu des mots, ni pour la tâche d'imitation de placement d'objets des volets 1999 et 2000. Ces variables permettent d'identifier les enfants de code de réponse « incapable » (code « 4 ») et qui ont été considérés comme répondants. Tout comme pour les enfants « inaptes », ces enfants doivent se voir imputer les résultats les plus faibles aux valeurs manquantes.

4. Référence bibliographique

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