

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATION D'UNE FONCTION  
DE PRODUCTION DE RÉSULTATS COGNITIFS  
CHEZ LES ENFANTS ÂGÉS ENTRE 7 ET 11 ANS : PRATIQUES PARENTALES  
ET CONTRIBUTION DU MILIEU SCOLAIRE

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN SCIENCES ÉCONOMIQUES

PAR  
MARTIN DESROSIERS-DROLET

NOVEMBRE 2006

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je tiens à remercier messieurs Pierre Lefebvre et Philip Merrigan, mes deux directeurs de mémoire, pour m'avoir judicieusement encadré lors de l'élaboration de cette présente recherche. Leurs conseils théoriques et pratiques m'ont été fort utiles. Je remercie aussi le CIQSS, pour m'avoir fourni un soutien financier par le biais de leur subvention du CRSHC, ainsi que Franck Larouche et Susan Carrothers pour leur support technique indispensable tout au long de mon analyse empirique. Par ailleurs, je ne peux laisser sous silence la contribution du fond FARE pour sa bourse d'excellence. J'aimerais également souligner l'apport financier de mes directeurs qui m'ont soutenu durant la transition entre mes études de bacc. et celles de maîtrise grâce aux fonds de recherche provenant du FQRSC. Enfin, j'aimerais remercier tout particulièrement mes collègues du groupe de recherche en politiques sociales du RHDSC pour leurs précieux conseils quant à l'analyse de données empiriques.

En terminant, mes remerciements vont à ma mère, à ma copine et au reste de ma famille, mais plus spécialement à mon père, Guy Drolet, pour m'avoir soutenu et m'avoir encouragé durant mes études

## TABLE DES MATIÈRES

LISTE DES FIGURES.....	v
LISTE DES TABLEAUX.....	vi
INTRODUCTION.....	1
CHAPITRE I	
REVUE DE LA LITTÉRATURE.....	4
1.1 Étude théorique.....	4
1.2 Études empiriques.....	8
1.2.1 Estimation d'une fonction de production d'outputs cognitifs.....	8
1.2.2 L'importance de la taille de la classe.....	11
1.2.3 La relation entre la performance en mathématiques et le rendement sur le marché du travail.....	15
CHAPITRE II	
MODÉLISATION.....	19
2.1 Cadre théorique.....	19
2.2 Modèles analytiques.....	23
2.2.1 La spécification contemporaine.....	23
2.2.2 La spécification cumulative (effets fixes ou aléatoires).....	24
2.2.3 Modèle à valeur ajoutée.....	28
2.2.4 Les effets d'élasticité.....	29
CHAPITRE III	
LES DONNÉES.....	31
3.1 La source des données.....	31
3.2 Caractéristiques des données.....	38
CHAPITRE IV	
ESTIMATIONS ET RÉSULTATS.....	47
4.1 La spécification contemporaine.....	48
4.2 Le modèle à valeur ajoutée.....	55

4.3 La spécification cumulative.....	59
4.4 Les élasticités.....	62
CONCLUSION.....	65
BIBLIOGRAPHIE.....	69
ANNEXE A	
SCHÉMATISATION DE L'ELNEJ.....	71
ANNEXE B	
AUTRES RÉSULTATS.....	76

## LISTE DES FIGURES

1.1 Liens causals entre le revenu d'emploi à l'âge adulte et les différent environnements de l'enfant.....	17
2.1 La fonction de production de résultats cognitifs.....	20
2.2 Niveau d'inputs optimal.....	22
3.1 Moyenne du score en mathématiques d'un enfant par rapport à son SFR.....	44
3.2 Score de mathématiques de l'enfant par région.....	45
3.3 Évolution du score de mathématiques de l'enfant en fonction de son âge.....	46
A-1 Contenu de l'ELNEJ.....	71
A-2 Organisation de l'ELNEJ.....	73
A-3 Répartition de l'échantillon entre les différentes provinces et territoires du Canada.....	74
A-4 Structure et composition de l'ELNEJ.....	75

## LISTE DES TABLEAUX

3.1	Nombre d'observations des échantillons des cycles 1 à 4.....	39
3.2	Nombre d'enfants ayant passé le test de Mathématiques par âge pour les 4 cycles.....	40
3.3	Moyennes et écarts-types des inputs familiaux et scolaires pour les cycles 1 à 4.....	42
3.4	Moyennes et écarts-types des variables de contrôle pour les cycles 1 à 4.....	43
4.1	Régressions transversales pour les cycles 1 à 4 (inputs familiaux et scolaires).....	54
4.2	Régressions pour les modèles à valeur ajoutée à un retard et à deux retards (inputs familiaux et scolaires).....	56
4.3	Tests des modèles à valeur ajoutée à un retard et à deux retards.....	58
4.4	Résultats du modèle avec effets fixes.....	63
4.5	Élasticité des inputs familiaux et scolaires pour les spécifications contemporaine et cumulative.....	64
B-1	Régressions transversales pour les cycles 1 à 4 (variables de contrôle).....	76
B-2	Régressions des modèles à valeur ajoutée à un retard et à deux retards (variables de contrôle).....	77
B-3	Test de Hausman (comparaison entre FE et RE).....	78
B-4	Résultats du modèle avec effets aléatoires.....	79

## INTRODUCTION

Le raisonnement économique suppose que les individus maximisent leur bien-être et ce, en optimisant leur niveau de préférences tout en tenant compte des multiples contraintes auxquelles ils font face. Ces contraintes découlent directement du phénomène de la rareté. En effet, les individus ont des besoins illimités, mais les ressources pour les combler sont, la plupart du temps, limitées. C'est pour cette raison que tout agent économique doit se confronter, un jour ou l'autre, à des choix déchirants entre différentes options. Néanmoins, toute personne rationnelle fera un choix optimal. Que ce soit le gouvernement, une entreprise ou un individu, chacun d'eux prendra la décision qui maximisera son utilité.

Durant la dernière décennie, la littérature économique a été parsemée de plusieurs débats concernant le rôle de l'école et de la famille dans le développement des enfants. Les économistes se sont divisés en deux branches bien distinctes. D'un côté, il y a les fervents de l'environnement familial, qui croient fortement que l'éducation préscolaire qu'apportent les parents est plus qu'essentielle au développement cognitif de leurs enfants. De l'autre côté, il y a les économistes qui croient plutôt que c'est l'éducation en milieu scolaire qui contribue davantage au développement des habiletés et des connaissances des enfants. La plupart des études empiriques qui traitent de ce sujet ont été réalisées aux États-Unis et leurs conclusions ne sont pas convergentes, de sorte qu'il est difficile de s'y fier pour orienter nos décisions en matière de politiques publiques.

Au printemps 2005, les enseignants des différentes commissions scolaires de la province de Québec ont débrayé devant les bureaux du Parlement afin de modifier le programme d'éducation en vigueur<sup>1</sup> et également dans le but d'améliorer leurs conditions de travail. D'un autre point de vue, le gouvernement incite les parents à se départir de l'éducation de leurs propres enfants en introduisant des politiques comme celle de la réduction considérable des tarifs de services de garde en 1997. Suite à cette baisse de tarification, le nombre d'enfants en milieu de garde a crû de façon considérable. Ceci montre

---

<sup>1</sup> Une des demandes des enseignants était de diminuer le ratio élèves-professeur.

clairement que plusieurs parents qui n'avaient pas nécessairement besoin de ce service ont commencé à l'utiliser intensivement. Il y a donc eu création d'un incitatif à la désappropriation de l'éducation des enfants, par leurs parents, vers un environnement hors-familial. En somme, les professeurs revendiquent pour obtenir davantage de ressources scolaires et le gouvernement incite les parents à éduquer de moins en moins leur progéniture. Or, qu'en est-il de l'éducation en milieu familial? Le gouvernement devrait-il allouer moins de ressources au milieu scolaire et se concentrer plutôt sur des politiques visant à accentuer l'apport des parents quant au développement de leurs enfants? À cet égard, la réponse à ces questions se trouve dans la solution du problème soulevé au tout début. Autrement dit, le gouvernement doit maximiser le bien-être des enfants en déterminant la quantité optimale de ressources familiales et scolaires qui contribuera au développement cognitif de l'enfant.

Bien sûr, les opinions divergent quant aux réponses à ces questions, mais ce thème est d'une très grande importance et sera au cœur des débats politiques au cours des prochaines années. En effet, les enfants d'aujourd'hui représentent la main d'œuvre de demain. Il faut donc tenter de déterminer précisément les choix efficaces à effectuer afin de maximiser le développement cognitif des jeunes canadiens.

Le principal objectif de ce texte est de mesurer la contribution de la famille, ainsi que celle du milieu scolaire quant au développement des aptitudes et des connaissances de l'enfant et ce, simultanément. En effet, plusieurs études ont analysé l'apport des inputs familiaux et des inputs scolaires, mais toujours dans une perspective unilatérale. Avec l'aide d'une enquête très élaborée réalisée par Statistique Canada (l'ELNEJ), on est en mesure de faire une analyse qui combine ces deux facteurs et par le fait même, de tirer des conclusions beaucoup plus intéressantes et pertinentes pour d'éventuelles décisions politiques.

Ce texte est divisé de la façon suivante. Au chapitre I, je fais un survol de la littérature afin de faire ressortir l'essentiel des études antérieures qui traite du sujet tout en mettant l'accent sur leurs forces et leurs faiblesses. Ensuite, au chapitre II, je spécifie les modèles empiriques qui serviront à l'analyse des résultats tout en me basant sur les notions

théoriques sous-jacentes. Par la suite, au chapitre III, je fais une brève description de la source de données utilisée et en illustre par le fait même les caractéristiques essentielles, par le biais de statistiques descriptives. L'avant dernière section, soit le chapitre IV, est constituée de l'interprétation des résultats provenant des différentes régressions effectuées lors de l'analyse empirique. Enfin, je conclus et suggère quelques améliorations à apporter pour d'éventuels projets de recherche, compte tenu des limites qu'impose la base de données de l'ELNEJ.

## CHAPITRE I

### REVUE DE LA LITTÉRATURE

Au cours des dernières années, plusieurs économistes se sont penchés sur le phénomène du développement de l'enfance. Cependant, leurs opinions se divisent en deux catégories bien distinctes : "The early childhood development" (ECD) et "The education production function" (EPF). La première s'intéresse au rôle des parents et à l'impact de l'environnement familial sur la production d'outputs cognitifs chez les enfants. L'autre catégorie de chercheurs examine plutôt les relations qui existent entre les inputs scolaires et les résultats aux différents tests cognitifs au niveau des élèves du primaire. La principale caractéristique de ces scientifiques est le caractère unilatéral de leurs études. En effet, ceux-ci se concentrent uniquement sur un des deux facteurs susceptibles d'influencer le développement des enfants. Rares sont ceux qui ont combiné inputs familiaux et scolaires dans leur analyse du développement cognitif des enfants. Or, cette faiblesse s'explique principalement par le manque de données dans les enquêtes avec lesquelles les chercheurs travaillent. La section suivante décrit brièvement les études des auteurs qui ont tenté d'avoir une vision plus globale de la problématique, ainsi que celles des auteurs qui favorisent uniquement un type d'inputs.

#### 1.1 ÉTUDE THÉORIQUE

Tout d'abord, Todd et Wolpin (2003) proposent un cadre théorique qui permettrait d'analyser correctement le développement cognitif des enfants en incluant les deux catégories de facteurs d'influence. Dans leur étude, les auteurs mentionnent la principale raison qui explique pourquoi les autres auteurs ont eu tant de difficulté à introduire les deux types d'inputs dans leur analyse : le manque de données. Autrement dit, les économistes du ECD utilisaient des bases de données riches en information sur la famille, mais pauvres en données sur les caractéristiques de l'école. Ceux qui s'intéressaient à la branche EPF rencontraient le même problème, mais à l'inverse. Plus précisément, les bases de données contenaient peu ou pas de données sur le passé de l'enfant, de sorte qu'il était pratiquement impossible pour les

chercheurs d'utiliser des variables retardées dans leurs modèles afin d'amener un aspect dynamique dans leur analyse.

Le principal objectif de leur analyse est de spécifier et d'estimer une fonction de production du développement cognitif des enfants, tout en étant cohérent avec la notion théorique suivante: le développement de l'enfant est un processus cumulatif qui dépend de l'historique des inputs familiaux et scolaires, ainsi que des dotations intellectuelles que l'enfant possède à la naissance<sup>2</sup>. Selon eux, pour tenir compte des différentes limites des données, il faut :

- i) identifier correctement les hypothèses;
- ii) vérifier si celles-ci sont plausibles et cohérentes avec le modèle défini;
- iii) s'assurer d'avoir les données nécessaires à l'utilisation d'un estimateur en particulier;
- iv) définir les conditions sous lesquelles les hypothèses du modèle peuvent être testées.

Todd et Wolpin discutent également du fait que les études empiriques se divisent en deux catégories: expérimentale et non expérimentale. Dans les études expérimentales, les divers inputs sont choisis aléatoirement, ce qui veut dire qu'ils ne dépendent pas des choix effectués par les parents ou par l'établissement scolaire de l'enfant. Or, dans les études non expérimentales, c'est tout le contraire. Les inputs associés au processus éducatif de l'enfant sont directement liés aux choix des parents ou à ceux des directions scolaires. Cette notion est importante, car très souvent, certains estimateurs peuvent être endogènes et créer un biais considérable quant à l'estimation des autres régresseurs. De plus, ces inputs peuvent dépendre directement de facteurs inobservables, comme l'intelligence de l'enfant. Par exemple, les ressources que les parents vont allouer à leur enfant peuvent être fonction de

---

<sup>2</sup> Todd P. et Wolpin K. (2003).

l'anticipation des capacités mentales que ce dernier possède. Autrement dit, il y a violation de l'hypothèse de moyenne conditionnelle nulle<sup>3</sup>.

L'essentiel de leurs idées se retrouvent dans trois types de modèles économétriques. Dans ces modèles, la production d'outputs cognitifs se mesure par l'addition d'inputs passés et courants combinés à la dotation intellectuelle innée de l'enfant. La fonction de production prend la forme suivante :

$$T_{ija} = T_a [F_{ij(a)}, S_{ij(a)}, \mu_{ij0}, \varepsilon_{ija}], \quad (1.1)$$

où  $T_{ija}$  représente le résultat au test de l'enfant  $i$  qui habite le ménage  $j$  et est d'âge  $a$ . Le vecteur  $F_{ij(a)}$  comprend les différents inputs familiaux, le vecteur  $S_{ij(a)}$  contient les ressources scolaires, la capacité mentale de l'enfant est dénotée par  $\mu$  et finalement,  $\varepsilon$  représente le terme d'erreur.

Le premier modèle utilisé par les auteurs est la spécification contemporaine. Cette dernière relie les résultats aux tests cognitifs uniquement aux inputs courants. Elle fait également l'hypothèse que les inputs ne varient pas à travers le temps, donc que les inputs courants captent l'effet des inputs passés. De plus, ces inputs ne sont pas corrélés avec la dotation cognitive de l'enfant non observable ( $\mu_{ij0}$ ), de sorte que le modèle se modifie et devient :

$$T_{ija} = T_a (F_{ija}, S_{ija}) + \varepsilon_{ija} \quad (1.2)$$

Ensuite, il y a le modèle à valeur ajoutée. Celui-ci n'inclut pas d'inputs retardés, mais introduit plutôt un retard au niveau du résultat aux tests des années antérieures. Cet estimateur, s'il se révèle significatif, a pour fonction de capter tous les effets des inputs

---

<sup>3</sup> L'hypothèse de moyenne conditionnelle nulle stipule que  $E(\varepsilon_i | X_i) = 0$ . En d'autres termes, il n'existe pas de corrélation entre le terme d'erreur et les variables indépendantes.

familiaux et scolaires passés de l'enfant et par le fait même, de combler pour le manque de données ou la nature inobservable de certaines variables. Or, cette nouvelle intrusion est susceptible de créer un biais d'endogénéité. Sans compter qu'elle émet de fortes restrictions sur la fonction de production. Ce modèle est donné par l'équation suivante :

$$T_{ija} = T_a [ X_{ija}, T_{a-1}(X_{ij(a-1)}), \mu_{ij0}, \eta_{ija} ], \quad (1.3)$$

où  $X_{ija}$  représente le vecteur d'inputs familiaux et scolaires de l'année courante,  $T_{a-1}$  est le résultat au test pour l'année précédente et finalement,  $\eta$  est le nouveau terme d'erreur.

Enfin, les auteurs décrivent le modèle cumulatif. Ce dernier suppose que toutes les données courantes et passées sont disponibles, mais que la dotation intellectuelle de l'enfant ( $\mu$ ) est constante à travers le temps, mais n'est pas observable. Todd et Wolpin mentionnent toutefois que les inputs varient avec l'âge de l'enfant. Ceci permet de calculer des estimateurs à effets fixes. En effet d'un côté, on prétend qu'il peut y avoir de la corrélation entre le terme  $\mu$  et les variables explicatives observées et d'un autre côté, on est en mesure d'éliminer l'élément constant  $\mu$ , grâce à la nature longitudinale des données. De cette façon, on arrive à enrayer le problème d'omission de variables, par conséquent, on évite d'obtenir des coefficients biaisés dans nos estimations. Par ailleurs, il est possible d'obtenir des résultats similaires si certaines familles interrogées comptent plus d'un enfant. En effet, on peut alors différencier les résultats entre frères et sœurs. Voici l'équation qui décrit ce type de modèle :

$$T_{ija} = X_{ija} \alpha_1 + X_{ij(a-1)} \alpha_2 + \dots + X_{ij1} \alpha_a + \beta \mu_{ij0} + \varepsilon_{ij(a)}. \quad (1.4)$$

En somme, toutes ces spécifications permettent d'illustrer les effets de substitution ou de complémentarité entre les inputs familiaux et scolaires.

## 1.2 ÉTUDES EMPIRIQUES

### 1.2.1 L'estimation d'une fonction de production d'outputs cognitifs

Premièrement, dans leur analyse empirique, Todd et Wolpin (2004) essaient de mettre en application les notions théoriques qu'ils ont développées dans le passé. Plus précisément, ils étudient les déterminants des résultats obtenus par des enfants à deux tests, mathématiques et lecture. En effet, les auteurs mentionnent que ces résultats, mesurés avant l'âge de sept ans, sont positivement corrélés avec l'éventuel succès de l'enfant sur le marché du travail<sup>4</sup>. D'où l'importance d'analyser l'impact des différents déterminants.

Todd et Wolpin considèrent trois types d'inputs pertinents, qu'ils nomment « premarket factors ». Ce sont l'influence de la famille, l'influence de l'école et la dotation innée d'habiletés et de connaissances que possède l'enfant. Ces trois facteurs représentent, d'une certaine manière, des investissements réalisés par les parents ou par le gouvernement. Or, il faut souligner que ces investissements peuvent dépendre des anticipations quant au futur succès de l'enfant lorsqu'il intégrera le marché du travail à l'âge adulte. En d'autres termes, il faut rester vigilant lors de l'analyse empirique, car ces variables ne sont pas totalement exogènes. Quoi qu'il en soit, avec l'aide de ces inputs, les auteurs estiment une fonction de production d'outputs cognitifs. Celle-ci est cohérente avec le modèle économique qui stipule que le développement cognitif d'un enfant dépend des facteurs antérieurs et contemporains relatifs à son éducation.

Ils font leur analyse en utilisant les trois mêmes types de spécifications économétriques qu'ils ont élaborés dans leur étude théorique afin d'estimer une fonction de production d'outputs cognitifs, soit :

- La spécification contemporaine :  $T_{ija} = T_a (F_{ija}, S_{ija}) + \varepsilon'_{ija}$  (1.5)

---

<sup>4</sup> Todd P. et Wolpin K. (2004).

- Le modèle à valeur ajoutée :  $T_{ija} = T_a [X_{ija}, T_{a-1}(X_{ij(a-1)}), \mu_{ij0}, \eta_{ija}]$ , (1.6)

- L'effet cumulatif :  $T_{ija} = X_{ija} \alpha_1 + X_{ij(a-1)} \alpha_2 + \dots + X_{ij1} \alpha_a + \beta \mu_{ij0} + \varepsilon_{ij(a)}$ . (1.7)

Todd et Wolpin se servent de la base de données NLSY79-CS. C'est un échantillon qui regroupe des milliers d'enfants vivant aux États-Unis. Les enquêteurs récoltent de l'information sur l'enfant (sexe, âge, poids à la naissance, fratrie, etc.), ainsi que sur la mère de celui-ci (scolarité, emploi, statut marital, etc.). Les résultats que l'enfant a obtenus (dans le temps) à des tests de mathématiques (PIAT-M) et à de lecture (PIAT-R) mesurent l'output de la fonction de production associée au développement cognitif de l'enfant. De plus, la NLSY79-CS comprend également une série de questions sur l'environnement familial du jeune. C'est à l'aide de ces éléments que les auteurs construisent la variable « home inputs ». Pour ce qui est des inputs scolaires, les données proviennent du Common Core Data (CCD) qui est un fichier contenant de l'information sur les caractéristiques des écoles dans l'État où habite l'enfant. Cependant, l'exactitude de cette information laisse à désirer, car les données sont agrégées et non reliées directement à l'enfant. Ainsi, les différentes spécifications sont susceptibles de contenir des erreurs de mesure et d'omettre des variables pertinentes, ce qui a pour effet de biaiser les coefficients obtenus dans l'estimation. En effet, puisque les auteurs détiennent uniquement de l'information sur la moyenne des salaires des enseignants et sur la moyenne du ratio pupitres-enseignant, il est évident que l'effet des inputs scolaires sera probablement sous-estimé. Somme toute, les auteurs font face à trois grandes difficultés en ce qui concerne l'échantillon avec lequel ils travaillent : la dotation intellectuelle de l'enfant est inobservable, l'ensemble de données sur les inputs est incomplet et les inputs peuvent être endogènes par rapport à l'intelligence de l'enfant.

Suite à l'évaluation des hypothèses d'identification<sup>5</sup> des trois modèles définis ci-dessus, les auteurs établissent l'équation avec laquelle ils travailleront, soit la spécification cumulative suivante :

$$T_a = \alpha_1 \text{current\_home}_a + \alpha_2 \text{lag\_home}_a I(\text{age} \geq 8) + \alpha_3 \text{lag\_lag\_avg}_a I(\text{age} \geq 10) + \alpha_4 \text{home}_{3-5} + \alpha_5 \text{home}_{0-2} + \delta \text{school\_inputs}_a + X_a \beta + \varepsilon_a, \quad (1.8)$$

où le vecteur  $X$  correspond aux variables de contrôle incluses dans la spécification. Les estimateurs des variables d'inputs familiaux en bas âge (0-2ans et 3-5ans) sont dissociés des autres inputs familiaux, car la formulation de leurs questions d'enquête les rend incompatibles.

Le  $T_a$ , qui mesure l'output cognitif, peut prendre différentes formes. Dans un premier temps, il mesure le résultat que l'enfant a obtenu à un test de lecture. Ensuite, il mesure le résultat au test de mathématiques. Finalement, il combine les deux premiers résultats pour former le PIAT total.

L'effet des deux catégories d'inputs est semblable d'un test à l'autre. Dans le modèle à effets fixes, qui suit l'enfant à travers le temps, l'effet retardé des inputs familiaux est légèrement supérieur à celui des inputs familiaux courants. Or, les auteurs n'obtiennent pas de résultats pour les « early home inputs », probablement à cause d'un nombre insuffisant d'observations. Pour ce qui est des inputs scolaires, les résultats sont peu concluants. En effet, l'hypothèse nulle est rarement rejetée, de sorte que la valeur des estimateurs n'est pas très significative.

Lorsque les auteurs utilisent les moindres carrés ordinaires, les conclusions sont un peu différentes. Dans l'analyse du PIAT-M (test de math), l'effet courant des inputs familiaux dépasse l'effet retardé de 0,002. Cependant, pour le PIAT-R (test de lecture) et le

---

<sup>5</sup> Les auteurs ont utilisé un test de spécification de Hausman-Wu afin de déterminer le modèle économétrique le plus consistant.

PIAT total (combinaison des deux tests), l'effet retardé reste nettement supérieur. Quant aux inputs scolaires, ceux-ci sont un peu plus significatifs. Leur effet est cependant plus important que ceux des inputs familiaux. Plus le ratio pupitres-enseignant est grand, plus il a un effet négatif sur le développement de l'enfant et ce, plus particulièrement lorsqu'on le mesure à l'aide du PIAT total.

En somme, les auteurs mentionnent que dans la plupart des spécifications, les inputs familiaux sont, de loin, plus significatifs quant aux déterminants influant les résultats des enfants aux divers tests. Todd et Wolpin discutent également du fait qu'il aurait été intéressant d'avoir des données pour les inputs scolaires du même niveau d'agrégation que celui des inputs familiaux. Par ailleurs, la possibilité d'être au prise avec un problème d'endogénéité oblige les auteurs à incorporer les variables instrumentales afin d'être plus rigoureux dans leur analyse. Or, l'ajout de ces instruments ne modifie pas les résultats de façon considérable. Néanmoins, notons que le simple fait de s'interroger sur ces inputs peut permettre de diminuer les écarts quant aux scores dans les tests d'aptitudes et par conséquent, réduire les inégalités salariales entre individus. Enfin, Todd et Wolpin concluent leur thèse en soulignant l'importance de bien déterminer le coût relatif des politiques publiques, car celles-ci se transposent directement dans la décision des ménages et des établissements scolaires et affectent le développement des enfants.

### 1.2.2 L'importance de la taille de la classe

Deuxièmement, dans le texte intitulé "Economic considerations and class size", Krueger (2003) défend l'importance des investissements dans les ressources académiques pour ce qui a trait à la réussite scolaire des enfants. Cette étude diffère des deux premières par sa trajectoire plutôt unilatérale. En effet, l'auteur accorde peu d'importance à l'environnement familial comme déterminant du développement cognitif des enfants. Il concentre ses efforts sur l'influence que la taille de la classe d'un enfant exerce sur sa réussite et sur son futur rendement sur le marché du travail.

Le texte de Krueger est divisé en deux parties. D'une part, il critique fortement une étude réalisée par Hanushek (1997) qui tente de démontrer l'inefficacité de la réduction de la taille des classes dans l'objectif d'améliorer le rendement académique des élèves. D'autre part, il effectue une analyse coûts-bénéfices qui estime le rendement interne de la réduction de la taille des classes.

Krueger reproche à Hanushek<sup>6</sup> son attitude pessimiste quant à l'efficacité des inputs scolaires. De plus, il mentionne que sa méthodologie est incorrecte et très imprécise. En effet, Hanushek analyse l'effet de la taille des classes d'un élève, sur la réussite scolaire de ce dernier et ce, en se fiant aux résultats obtenus dans neuf études qui traitent directement ou indirectement du sujet. Krueger l'accuse dans la façon dont il détermine les poids accordés aux estimateurs extraits des différentes études. Il juge que la procédure utilisée met l'emphase sur la négativité et le faible niveau de signification des coefficients associés aux classes à petite taille. Il mentionne également que la qualité des études retenues par Hanushek laisse à désirer. En effet, les échantillons d'analyse sont souvent trop petits et les modèles sont parfois mal spécifiés. Tout au long du texte, l'auteur argumente clairement ses propos et conclut en stipulant que plus la quantité de ressources par pupitre sera grande, plus grand sera le rendement académique des élèves concernés.

La deuxième partie du texte est consacrée à l'analyse coûts-bénéfices associée au rendement interne d'une réduction de la taille des classes. Plus précisément, l'auteur tente de répondre à la question suivante : quelle amélioration dans la performance des élèves peut justifier les coûts de la réduction de la taille des classes ? En d'autres termes, Krueger tente de calculer un taux de rendement interne. Cette analyse s'appuie sur le modèle de Lazear (2001). Celui-ci met l'accent sur le fait que les classes à petite taille bénéficient de plus de temps d'apprentissage, car il y a moins d'élèves susceptibles d'être perturbants durant le temps de classe. Évidemment, ceci est valide si les perturbations des enfants suspendent temporairement l'enseignement du professeur. Lazear mentionne que pour éliminer le plus

---

<sup>6</sup> Hanushek E. A. (1997), dans cet article, cet auteur s'oppose farouchement à la réduction de la taille des classes.

possible cette externalité négative, il faut optimiser la taille de la classe<sup>7</sup>. Autrement dit, il faut la réduire à un point tel que le bénéfice associé à cette réduction soit égal à son coût (bénéfice marginal = coût marginal). Ensuite, Krueger cite quelques exemples tirés de diverses études qui montrent la relation entre le rendement d'élèves à des tests de mathématiques et de lecture et le revenu qu'ils obtiendront sur le marché du travail. Par exemple, les élèves qui se classent dans le quartile le plus fort suite à un examen de lecture, gagneront 20% de plus que ceux qui se situent dans le quartile le plus faible. Pour ce qui est du test de mathématiques, c'est de l'ordre de 19%. De plus, une augmentation d'un écart-type du score aux tests dans les écoles primaires serait associée à des revenus futurs d'environ 8% plus élevés.

Krueger fait tout de même une mise en garde sur le fait qu'il est difficile de quantifier de façon précise les bénéfices attendus des rendements de la réduction de la taille des classes. Autrement dit, il peut être ardu de déterminer quelle sera l'augmentation de salaire (liée directement à la modification du statu quo) d'un individu qui a bénéficié d'une réduction de la taille de sa classe lorsqu'il était à l'école primaire. D'un autre côté, les dépenses encourues pour ces restructurations sont facilement visibles, car elles s'effectuent au moment même où la décision est prise. Par exemple, si le gouvernement du Québec décide de réduire la taille des classes au primaire à 15 élèves par enseignant, parce qu'il juge que cette politique est valable et qu'elle procurera de nombreux avantages dans le futur, les contribuables ne seront peut-être pas du même avis si cette réduction est accompagnée d'une facture de plusieurs centaines de millions de dollars. Par la suite, l'auteur détermine alors les valeurs présentes respectives des coûts et bénéfices associés à la réduction de l'effectif des classes :

- PV des coûts =  $\sum_{i=1}^T C_i / (1+r)_i$  ; (1.9)

- PV des bénéfices =  $\sum_{i=1}^T E_i \times \beta (\delta M + \delta R) / (1+r)$  ; (1.10)

---

<sup>7</sup> Lazear E. P., (2001)

où  $C_t$  représente le coût de la réduction de la taille d'une classe au temps  $t$ ,  $E_t$  est le revenu réel moyen suite à l'introduction de l'élève sur le marché du travail,  $\delta M$  et  $\delta R$  représentent l'augmentation des scores aux tests (mathématiques et lecture) et le  $\beta$  est l'augmentation du revenu associée à une augmentation d'un écart-type dans les scores obtenus à un ou l'autre des deux tests. Enfin, le  $r$  est le taux d'escompte réel. Le taux de rendement optimal<sup>8</sup> se calcule en égalisant les deux équations définies ci-dessus. Concrètement, il suffit de résoudre pour le taux d'escompte ( $r^*$ ) dans l'équation suivante :

$$\sum_{i=1}^T C_t / (1 + r^*) = \sum_{i=1}^T E_t \times \beta (\delta M + \delta R) / (1 + r^*). \quad (1.11)$$

Krueger évoque également d'autres faits très captivants. Par exemple, si le taux d'escompte est établi à 4 %, une réduction de l'effectif d'une classe de 22 à 15 élèves du primaire a pour effet de générer des bénéfices supérieurs de 43 % aux coûts et ce, sans croissance du salaire réel. Si on tient compte d'une croissance d'environ 1 %, les bénéfices deviennent 100 % supérieurs aux coûts.

Tout en demeurant confiant en ses résultats, Krueger fait une petite mise en garde quant à l'interprétation de ces derniers et aux conséquences que peut impliquer leur utilisation dans d'éventuelles décisions politiques. En voici une liste non exhaustive :

- Le coefficient  $\beta$  peut être surestimé, car les ressources scolaires peuvent améliorer des habiletés non cognitives qui, à leur tour, influenceront les revenus futurs de l'enfant. Il faut donc nuancer;
- l'indexation annuelle de [0% à 2%] de la productivité des travailleurs est subjective;
- le coût de réduction d'une classe peut être différent dans les premières années;
- l'analyse coûts-bénéfices ne tient pas compte des effets de redistributions. Ces derniers peuvent s'avérer très positifs pour les moins bien nantis;

---

<sup>8</sup> Équivalent au taux de rendement interne.

- la taille des classes peut être corrélée négativement au taux de criminalité et à la dépendance aux revenus de transfert. Dès lors, une réduction de la taille des classes pourrait amener des externalités positives provenant de l'accroissement du capital humain.

En somme, l'auteur conclut en rappelant que l'attitude pessimiste quant aux résultats relatifs à la réduction de la taille des classes de Hurnashek est malheureusement fondée sur des estimations inexactes et imprécises. Les propos de Kreguer se résument en quelque sorte à ceci : les enfants, inscrits dans des écoles où l'effectif des classes est réduit, tendent à obtenir de meilleurs résultats académiques et par conséquent, ils ont de meilleures aspirations pour ce qui a trait à leur rendement sur le marché du travail. L'auteur renforce son argumentation en mentionnant que les écoles privées sont munies de classes de plus petites tailles et que les parents qui y inscrivent leurs enfants déboursent d'importantes sommes d'argent pour que leurs enfants puissent bénéficier de cet avantage.

### 1.2.3 La relation entre la performance en mathématiques et le rendement sur le marché du travail

Enfin, pour justifier le choix de la variable dépendante du modèle (le résultat au test de mathématiques), il faut également faire part de la littérature qui traite du sujet. Dans leur rapport de recherche s'intitulant "Aptitude or achievement : why do test scores predict educational attainment and earnings ?", Jencks et Phillips (1999) démontrent que les résultats obtenus à des tests de mathématiques au niveau du primaire influencent de façon significative les revenus d'emploi futurs de l'enfant.

Les auteurs commencent par faire la distinction entre deux modèles théoriques qui illustrent les facteurs déterminants du revenu sur le marché du travail. D'un côté, il y a « the aptitude model » qui stipule que les scores aux tests cognitifs sont corrélés avec les gains financiers à la hausse d'un individu, car ils mesurent sa capacité à assimiler facilement et rapidement de nouvelles connaissances. Par conséquent, ces individus sont considérés plus utiles et sont susceptibles de mieux répondre aux besoins des employeurs que les gens qui ne

possèdent pas cette capacité d'apprentissage<sup>9</sup>. De plus, cette explication se base sur le fait que les scores aux tests sont corrélés avec les années d'études que l'individu a complétées. En effet, ceux qui apprennent rapidement et facilement ont tendance à obtenir de meilleures notes, ce qui les motive à rester à l'école plus longtemps que ceux qui ont plus de difficulté dans leur apprentissage. Par exemple, ce modèle prédit qu'une personne qui maîtrise bien le Latin devrait gagner autant qu'une personne qui maîtrise bien l'Algèbre. En d'autres termes, il ne distingue pas la nature des compétences d'un individu, mais plutôt l'effort et le travail ardu que celui-ci a donné au fil du temps. D'un autre côté, il y a « The achievement model » qui défend le principe que ce qui compte réellement, c'est ce qu'une personne connaît ou ce qu'elle sait faire et non son effort d'apprentissage. Par exemple, le score d'un employé à un test de mathématiques peut prédire son salaire, car l'employeur attribue de la valeur à sa capacité à résoudre des problèmes ou raisonner mathématiquement. Autrement dit, ces deux modèles se comparent avec la différence entre la théorie du capital humain (achievement model) et la théorie du signalement (aptitude model). La première assume que les gens consacrent du temps, de l'effort et de l'argent pour acquérir des connaissances mathématiques, car celles-ci leurs permettent d'être plus productifs et de gagner un meilleur salaire chez les employeurs qui recherchent de telles compétences. Or, le signalement expose plutôt le fait que les gens mettent du temps, de l'effort et de l'argent à apprendre, car ces acquis sont une façon de montrer aux employeurs qu'ils possèdent la motivation et la stabilité qu'ils recherchent<sup>10</sup>.

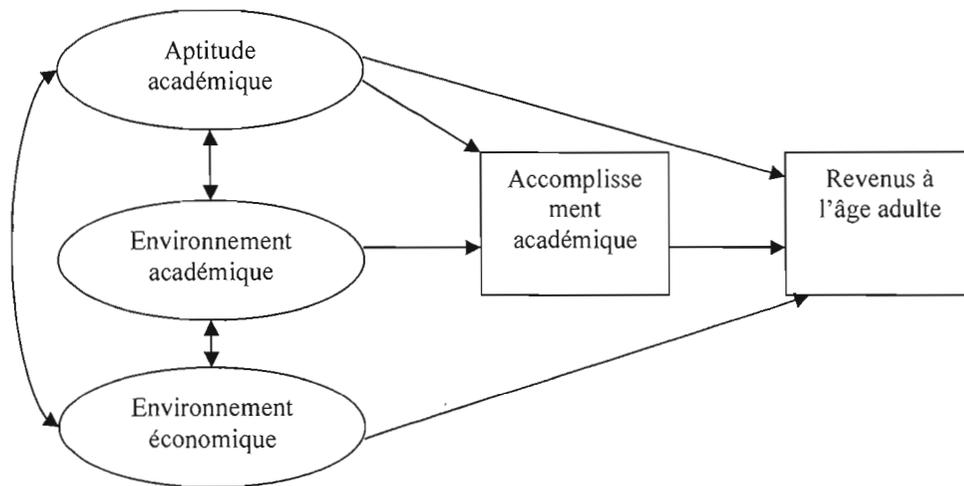
Afin de bien illustrer les effets de l'aptitude et de l'accomplissement (achievement) sur les gains financiers à la hausse à l'âge adulte, Jencks et Phillips présentent une figure qui décrit les liens causals entre les différentes variables.

---

<sup>9</sup> Jencks C. et Phillips M., (1999).

<sup>10</sup> Jencks C. et Phillips M., (1999)

**Figure 1.1 :** *Liens causals entre le revenu d'emploi à l'âge adulte et les différents environnements de l'enfant*



Source : Jencks C. et Phillips M. (1999), figure 2-1.

L'aptitude académique mesure l'effet cumulatif des gènes qui permettent aux individus d'apprendre une certaine quantité d'information lorsqu'ils ont été exposés à un environnement identique. L'intelligence de l'enfant fait partie de cette catégorie. L'environnement académique mesure plutôt l'effet cumulatif de toutes les expériences et circonstances vécues à l'intérieur du milieu scolaire qui ont influencées la performance des individus aux tests. Par exemple, le niveau de compétence de l'enseignant de l'enfant. Ensuite, l'environnement économique mesure l'effet cumulatif associé aux expériences et circonstances qui influencent les revenus indépendamment de l'aptitude et de l'accomplissement académique. Cette catégorie pourrait inclure le statut socio-économique de la famille de l'enfant, le taux de chômage du quartier ou les anticipations quant aux conditions économiques auxquelles l'enfant fera face lors de son introduction dans le marché du travail. Enfin, la réussite représente le score d'un individu aux divers tests cognitifs. La flèche reliant cette dernière à la boîte du revenu à l'âge adulte montre clairement le lien direct qui existe entre ces deux composantes.

Afin de vérifier les fondements théoriques qu'ils ont développés, Jencks et Phillips ont élaboré un modèle analytique servant à estimer le rendement futur des scores aux tests de mathématiques. La relation entre le revenu d'emploi et la réussite scolaire se traduit par cette simple spécification :

$$Y = \beta_0 + \beta_{M_{12}} M_{12} + \beta_{V_{12}} V_{12} + \varepsilon, \quad (1.12)$$

où  $Y$  est le revenu à l'âge adulte,  $M_{12}$  est le résultat d'un élève de 12<sup>ième</sup> année à un test de mathématiques,  $V_{12}$  est le résultat du même élève à un test verbal (vocabulaire et lecture) et  $\varepsilon$  est le terme d'erreur. Les auteurs obtiennent les résultats attendus. Autrement dit, le  $\beta$  associé au test de mathématiques est positif et significatif. De plus, ce coefficient est également positif et significatif lorsque  $Y$  mesure l'éventuelle réussite scolaire de l'élève. Dès lors, cette variable est donc pertinente et peut constituer une bonne mesure du développement cognitif d'un enfant. Il est vrai qu'à première vue, ce modèle semble un peu trop simpliste et pourrait contenir des estimateurs biaisés<sup>11</sup>. À cet égard, les auteurs ont ajouté quelques variables de contrôle et ces dernières n'ont pas modifié la nature de l'interprétation des coefficients originaux.

Cette section a montré de quelle manière les économistes abordent le sujet. De plus, les différents modèles énumérés ci-dessus serviront de références dans l'analyse empirique qui sera effectuée. En d'autres termes, ils permettront de déterminer si leurs conclusions convergent dans le même sens que celles auxquelles je parviendrai.

---

<sup>11</sup> Le biais pourrait être causé par l'omission de variables explicatives non négligeables.

## CHAPITRE II

### LA MODÉLISATION

#### 2.1 CADRE THÉORIQUE

De façon analogue, la fonction de production de résultats cognitifs chez les enfants se compare avec les fonctions de production d'outputs des entreprises de biens et services. En effet, tout comme les entreprises de biens intermédiaires<sup>12</sup>, le développement cognitif de l'enfant représente, en quelque sorte, un bien qui servira à la production d'autres biens en aval dans la chaîne de production. En d'autres termes, le développement intellectuel du jeune peut prendre la forme de capital humain, donc un input plus qu'important dans la plupart des industries. De ce point de vue, les parents du jeune ou le gouvernement chercheront à maximiser l'output, soit le résultat qu'obtiendra l'enfant au test de mathématiques<sup>13</sup>. Le graphique qui suit (page suivante) illustre la fonction de production d'outputs cognitifs  $T_{it}$ <sup>14</sup> qui dépendent de plusieurs inputs (familiaux, scolaires et autres).

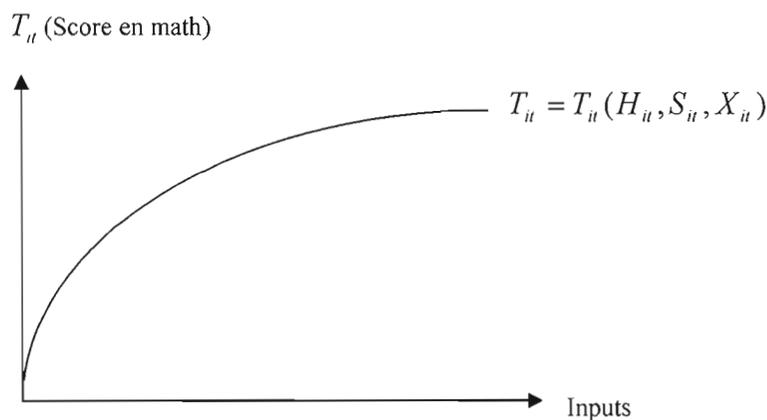
---

<sup>12</sup> Catégorie de biens qui sont vendus en aval à d'autres entreprises qui s'en servent pour fabriquer des biens destinés à la consommation finale.

<sup>13</sup> Comme, il a été mentionné plutôt, le résultat au test de mathématique est corrélé avec le rendement futur sur le marché du travail.

<sup>14</sup> L'indice  $i$  fait référence à l'enfant, alors que l'indice  $t$  fait référence à la période.

**Figure 2.1 :** *La fonction de production<sup>15</sup> de résultats cognitifs*



Il va de soi que chaque unité d'output produit est fonction d'une combinaison d'inputs lors du processus de production. Dès lors, il faut choisir la quantité optimale d'inputs qui maximisera l'output. Dans ce cas-ci, il suffit de choisir les quantités d'inputs familiaux et scolaires qui maximiseront le développement cognitif de l'enfant. Or, comme il a été mentionné au tout début, le phénomène de rareté implique que les agents économiques font face à des contraintes, de sorte que les quantités optimales peuvent être limitées. Notamment, les parents de l'enfant font face d'une part à des contraintes financières, car ils ne peuvent pas attribuer la totalité de leur budget au développement de leur progéniture. D'autre part, ils font face à une contrainte de temps, car ces derniers doivent travailler et avoir du temps de loisir, ils ne peuvent pas être constamment avec leur enfant. Sans compter qu'ils ont peut-être d'autres enfants à éduquer. Pour ce qui est du gouvernement, celui-ci fait face à une contrainte budgétaire, si bien qu'il ne peut pas allouer la totalité de son budget aux ressources scolaires qui favorisent le développement des jeunes. Enfin, le jeune possède une certaine capacité à assimiler des nouvelles connaissances, ce qui peut venir amoindrir les effets positifs d'une augmentation des ressources familiales ou scolaires. Cette dernière contraintes est semblable aux contraintes technologiques auxquelles font face les entreprises productrices

<sup>15</sup> La forme concave de cette fonction démontre qu'après une certaine quantité d'inputs, ces derniers n'influencent plus le développement de l'enfant.

de biens et services. En somme, les parents (ou le gouvernement) doivent résoudre un problème de maximisation sous contraintes, qui se définit comme suit :

$$\underset{SI, HI}{MAX} T_{ii} = T_{ii}(SI_{ii}, HI_{ii}, X_{ii}) \quad \text{tel que :} \quad (1.13)$$

$$P_{ii}^{SI} SI_{ii} + P_{ii}^{HI} HI_{ii} + P_{ii}^C C_{ii} = Y_{ii} \quad (1.14)$$

$$\sigma_{ii}^{SI} + \sigma_{ii}^{HI} + l_{ii} + L_{ii} = 1, \quad (1.15)$$

$$T_{ii} \leq T_{ii}(SI_{ii}, HI_{ii}, X_{ii}). \quad (1.16)$$

Pour résumer, cette expression représente la maximisation de la fonction de production de résultats cognitifs ( $T_{ii}$ ) sous trois types de contraintes. La première étant la contrainte budgétaire. Les parents de l'enfant possèdent un certain revenu familial  $Y_{ii}$  et ils doivent répartir ce montant entre les biens de consommation  $C_{ii}$ , les inputs scolaires  $SI_{ii}$ <sup>16</sup> et les inputs familiaux  $HI_{ii}$ , étant donnés les prix relatifs de chacun de ces éléments. La seconde est la contrainte temporelle. Ainsi, au cours d'une journée, les parents doivent partager leur temps entre le travail  $L_{ii}$ , le loisir  $l_{ii}$  et les deux types d'inputs<sup>17</sup>. Par exemple,  $\sigma_{ii}^{HI}$  pourrait être le temps que le père prend pour jouer au ballon avec son enfant et  $\sigma_{ii}^{SI}$  peut représenter le temps que les parents investissent dans la recherche d'un bon établissement scolaire pour leur enfant. En effet, nous pourrions émettre l'hypothèse qu'une bonne école permettra à l'enfant de recevoir une plus grande quantité d'inputs scolaires. Enfin, la dernière est la contrainte d'habileté. Elle stipule que le résultat au test de mathématiques ( $T_{ii}$ ) doit se situer sur la courbe de la fonction de production<sup>18</sup>. Autrement dit, cette courbe nous fournit l'ensemble des scores réalisables étant données les quantités d'inputs et le niveau d'habileté de l'enfant. Selon Todd et Wolpin (2001), cette habileté innée que l'enfant possède est inobservable, mais elle influence grandement son développement

<sup>16</sup> Supposons, pour le moment, que les dépenses scolaires encourues pour l'enfant sont entièrement assumées par les parents et non par le gouvernement.

<sup>17</sup> La somme du temps disponible dans une journée a été normalisée à 1.

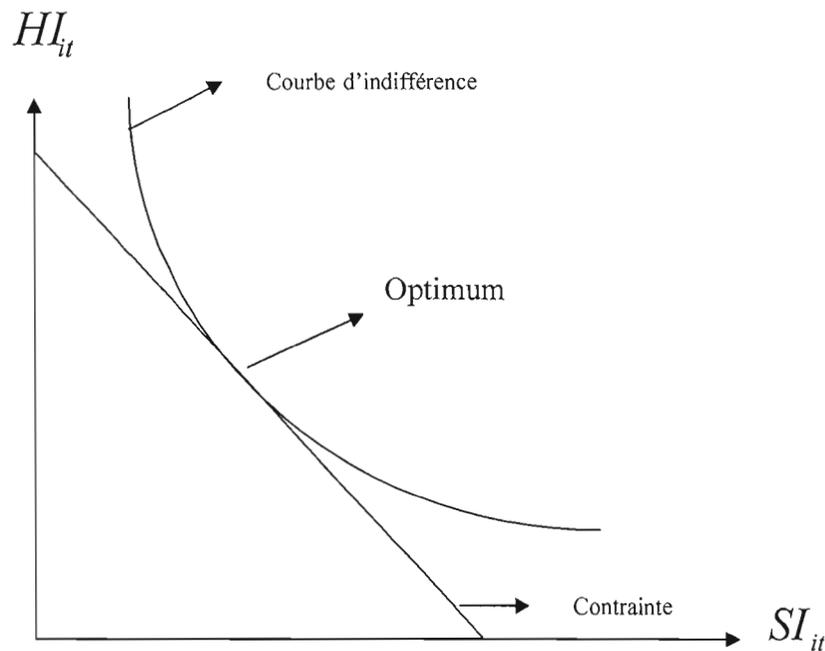
<sup>18</sup> Le résultat peut également se situer en dessous de la courbe, mais ce serait sous-optimal.

cognitif. Donc, ceci détermine, en quelque sorte, la quantité d'information que le jeune sera en mesure d'assimiler. Compte tenu de ce qui précède, il suffit de résoudre le Lagrangien suivant :

$$L = T_{it}(SI_{it}, HI_{it}, X_{it}) + \sum_{j=1}^3 \lambda_j (G_j(SI_{it}, HI_{it}, X_{it})) \quad (1.17)$$

où  $G_j$  représente la fonction qui caractérise les trois contraintes définies plutôt. Le terme  $\lambda_j$  est tout simplement le multiplicateur de Lagrange. Graphiquement, la solution se trouve à l'intersection de la courbe d'indifférence et des différentes contraintes.

**Figure 2.2 :** Niveau d'inputs optimal



Note : Afin de simplifier l'illustration, l'auteur a préféré donner une forme convexe à la courbe d'indifférence et inclure une seule des contraintes.

À l'optimum, le taux marginal de substitution entre les inputs familiaux et scolaires est égal au ratio du prix de ces inputs. C'est-à-dire, combien faut-il sacrifier d'unités d'inputs scolaires pour « consommer » une unité de plus d'input familial ? Autrement dit, cette situation représente un arbitrage entre deux investissements en capital. En effet, ces deux types d'inputs prennent la forme de capital qui générera un rendement dans le futur lorsque l'enfant intégrera le marché du travail. D'une part, les ressources scolaires peuvent être considérées comme du capital physique (l'établissement et ses infrastructures) ou comme du capital humain (les enseignants). D'autre part, les ressources familiales peuvent être perçues comme du capital social<sup>19</sup> (relation entre les parents et les enfants), du capital financier (revenu familial) ou du capital humain (éducation des parents). Voyons maintenant, comment estimer, de différentes façons, les coefficients de la fonction de production de résultats cognitifs définie ci-dessus.

## 2.2 MODÈLES ANALYTIQUES

Les auteurs cités dans la revue littéraire ont leur propre méthode d'analyse. Certains utilisent les MCO, d'autres ont recours aux variables instrumentales et quelques uns empruntent même le modèle à valeur ajoutée. Dans le but d'être soucieux de la précision des résultats, il est nécessaire d'estimer l'impact des deux catégories d'inputs (familiaux et scolaires) à l'aide de plusieurs modèles économétriques.

### 2.2.1. La spécification contemporaine

Premièrement, on peut estimer les coefficients d'intérêt de façon contemporaine en utilisant les moindres carrés ordinaires. En effet, ce type de modèle permet de mesurer l'effet des variables indépendantes sur la variable dépendante pour une période donnée. Par exemple, pour l'année 1994 (cycle 1), on pourrait estimer l'impact des inputs familiaux sur le développement cognitif de l'enfant, tout en conservant les autres variables constantes. En répétant cette formule pour 1998 (cycle 3), on peut observer la différence entre l'effet de

---

<sup>19</sup> Coleman James (1994). Le capital social représente concrètement les inputs familiaux, alors que le capital financier et humain est davantage associé aux variables de contrôle.

certain inputs lorsque l'enfant est âgé de 7 ans et lorsqu'il est âgé de 11 ans, soit deux cycles plus tard. L'équation à régresser prendrait la forme suivante :

$$T_i = \beta_0 + \beta_1 HI_i + \beta_2 SI_i + X_i' \gamma + \varepsilon_i, \quad (1.18)$$

où  $T_i$  est le résultat de l'enfant  $i$  à un test de mathématiques. Le coefficient  $\beta_0$  est la constante du modèle, alors que  $\beta_1$  mesure l'impact des inputs familiaux ( $HI_i$ ) sur le score au test de mathématiques et  $\beta_2$  mesure plutôt l'impact des inputs scolaires ( $SI_i$ ). Le vecteur de coefficients  $\gamma$  associé au vecteur  $X_i'$  estime l'effet de chacune des variables de contrôle qui composent ce vecteur. Enfin,  $\varepsilon_i$  est le terme d'erreur non observé. Dans le but d'obtenir des coefficients précis et sans biais, les deux hypothèses suivantes doivent être respectées :

- *Homogénéité du terme d'erreur* :  $E(\varepsilon_i \varepsilon_j) = 0 \rightarrow$  le terme d'erreur possède la même variance pour tous les individus;
- *Exogénéité contemporaine* :  $E(\varepsilon_i | X_i, HI_i, SI_i) = 0 \rightarrow$  le terme d'erreur n'est pas corrélé avec les variables indépendantes du modèle<sup>20</sup>.

Or, les chercheurs doivent être prudents lorsqu'ils utilisent cette spécification, car celle-ci est justifiée par de fortes hypothèses, souvent difficiles à respecter. Néanmoins, elle s'avère utile lorsque l'information désirée est plus ou moins disponible.

### 2.2.2. Spécification cumulative (effets fixes ou aléatoires)

Deuxièmement, grâce à la nature longitudinale des données, il est possible d'utiliser les méthodes d'analyse pour panels. En effet, puisque pour chaque enfant  $i$ , on observe  $HI$ ,  $SI$ ,  $X$  et  $T$  plusieurs fois à travers le temps. Le modèle de régression (1.18) devient donc :

---

<sup>20</sup> Merrigan P., (2005).

$$T_{it} = \beta_0 + \beta_1 HI_{it} + \beta_2 SI_{it} + X'_{it} \gamma + c_i + \varepsilon_{it} , \quad (1.19)$$

où l'indice  $t$  sert à identifier le cycle auquel on observe les variables. De plus, le terme  $c_i$  s'est ajouté à la régression. Ce dernier ne comporte pas d'indice  $t$ , car il ne varie pas dans le temps (d'un cycle à l'autre). Par exemple, le sexe de l'enfant, l'âge de la mère à la naissance ou tout autre facteurs inobservables qui influencent le développement cognitif de l'enfant. Comme Belzil et Hensen (2002) l'ont mentionné dans leur étude, l'habileté que possède l'enfant dès sa naissance influence grandement son développement. Or, il arrive fréquemment qu'il soit impossible de mesurer cette habileté. Par conséquent, les chercheurs doivent exclure cette variable du modèle. Ceci implique que l'hypothèse d'exogénéité contemporaine n'est pas respectée, de sorte qu'il y a de la corrélation entre le terme d'erreur et les variables explicatives du modèle. C'est pourquoi la spécification contemporaine souffre d'un biais d'omission de variables ou d'endogénéité. En effet, l'habileté  $\mu_i$  fait alors parti du terme d'erreur inobservable, donc :  $E(\varepsilon_i | X_i, HI_i, SI_i) \neq 0$ .

Néanmoins, lorsque les facteurs non observables ne varient pas d'un cycle à l'autre, l'utilisation d'estimateurs à effets aléatoires ou à effets fixes permet d'éviter ces biais et d'obtenir des estimations plus rigoureuses. Ainsi, ceci évite d'avoir de la corrélation entre le terme d'erreur composé ( $c_i + \varepsilon_{it}$ ) et les variables explicatives du modèle, car cette catégorie d'estimateurs permet d'éliminer les effets individuels fixes  $c_i$  après avoir effectué les régressions. En revanche, pour s'assurer de la validité des estimateurs décrits ci-dessus, il est nécessaire d'imposer les conditions suivantes :

- RE (effets aléatoires)
  1.  $E(\varepsilon_i | X_i, HI_i, SI_i, c_i) = 0, \quad t = 1, \dots, T \rightarrow$  exogénéité stricte des variables explicatives conditionnelle à la présence du terme  $c$ .
  2.  $E(c_i | X_i, HI_i, SI_i) = E(c_i) = 0 \rightarrow$  les inputs (familiaux et scolaires), ainsi que les variables de contrôle ne sont pas influencés par les  $c_i$ .

- FE (effets fixes)
  1.  $E(\varepsilon_{it} | X_{it}, HI_{it}, SI_{it}, c_i) = 0$ ,  $\rightarrow$  exogénéité contemporaine des variables explicatives conditionnelle à la présence du terme  $c$ .
  2.  $E(c_i | X_{it}, HI_{it}, SI_{it}) = E(c_i) \neq 0 \rightarrow$  les effets individuels fixes peuvent être corrélés avec les variables explicatives du modèle<sup>21</sup>.

La principale différence entre l'estimateur à effets aléatoires et l'estimateur à effets fixes découle de la deuxième condition, soit l'indépendance entre les  $c_i$  et les variables explicatives du modèle. Lorsqu'on se situe dans un cadre d'analyse expérimentale, l'estimateur RE est plus adéquat, car toutes les variables sont déterminées de façon aléatoire. À l'inverse, en sciences sociales, il est plus probable de se trouver dans un contexte non expérimental, si bien qu'il peut y avoir présence d'endogénéité. Autrement dit, une des variables du modèle peut influencer le choix d'une autre variable. Par exemple, si les parents s'aperçoivent que leur enfant est doté d'une grande intelligence, ils auront tendance à inscrire celui-ci dans les meilleurs établissements scolaires possibles, afin de lui assurer le développement le plus optimal. Or, l'estimateur FE tient compte de cette corrélation et il est en mesure de corriger le biais qu'elle peut provoquer.

Le test de Hausman compare le modèle avec effets fixes et le modèle avec effets aléatoires en déterminant si les coefficients des deux estimations sont statistiquement différents. Autrement dit, il indique quel est le modèle le plus approprié dans une circonstance donnée. La statistique de test suivante compare la matrice de variances-covariances des deux estimateurs :

$$H = (\beta_{FE} - \beta_{RE})' (\text{var}(\beta_{FE} - \beta_{RE}))^{-1} (\beta_{FE} - \beta_{RE}) \rightarrow \text{distribuée sous une } \chi^2. \quad (1.20)$$

---

<sup>21</sup> C'est pour cette raison, que dans le modèle avec effets fixes, le terme  $c_i$  n'est pas considéré comme aléatoire, mais plutôt endogène.

où  $\beta_{FE}$  est le vecteur de coefficients du modèle avec effets fixes et  $\beta_{RE}$  est le vecteur de coefficients pour le modèle avec effets aléatoires. Sous l'hypothèse nulle, les deux estimateurs sont non biaisés. Par conséquent, on peut utiliser l'un ou l'autre des modèles puisque les coefficients estimés seront quasiment les mêmes. À l'inverse, si on rejette l'hypothèse nulle, l'estimateur RE est biaisé et on doit se tourner vers l'estimateur FE pour obtenir des résultats convenables. Or, dans la littérature, la plupart des modèles utilisés pour estimer la fonction de production de résultats cognitifs emploient la méthode avec effets fixes. En effet, la plupart du temps, le modèle avec effets aléatoires est rejeté, car il est pratiquement impossible de respecter l'hypothèse d'exogénéité stricte. De plus, conformément à ce qui précède, les chercheurs en sciences sociales sont presque toujours confrontés à traiter avec des variables endogènes. Par exemple, Todd et Wolpin estime leur spécification cumulative par le biais d'estimateurs avec effets fixes, mais également avec des estimateurs avec effets aléatoires. Or, suite au test de Hausman, il rejette fortement l'utilisation du modèle avec effets aléatoires.

Le modèle cumulatif représente le centre de cette étude. En effet, il tient compte de l'aspect dynamique de la fonction de production et, grâce à l'estimateur FE ou RE, il permet d'éliminer le biais engendré par l'omission de l'habileté innée de l'enfant. Bref, cette spécification est en mesure de fournir des estimateurs beaucoup plus précis. Partant de ce fait, les recommandations en matière de politiques publiques tirées de ces estimations sont d'autant plus rigoureuses et permettent d'atteindre les objectifs de développement social que notre société s'est fixée. En ce qui concerne les enfants, le fait d'obtenir une juste mesure de l'impact de l'environnement familial et scolaire sur leur développement cognitif fait en sorte que le gouvernement possède les outils de décision nécessaires quant à l'élaboration de programmes ou politiques visant à promouvoir le développement intellectuel des jeunes.

### 2.2.3. Modèle à valeur ajoutée

Troisièmement, comme dans l'étude empirique de Todd et Wolpin, il est utile de voir si le modèle à valeur ajoutée est une spécification concevable. Plus précisément, il faut tester le niveau de signification du coefficient associé au résultat que l'enfant a obtenu au test de mathématiques et ce, une année avant l'année courante. En incluant cette nouvelle variable explicative, la spécification contemporaine devient :

$$T_{it} = \beta_0 + \alpha_1 HI_{it} + \delta_1 SI_{it} + \lambda_1 T_{i,t-1} + \beta_1 X_{it} + \varepsilon_{it}, \quad (1.21)$$

où  $T_{i,t-1}$  représente le score au test de l'année antérieure et les autres variables ont la même signification que dans le modèle contemporain original. Ce type de spécification a pour principal objectif de combler les lacunes associées au manque de données sur l'historique des inputs familiaux et scolaires. Par surcroît, l'introduction du score retardé permet également de pallier pour l'omission de variables inobservables. En d'autres termes, cette spécification représente un compromis acceptable au sein de la communauté scientifique. Quoi qu'il en soit, si nous voulons savoir si l'utilisation de ce modèle est adéquate, il faut, dans un premier temps, faire une statistique de test  $t$  (Student), afin de voir si on rejette ou non l'hypothèse nulle suivante :

- $H_0 : \lambda_1 = 0 \rightarrow$  hypothèse nulle;
- $H_1 : \lambda_1 \neq 0 \rightarrow$  hypothèse alternative.

Le rejet de l'hypothèse nulle implique que le résultat au test de mathématiques de l'année antérieure  $T_{i,t-1}$  capte tous les effets retardés qui influencent le score au test en période courante. Dans un deuxième temps, il faut introduire les inputs retardés disponibles et tester leur niveau de signification. Par hypothèse, celles-ci devraient être non significatives si  $\lambda_1$  s'avérerait significatif. En d'autres termes, il est nécessaire de faire une double vérification avant de valider l'utilisation de cette spécification.

Somme toute, si ce modèle s'avère approprié, l'ajout de quelconques variables retardées n'est pas nécessaire et viendrait biaiser le véritable effet des estimateurs de la période courante. En effet, puisque  $T_{i,t-1}$  capte tous les effets antérieurs des inputs, ce dernier comble le manque de données observables dans le passé et permet d'éliminer certains problèmes liés au biais d'omission de variables (inobservables). Par contre, le modèle à valeur ajoutée est susceptible d'être au prise avec des problèmes d'endogénéité et ce, même si les variables omises sont orthogonales aux variables explicatives incluses dans l'analyse. Étant donné que certaines variables inobservées affectent  $T_{it}$ , il va de soi que ces dernières influencent également  $T_{i,t-1}$ . Donc, l'hypothèse d'exogénéité ne peut pas être respectée, puisque le terme d'erreur composée et le score de mathématiques retardé sont corrélés. Ceci aura pour effet non seulement de biaiser le coefficient  $\lambda_1$ , mais aussi ceux des inputs familiaux et scolaires contemporains. Dans l'ensemble, il s'agit de peser le pour et le contre dans le but de déterminer si ce modèle nous donne des résultats rigoureux et justifiables autant que possible.

#### 2.2.4 Les effets d'élasticité

Le calcul de l'élasticité des coefficients des divers modèles de régressions énumérés ci-dessus permet de mesurer la contribution marginale de chacun des inputs. De plus, pour interpréter ces effets marginaux en pourcentage, il suffit de transformer les variables des modèles en forme logarithmique et de calculer leur élasticité. Algébriquement, les élasticités partielles du score en mathématiques ( $T_{it}$ ) par rapport aux inputs s'obtiennent de la façon suivante :

$$\eta_{T,HI} \approx \frac{\partial \ln \bar{T}_{it}}{\partial \ln \bar{HI}_{it}} \Rightarrow \text{Élasticité du } T_{it} \text{ par rapport aux inputs familiaux,} \quad (1.22)$$

$$\eta_{T,SI} \approx \frac{\partial \ln \bar{T}_{it}}{\partial \ln \bar{SI}_{it}} \Rightarrow \text{Élasticité du } T_{it} \text{ par rapport aux inputs scolaires.} \quad (1.23)$$

En d'autres termes, ces élasticités partielles<sup>22</sup> nous permettent de mesurer le pourcentage de variation de la variable dépendante par rapport au pourcentage de variation des variables indépendantes. Par exemple, si  $\eta_{T,HI} = 1$ , ceci implique qu'une augmentation de 5 % de la quantité d'inputs familiaux sera suivie d'une augmentation de 5 % du score en mathématiques. Dans cette éventualité, il est question d'élasticité unitaire. Or, si  $\eta_{T,HI} \rightarrow \infty$  ou  $\eta_{T,HI} \rightarrow 0$ , alors nous sommes plutôt en présence de parfaite élasticité ou de parfaite inélasticité. En somme,  $\eta_{T,HI}$  et  $\eta_{T,SI}$  mesurent le degré de sensibilité de la «quantité offerte» de résultats cognitifs lorsque la quantité d'inputs varie.

---

<sup>22</sup> Ces élasticités sont calculées à partir des valeurs moyennes des variables.

## CHAPITRE III

### LES DONNÉES

#### 3.1 LA SOURCE DE DONNÉES

La principale source de données provient de l'enquête longitudinale sur les enfants et les jeunes (ELNEJ). Cette enquête, réalisée par Statistique Canada, a pour objectif à long terme de suivre le développement et le bien-être d'enfants canadiens et ce, de leur naissance jusqu'au début de l'âge adulte<sup>23</sup>. Cette enquête a débuté en 1994 et elle est toujours active. Elle est principalement utile dans l'analyse des facteurs contribuant au développement social, émotionnel et cognitif des enfants et des jeunes. Par exemple, le comportement des enfants, leur développement physique, leur environnement social et familial, leur cheminement académique, etc<sup>24</sup>.

La structure longitudinale de l'ELNEJ permet de suivre un enfant à travers le temps. En effet, les enquêteurs chargés de recueillir l'information reviennent interroger le même enfant (ou sa famille) tous les deux ans. Autrement dit, la base de données est divisée en cycles regroupés sous forme de fichiers. Jusqu'à maintenant, l'ELNEJ en est à son cinquième cycle. La figure A-2 de l'annexe A illustre clairement l'organisation de cette base de données.

L'ELNEJ enquête sur les comportements des jeunes de 0 à 11 ans. Évidemment, certains sont plus âgés (jusqu'à 19 ans) lors des cycles subséquents. Cependant, comme dans la plupart des enquêtes, il y a de l'attrition, de sorte que le nombre d'observations diminue d'un cycle à l'autre. Or, pour remédier à la situation, il y a eu ajout de nouvelles cohortes d'enfants à compter du cycle 2, ainsi que pour les cycles suivants. Toutefois, ces enfants ne sont suivis que jusqu'à l'âge de 5 ans. L'analyse empirique présentée dans le chapitre IV ne s'adresse pas à la population totale d'enfants de 0 à 11 ans, mais bien à un échantillon représentatif de cette population. Par conséquent, des poids transversaux et longitudinaux ont été incorporés dans

---

<sup>23</sup> [www.statcan.ca/cgi-bin/imdb](http://www.statcan.ca/cgi-bin/imdb)

<sup>24</sup> La figure A-1 de l'annexe A énumère tous les éléments d'information disponibles dans la base de données de l'ELNEJ.

l'analyse afin d'obtenir des estimations fiables et représentatives de la population. Compte tenu de ce qui précède, il importe de décrire comment les enquêteurs ont procédé lors de leur échantillonnage. Statistique Canada fournit les explications suivantes relatives à son échantillonnage :

l'ELNEJ est une enquête longitudinale composée de plusieurs échantillons longitudinaux et transversaux. Les échantillons longitudinaux sont représentatifs des populations longitudinales initiales (c'est-à-dire les populations au moment de la sélection des échantillons). Des poids transversaux sont fournis quand une cohorte d'âge peut aussi être considérée comme représentative d'une population transversale. La plupart des échantillons ont été pris à même l'échantillon des ménages répondants de l'Enquête sur la population active (EPA) et à l'aide des données du registre des naissances des provinces. Le plan d'échantillonnage relatif aux enfants sélectionnés à partir des registres de naissances est le suivant : chaque province a été divisée en une strate rurale et une strate urbaine. Un échantillon aléatoire simple a été sélectionné dans la strate rurale et un plan à deux degrés a été utilisé pour la strate urbaine. Au premier degré, on a tiré un échantillon de régions géographiques, et au second, on a tiré un échantillon d'enfants dans chacune de ces régions<sup>25</sup>.

La figure A-3 en annexe montre la répartition de l'échantillon entre les différentes provinces et territoires du Canada.

La structure de l'ELNEJ est particulièrement complexe<sup>26</sup>. Afin de mieux comprendre son fonctionnement, il est utile d'en faire la décomposition. Avant tout, les enquêteurs identifient le ménage dans lequel l'enfant interrogé habite. Lorsque cela est fait, les sondeurs de Statistique Canada sont alors en mesure de créer un dossier pour chaque ménage<sup>27</sup>. Ce dossier est composé de plusieurs éléments d'information. Premièrement, il y a un questionnaire pour les parents<sup>28</sup>. Celui-ci vise à recueillir de l'information sur leur éducation, leur santé, leurs habitudes de vie, leurs revenus, etc. Ensuite, il y a un questionnaire pour l'enfant. Toutefois, il s'adresse principalement aux parents du jeune, puisqu'il tente d'obtenir l'information la plus objective possible. En revanche, si le jeune est âgé de 10 ans et plus, il aura à remplir un

---

<sup>25</sup> Extrait tiré du document électronique d'information sur l'ELNEJ. Site web : [www.statcan.ca](http://www.statcan.ca)

<sup>26</sup> La figure A-4 de l'annexe A présente un schéma illustrant les différents éléments qui composent chacun des dossiers de l'ELNEJ.

<sup>27</sup> Il est possible qu'il y ait plus d'un enfant interrogé par ménage.

<sup>28</sup> Le terme « parents » sera utilisé tout au long du texte pour faciliter sa lecture. Pour être plus précis, nous devrions plutôt utiliser le terme « personne la mieux renseignée au sujet de l'enfant ».

questionnaire par lui-même. Celui-ci permet de recueillir des éléments ignorés par les parents. Enfin, un questionnaire est administré à l'enseignant de l'enfant (si ce dernier fréquente l'école), ainsi qu'à la direction de l'école. Ceci permet d'obtenir des renseignements sur l'environnement académique du jeune. Outre les questionnaires, le dossier du ménage comprend différents types de tests que l'enfant doit subir. Premièrement, si l'enfant est âgée de 4 ou 5 ans, il doit passer le test d'Échelle vocabulaire en image Peabody. Ensuite, si l'enfant est en deuxième année du primaire ou plus, il doit passer un test d'habileté en mathématiques<sup>29</sup>. Toutes ces informations provenant de différentes personnes liées directement ou indirectement à l'enfant sont indispensables pour une analyse précise et cohérente de l'évolution cognitive du jeune. En ce qui concerne la présente étude, les caractéristiques suivantes seront retenues :

- caractéristiques physiques de l'enfant;
- caractéristiques familiales;
- caractéristiques géographiques;
- caractéristiques des parents;
- résultats à des tests de mathématiques;
- relations et fonctionnement de la famille;
- caractéristiques relatives au milieu scolaire.

Ces éléments servent à mesurer l'impact des inputs familiaux et scolaires sur les résultats obtenus au test de mathématiques. Les autres caractéristiques sont utilisées, comme variables de contrôle, afin de réduire la corrélation entre les deux types d'inputs et le terme d'erreur. Or, comme dans la plupart des études faites sur ce sujet d'analyse, un problème majeur se pose : les informations manquantes. En effet, il arrive fréquemment que certains répondants ne soient pas en mesure de répondre aux questions, car ils ne détiennent pas l'information s'y rattachant. De plus, la question posée peut, dans certains cas, ne pas s'appliquer à la situation de l'enfant. Enfin, il arrive parfois que les individus refusent tout simplement de répondre à la question. En somme, le nombre d'observations est insuffisant

---

<sup>29</sup> Tous ces tests se font sous le consentement des parents de l'enfant.

pour que je puisse utiliser exactement les modèles de régression présentés au chapitre précédent. En d'autres termes, le nombre d'inputs familiaux ou scolaires utilisés est restreint. Par conséquent, je dois m'en remettre à des spécifications un peu moins précises que voulues.

Étant donnée l'utilisation limitée de certaines variables explicatives, les différents modèles analytiques feront usage des variables disponibles suivantes :

### *Variable dépendante*

La variable MATHSCORE représente l'output de la fonction de production de résultats cognitifs. Elle constitue le résultat échelonné à un test de mathématiques et par ailleurs, cette variable est ajustée selon le niveau de scolarité de l'enfant<sup>30</sup>. Cependant, Statistique Canada mentionne qu'il faut agir avec prudence lorsqu'on analyse les résultats que les enfants ont obtenus au cycle 1, car il se peut que le test de mathématiques ait été trop facile pour certains groupes d'âge. Toutefois, cette petite défaillance a été corrigée pour les cycles subséquents.

### *Variables indépendantes*

Les *inputs familiaux* sont mesurés par cinq scores calculés directement par les statisticiens de l'enquête. Premièrement, il y a la variable SCOREFAM. Celle-ci représente le score de fonctionnement familial. Ce score varie entre 0 et 36 et un score élevé est un indicateur que la famille a un disfonctionnement familial. Deuxièmement, il y a la variable INTERACTIONSCORE. Celle-ci varie entre 0 et 20 et mesure la fréquence d'interactions positives entre les parents et leur enfant. Évidemment, un score élevé indique un plus haut degré d'interactions positives. Troisièmement, la variable INEFFECTIVESCORE révèle le niveau d'hostilité ou l'inefficacité du style parental. Ce score oscille entre 0 et 25, plus le score est élevé, moins les parents sont efficaces dans leur façon d'éduquer leur enfant. Quatrièmement, la variable CONSISTENCYSCORE relate la cohérence du style parental.

---

<sup>30</sup> Le score échelonné est calculé à partir du score brut qui est le nombre de bonnes réponses que l'enfant a obtenu lors du test en question.

Par exemple, si un parent décide de punir son enfant, va-t-il revenir quelques minutes après sur sa décision et retirer du même coup la sentence ? En d'autres termes, ce score nous permet de voir si les parents sont contradictoires ou judicieux. Notons que ce score peut se situer entre 0 et 20. Enfin, la variable *AVERSIVESCORE* indique le niveau de sévérité des parents. Le score varie entre 0 et 20, plus il est élevé, plus les parents sont sévères avec leur enfant.

La plupart des *inputs scolaires* utilisés dans notre analyse empirique sont également mesurés à l'aide de scores dérivés par les responsables de l'enquête. Tout d'abord, la variable *DISCIPLINESCHOOL* est une mesure du climat de discipline qui règne au sein de l'établissement scolaire de l'enfant. Elle varie entre 0 et 16 et plus elle est élevée, plus l'école est disciplinée. Ensuite, *PROFACADSCORE* représente la qualité d'enseignement du professeur de l'enfant. Ce score fluctue entre 0 et 16 et plus le professeur enseigne bien, plus haut sera son score. Cependant, cette variable n'est disponible que pour les deux premiers cycles de l'enquête. Par conséquent, cet input ne sera utilisé que dans les régressions transversales des cycles 1 et 2. Puis, la variable *PARTICIPESCHOOL*<sup>31</sup> nous donne un indice du niveau d'implication des professeurs dans la réussite de leur établissement scolaire. De plus, elle fait état de leur solidarité, ainsi que de leur esprit d'équipe. Cette cohésion peut être très bénéfique pour le succès académique des élèves. Finalement, la variable *SUPPORTESCHOOL* représente le niveau de support administratif sur lequel peuvent compter les enseignants de l'école. Par exemple, les professeurs ont-ils les ressources matérielles nécessaires ou le directeur les appuie-t-il dans leurs tâches ? Ce score varie entre 0 et 20 et une école qui offre un bon support est associée à un score élevé. Mise à part les scores dérivés, les inputs scolaires sont composés également d'autres variables pertinentes comme la quantité de devoirs que l'enseignant exige à ses élèves. Cette variable est divisée en trois catégories, soit entre 0 et 15 minutes de devoirs par jour, entre 15 et 35 minutes ou plus de 45 minutes. Dans le modèle analytique, elle apparaissent respectivement comme : *LOWHW*, *MIDHW* et *HIGHHW*. De plus, il y a la variable *CLASSIZE* représentant la taille de la

---

<sup>31</sup> Ce score varie entre 0 et 28. Plus il est élevé, plus le niveau de solidarité et d'entraide entre les professeurs est élevé.

classe; rappelons que celle-ci est indispensable selon les résultats de Krueger<sup>32</sup>. Enfin, j'ai inclus les variables EXPPROF et EXPRONIV<sup>33</sup> qui représentent respectivement l'expérience de l'enseignant et son expérience dans le niveau scolaire de l'enfant interrogé. Sous prétexte que l'expérience n'agit pas de façon linéaire sur le développement cognitif de l'enfant, le modèle inclut également ces deux dernières variables au carré afin de capter l'effet quadratique<sup>34</sup>.

L'utilisation de *variables de contrôle* sert à éviter des problèmes de biais d'omission des variables et à s'assurer que la spécification fait preuve d'exogénéité. Parmi celles-ci, on compte les variables MALE et FEMALE qui servent à déterminer le sexe de l'enfant. Ensuite, le modèle comprend également l'âge de l'enfant (CHILDAGE) et son âge au carré (CHILDAGE2). Ceci me permet de capter l'effet non linéaire des rendements de l'éducation, comme l'ont démontré Belzil et Hansen (2002) dans leur étude sur l'habileté inobservable et les rendements de l'éducation. Dans un autre ordre d'idées, les diverses régressions incluent la variable DISABLE. Celle-ci est dichotomique et prend la valeur de 1 lorsque l'enfant bénéficie d'une aide spécialisée due au fait qu'il a des troubles émotionnels, physiques ou comportementaux qui l'empêchent de faire ses travaux scolaires normalement. De plus, l'analyse fait également appel aux variables de nature géographique comme la province d'origine de l'enfant, ainsi que la taille de sa population. Par exemple un enfant qui habite au Québec et dans une région où le nombre d'habitants se situe entre 100 000 et 500 000 se verra attribué la valeur 1 aux variables QUEBEC ET MIDURBAN<sup>35</sup>. Parallèlement, la variable IMMIGRANT indique si la mère de l'enfant est née au Canada (IMMIGRANT=0) ou ailleurs dans le monde (IMMIGRANT=1). De surcroît, les différentes spécifications font

---

<sup>32</sup> Krueger A., (2003)

<sup>33</sup> Étant donné que l'expérience du professeur, ainsi que son expérience dans le niveau scolaire de l'enfant sont comptabilisés en mois, celles-ci seront divisées par 100. Cette petite modification permettra d'obtenir des coefficients qui s'interprètent plus facilement.

<sup>34</sup> EXPPROF2 est l'expérience du professeur au carré et EXPRONIV2 est l'expérience du professeur dans le niveau de l'enfant mise au carré.

<sup>35</sup> Les variables de provinces sont définies comme suit : Île du Prince Édouard ou Nouvelle-Écosse ou Nouveau-Brunswick ou Terre-Neuve = MARITIMES, Québec = QUEBEC, Ontario = ONTARIO, Manitoba ou Saskatchewan ou Alberta = PRAIRIES, Colombie-Britannique = BC et autres = OTHERS. De plus la taille de la région est définie de la façon suivante : région rurale = RURAL, région urbaine de moins de 30 000 hab. = SMALLURBAN, région urbaine entre 30 000 et 500 000 hab. = MIDURBAN et région urbaine de plus de 500 000 hab. = BIGURBAN.

état de quelques variables relatives à la famille, comme le nombre d'enfants dans la famille (SIBLINGS) et le statut parental<sup>36</sup>. De plus, selon la littérature économique, les variables relatives aux caractéristiques des parents ont beaucoup d'influence sur le rendement académique des enfants. Donc, j'ai cru nécessaire d'inclure les variables suivantes pour éviter d'omettre des variables indispensables dans la fonction de production définie plus haut : l'âge de la mère<sup>37</sup> à la naissance (PMKAGEBIRTH) et son nombre d'années d'éducation (PMKEDUC). Dans le but de contrôler pour l'année de l'enquête, la spécification cumulative inclut des variables dichotomiques pour chacun des cycles, à l'exception d'un entre eux (CYCLE1), afin d'éviter d'avoir de la colinéarité parfaite. Pour terminer, comme l'ont souligné plusieurs auteurs<sup>38</sup>, le statut socio-économique d'une famille peut avoir des répercussions importantes sur le comportement et le développement des enfants vivant dans celle-ci. Dès lors, il importe d'inclure une mesure de revenu dans la modélisation. C'est ainsi que la variable LOWINCOMEFAM prend la valeur de 1 si le revenu familial avant impôts se situe en dessous du SFR<sup>39</sup> (seuil de faible revenu avant impôts personnels), alors que la variable NORMALINCOMEFAM prend la valeur 1 si le revenu familial avant impôts est plus grand ou égal au SFR. Or, cette dernière variable servira uniquement à différencier les résultats mathématiques des jeunes provenant de statut sociaux différents. C'est plutôt la variable REALAVGINCOME qui sera utilisée dans les régressions transversales et dans le modèle à valeur ajoutée. Celle-ci représente le revenu familial moyen réel<sup>40</sup> sous forme logarithmique. À la différence des deux dernières spécifications, le modèle avec effets fixes fera usage du revenu réel courant, soit la variable LNREALINCOME. Évidemment, le revenu permanent serait évincé par ce type de régression, étant donné qu'on fait l'hypothèse qu'il ne varie pas dans le temps.

<sup>36</sup> Le statut parental est divisé en 4 catégories soit les variables MONOP (famille monoparentale), BIO (2 parents biologiques) et AUTRES (familles reconstituées et autres).

<sup>37</sup> L'emploi du terme « mère » sert à alléger le contenu du texte, car ce terme fait allusion à la personne la plus proche de l'enfant. Or, dans la majorité des cas c'est la mère qui joue ce rôle.

<sup>38</sup> Currie J. & Thomas D., (1999).

<sup>39</sup> Le SFR (Low Income Cut-Off) est un seuil de faible revenu calculé par Statistique Canada. Il représente le seuil en deçà duquel une famille est susceptible de consacrer une part beaucoup plus importante de son revenu au logement, à la nourriture et à l'habillement qu'une famille moyenne.

<sup>40</sup> Cette variable constitue la somme des revenus de la famille de l'enfant pour chacun des cycles divisée par le nombre de cycles où ce dernier est présent dans l'enquête. De plus, ce revenu nominal est divisé par l'indice des prix à la consommation afin d'observer le réel pouvoir d'achat des ménages. Enfin, j'utilise le revenu moyen, car la théorie économique du revenu permanent stipule que les individus ont tendance à lisser leur consommation.

### 3.2 CARACTÉRISTIQUES DES DONNÉES

Afin de bien illustrer les caractéristiques des données, il est nécessaire de faire une analyse descriptive exhaustive<sup>41</sup>. Ceci permet de voir comment ces dernières se comportent et évoluent dans le temps. De plus, il est également possible de déceler des différences entre les jeunes provenant de différentes régions ou de différents statuts socio-économiques. Il importe également de voir l'aspect dynamique des données afin de déterminer si celles-ci varient considérablement avec l'âge de l'enfant.

Tout d'abord, le tableau qui suit permet de comprendre comment l'échantillon final a été déterminé. En effet, pour chacun des cycles, le nombre initial de jeunes interrogés varie. Ceci provient de l'attrition et de l'ajout de nouvelles cohortes de jeunes. Par exemple, le cycle 1 compte 22 831 enfants, alors que le cycle 3 en compte 32 158. Cependant, mon modèle s'adresse plus précisément aux enfants du primaire ayant passé le test de mathématiques de l'ELNEJ. Par conséquent, je dois restreindre l'échantillon initial à un sous-échantillon d'enfants compris entre 7 et 11 ans<sup>42</sup>. Or, l'ajout de nouvelles cohortes n'affecte pas le nombre d'observations que contiennent les sous échantillons finaux des cycles subséquents, puisqu'il s'agit d'ajouts d'enfants non longitudinaux compris entre 0 et 5 ans. Enfin, je conserve seulement les enfants ayant une femme comme PCM (personne qui connaît mieux l'enfant).

---

<sup>41</sup> Il est à noter que les statistiques descriptives ont été pondérées à l'aide d'un poids analytique.

<sup>42</sup> L'échantillon est caractérisé par des enfants de 7 à 11 ans, car la quasi-totalité des enfants commencent à passer le test de mathématique exigé par l'ELNEJ lorsqu'ils sont en deuxième année du primaire. De plus, les enfants de plus de 11 ans qui étaient toujours au primaire n'ont pu être analysés, car certaines questions relatives aux inputs familiaux ne leurs étaient pas posées.

**Tableau 3.1 : Nombre d'observations des échantillons des cycles 1 à 4**

Nombre d'observations / Cycles	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Échantillon avec enfants ayant passé le test de mathématiques*	4 239	5 609	5 143	3 843
Échantillon avec enfants ayant une femme comme PCM	3 898	5 137	4 796	3 573
Échantillon avec enfants de 7 à 11 ans	3 882	3 595	2 743	2 169
Après élimination des données manquantes	3 816	3 474	2 648	2 132
<b>Échantillon final</b>	<b>3 816</b>	<b>3 474</b>	<b>2 648</b>	<b>2 132</b>

\* : Peu importe l'âge de l'enfant

Source : Calculs et manipulations de l'auteur à partir de la base de données de l'ELNEJ

Malgré l'imposition de ces restrictions, certaines variables restent au prise avec le problème de non-réponse. Ceci fait en sorte qu'il m'est impossible de leur attribuer une valeur quelconque, à moins que je n'aie recours à l'imputation. Néanmoins, quelques-unes de ces variables ont des taux de non-réponse extrêmement faibles. Dans ce cas-ci, j'ai tout simplement éliminé les observations non pertinentes, ce qui a eu pour effet de diminuer, une fois de plus, le nombre d'observations résiduelles. Cependant, lorsque le taux de non-réponse était considérable<sup>43</sup>, j'ai dû faire appel à de l'imputation afin d'éviter de perdre une quantité importante d'observations. La méthode d'imputation utilisée est fort simple, si la variable à imputer est continue, je ramène les éléments de non-réponse à la moyenne de la distribution<sup>44</sup>. Si la variable est dichotomique, je crée une autre variable binaire qui indique si

<sup>43</sup> L'imputation a été appliquée aux variables de mon sous-échantillon contenant plus de 50 non répondants, soient les personnes ayant des réponses classées dans les catégories suivantes : SANS OBJET, REFUS, NE SAIT PAS ou NON DÉCLARÉ.

<sup>44</sup> Tout d'abord, je calcule la moyenne de la variable excluant les éléments de non réponse. Ensuite, je remplace les différentes codifications par la valeur de la moyenne de la distribution. Enfin, je crée une variable dichotomique qui prend la valeur de 1 si le répondant se situe dans l'une ou l'autre des catégories de non réponse.

le répondant se situe dans l'une ou l'autre des catégories de non-réponse. Bref, après toutes ces manipulations, voici combien d'enfants par âge ont passé le test d'aptitudes numériques à chacun des cycles de l'enquête.

**Tableau 3.2 : Nombre d'enfants ayant passé le test mathématique par âge pour les 4 cycles**

Âge des enfants	<i>cycle 1</i>	<i>cycle 2</i>	<i>cycle 3</i>	<i>cycle 4</i>
<i>7 ans</i>	619	633	506	458
<i>8 ans</i>	813	749	662	399
<i>9 ans</i>	754	676	444	456
<i>10 ans</i>	820	769	621	414
<i>11 ans</i>	810	647	415	405
Nombre d'observations	<b>3816</b>	<b>3474</b>	<b>2648</b>	<b>2132</b>

Source : Calculs et manipulations de l'auteur à partir de la base de données de l'ELNEJ

Les deux prochains tableaux font état des moyennes et des écarts-types de chacune des variables de la spécification contemporaine et ce, pour les quatre premiers cycles de l'enquête. Cela me permet de voir les principales caractéristiques des variables de mes sous-échantillons transversaux. Ainsi, je suis en mesure de repérer les traits spécifiques de la distribution des données, ce qui m'aide grandement à formuler mes hypothèses quant aux résultats produits par les différentes régressions. Premièrement, le tableau 3.2 nous montre que la variable dépendante MATHSCORE varie considérablement d'un cycle à l'autre. Ceci est peut-être dû au fait que la version du test de mathématiques imposé aux jeunes a changé à travers le temps. Deuxièmement, les inputs familiaux n'ont pas vraiment fluctué pendant cet intervalle. La plupart d'entre eux ne varient que d'un demi-point entre le cycle 1 et le cycle 4. Troisièmement, le même scénario se répète pour les inputs scolaires, sauf pour ce qui est de la quantité de devoirs que l'enseignant donne à ses élèves. En effet, cette catégorie de variables dichotomiques varie énormément, de sorte qu'il est difficile d'établir une tendance quant à la distribution des enfants entre les trois différentes situations. Par exemple, au cycle 1, en moyenne, 29 % des professeurs donnaient moins de 15 minutes de devoirs par soir à

leurs élèves, 62 % d'entre eux en donnaient entre 15 et 45 minutes, alors que seulement 9 % en donnaient pour plus de 45 minutes. Au cycle suivant, la distribution des moyennes devient plutôt 14 %, 21 % et 62 %. Cette volatilité est probablement causée par la proximité entre les différentes situations. Autrement dit, il est facile pour un enseignant de passer d'une catégorie à une autre. Par ailleurs, la majorité des variables de contrôle semblent se comporter de la même manière à travers le temps. C'est le cas entre autre pour la proportion d'enfants de sexe masculin ou féminin. C'est-à-dire que les quatre sous échantillons se composent en moyenne de 49 % de filles et de 51 % de garçons. De plus, l'éducation de la mère des enfants varie très peu, soit entre 12,52 à 12,74 années, ce qui équivaut à peu près à un diplôme d'études secondaires<sup>45</sup>. Pour ce qui est des provinces d'origine des enfants, l'Ontario est celle qui détient la plus grande part du sous-échantillon avec un peu moins de 40 % de la distribution. Les variables définissant la taille de la région où habite l'enfant changent relativement peu, sauf lors du dernier cycle 4 où l'on peut percevoir une certaine migration des individus des milieux ruraux vers les régions plus urbaines. Par exemple, la variable BIGURBAN passe de 37 % à 46 % entre le cycle 3 et le cycle 4. De plus, les familles étant composées d'un ou des enfants vivant avec leurs parents biologiques constituent environ 75 % de la distribution. On retrouve également une quantité relativement faible (moins de 10 %) d'enfants vivant dans des familles reconstituées ou autres. Les enfants natifs d'une mère ayant immigré au Canada représentent en moyenne environ 16-18 % de ces sous-échantillons. Enfin, la moyenne d'âge de l'enfant lors de l'interview ( $\approx$  9ans), ainsi que celle de sa mère à sa naissance ( $\approx$  28 ans) sont assez stables durant cet intervalle de temps. En revanche, la variable dichotomique DISABLE qui prend la valeur de 1 si l'enfant nécessite une aide académique supplémentaire bondit subitement au cycle 4. Ceci pourrait peut-être s'expliquer par l'introduction de programmes gouvernementaux ayant pour objectif de venir en aide aux enfants ayant des difficultés d'apprentissage en milieu scolaire. Finalement, la proportion d'enfants provenant de familles qui vivent sous le seuil de faible revenu a chuté considérablement à partir du troisième cycle, soit en 1998. Durant cette période, le revenu réel courant passe de 52 891\$ à 59 628\$. Parallèlement, au Canada, cette période a été marquée par une forte croissance

---

<sup>45</sup> Pour la province de Québec, c'est plutôt une année de plus qu'un diplôme d'études secondaires étant donné que le système d'éducation de cette province est légèrement différent de ceux des autres provinces canadiennes.

économique et des taux de chômage relativement bas. Ce fait peut nous aider à comprendre pourquoi le pourcentage de personnes vivant sous le SFR a tant diminué entre 1996 et 1998.

**Tableau 3.3 : Moyennes et écarts-types des inputs familiaux et scolaires pour les cycles 1 à 4**

	<i>cycle 1</i>	<i>cycle 2</i>	<i>cycle 3</i>	<i>cycle 4</i>
mathscore	419,458 (1.421)	414,127 (1.445)	390,960 (1.529)	381,334 (1.597)
<i>inputs familiaux</i>				
scorefam	8,001 (0.086)	8,205 (0.084)	8,490 (0.096)	8,400 (0.105)
interactionscore	11,812 (0.044)	12,194 (0.045)	11,964 (0.051)	12,445 (0.057)
ineffectivescore	8,985 (0.064)	9,002 (0.066)	8,891 (0.071)	8,583 (0.077)
consistencyscore	15,197 (0.055)	15,125 (0.053)	15,286 (0.059)	15,400 (0.065)
aversivescore	8,887 (0.033)	8,738 (0.033)	8,675 (0.034)	8,373 (0.041)
<i>inputs scolaires</i>				
classize	25,167 (0.071)	25,405 (0.075)	24,766 (0.085)	24,856 (0.086)
expprof	2,096 (0.018)	2,129 (0.019)	1,994 (0.021)	1,881 (0.026)
expprofniv	0,870 (0.013)	0,910 (0.014)	0,803 (0.014)	0,645 (0.015)
expprof2	5,618 (0.077)	5,825 (0.085)	5,198 (0.091)	4,954 (0.108)
expprofniv2	1,359 (0.039)	1,536 (0.050)	1,193 (0.045)	0,869 (0.040)
lowhw	0,293 (0.007)	0,136 (0.006)	0,086 (0.005)	0,063 (0.005)
midhw	0,622 (0.008)	0,205 (0.007)	0,537 (0.010)	0,555 (0.011)
highhw	0,085 (0.005)	0,615 (0.008)	0,247 (0.008)	0,335 (0.010)
profacadscore	12,445 (0.032)	12,766 (0.031)	/	/
participschool	19,204 (0.075)	19,338 (0.079)	19,916 (0.085)	19,942 (0.091)
supportschool	14,132 (0.068)	14,388 (0.069)	14,778 (0.073)	14,453 (0.084)
disciplineschool	10,939 (0.049)	11,205 (0.050)	11,255 (0.052)	11,131 (0.069)
Observations	3816	3474	2648	2132

Écarts-types log-linéarisés entre parenthèses

Source : Calculs de l'auteur à partir des données pondérées de l'ELNEJ

Tableau 3.4 : Moyennes et écarts-types des variables de contrôle pour les cycles 1 à 4

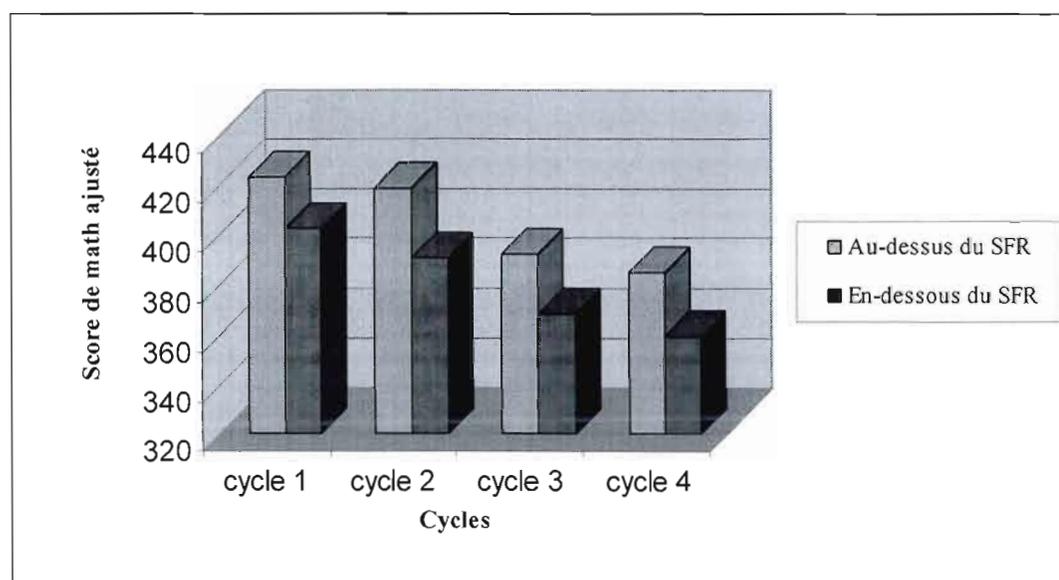
	<i>cycle 1</i>	<i>cycle 2</i>	<i>cycle 3</i>	<i>cycle 4</i>
childage	9,155 (0.022)	9,036 (0.024)	8,950 (0.027)	9,001 (0.029)
childage2	85,726 (0.407)	83,591 (0.428)	81,984 (0.480)	82,873 (0.531)
female	0,492 (0.008)	0,496 (0.008)	0,492 (0.010)	0,475 (0.011)
male	0,508 (0.008)	0,504 (0.008)	0,508 (0.010)	0,525 (0.011)
maritimes	0,112 (0.005)	0,081 (0.005)	0,107 (0.006)	0,073 (0.006)
ontario	0,401 (0.008)	0,388 (0.008)	0,369 (0.009)	0,393 (0.011)
prairies	0,206 (0.007)	0,194 (0.007)	0,199 (0.008)	0,184 (0.008)
bc	0,111 (0.005)	0,141 (0.006)	0,132 (0.007)	0,117 (0.007)
quebec	0,170 (0.006)	0,196 (0.007)	0,194 (0.008)	0,234 (0.009)
bigurban	0,395 (0.008)	0,409 (0.008)	0,369 (0.009)	0,458 (0.011)
midurban	0,269 (0.007)	0,286 (0.008)	0,297 (0.009)	0,268 (0.010)
smallurban	0,130 (0.005)	0,161 (0.006)	0,168 (0.007)	0,140 (0.008)
rural	0,205 (0.007)	0,136 (0.006)	0,155 (0.007)	0,123 (0.007)
monop	0,141 (0.006)	0,164 (0.006)	0,153 (0.007)	0,194 (0.009)
autres	0,094 (0.005)	0,085 (0.005)	0,085 (0.005)	0,092 (0.006)
bio	0,765 (0.007)	0,751 (0.007)	0,763 (0.008)	0,714 (0.010)
disable	0,058 (0.004)	0,086 (0.005)	0,063 (0.005)	0,168 (0.008)
pmkagebirth	27,372 (0.076)	28,111 (0.082)	28,565 (0.098)	29,213 (0.106)
pmkeduc	12,518 (0.037)	12,648 (0.035)	12,660 (0.039)	12,742 (0.043)
immigrant	0,176 (0.006)	0,178 (0.006)	0,155 (0.007)	0,162 (0.008)
lowincomefam	0,166 (0.006)	0,178 (0.006)	0,119 (0.006)	0,129 (0.007)
normalincomefam	0,834 (0.006)	0,807 (0.007)	0,854 (0.007)	0,871 (0.007)
realincome	53 626,496 (583.354)	52 891,029 (662.316)	59 627,509 (968.437)	59 798,569 (1098.468)
Inrealincome	10,694 (0.010)	10,666 (0.011)	10,793 (0.012)	10,772 (0.014)
Observations	3816	3474	2648	2132

Écarts-types log-linéarisés entre parenthèses

Source : Calculs de l'auteur à partir des données pondérées de l'ELNEJ

La figure qui suit illustre bien le fait que les enfants vivant dans des familles pauvres sont désavantagés par rapport à ceux qui vivent dans des familles plus aisées. En effet, on peut voir, qu'en moyenne, le résultat au test de mathématiques des enfants faisant partie de la catégorie LOWINCOMEFAM est beaucoup plus faible que ceux faisant plutôt partie de l'autre catégorie de revenu familial (NORMALINCOMEFAM). Étant donné que les parents ont un revenu net disponible moins élevé, ils ne sont pas en mesure de fournir les mêmes outils d'apprentissage à leur enfant, car ils sont beaucoup plus contraints financièrement. Par exemple, ce type d'enfant est moins susceptible de fréquenter des écoles privées ou d'avoir plusieurs livres éducatifs à la maison. De plus, la littérature économique nous a souvent démontré que le niveau d'éducation est fortement corrélé avec le revenu. Donc, il se peut fort bien que les mères de ces familles pauvres aient peu d'éducation ou du moins, un niveau d'éducation moyen inférieur à celui des mères un peu plus riches. Par ailleurs, les études antérieures qui traitent de ce sujet ont également démontré que l'éducation de la mère avait une influence fondamentale sur le développement cognitif de leur enfant. Dès lors, ces deux derniers éléments nous expliquent en quelque sorte pourquoi ces jeunes sont moins performants et dans quel genre de cercle vicieux ils se situent.

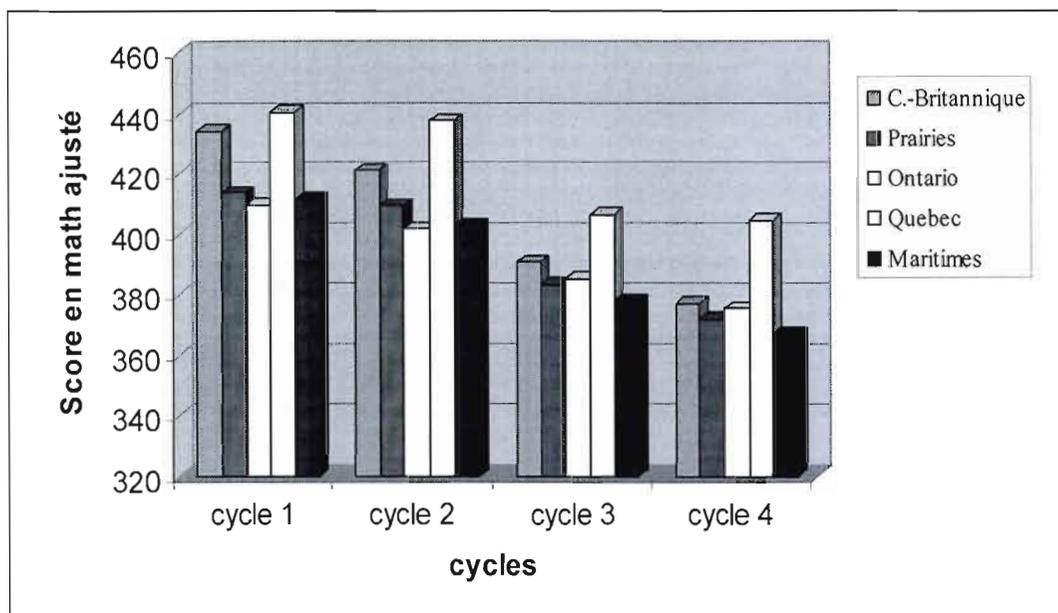
**Figure 3.1 :** *Moyenne du score en mathématiques d'un enfant par rapport à son SFR*



Source : Calculs de l'auteur à partir des données pondérées de l'ELNEJ

La figure 3.2 illustre les différences interprovinciales concernant le score de mathématiques ajusté. D'après celle-ci, la province de Québec détient un net avantage à ce sujet et ce, pour les quatre cycles de l'enquête. À l'inverse, les provinces des Maritimes tirent de l'arrière dans le trois quarts du temps. Ces résultats m'amènent à former des hypothèses pour ce qui est de leurs coefficients respectifs dans les divers modèles de régression qui seront effectués. Si on détermine que l'Ontario est le groupe de base, on peut émettre l'hypothèse que le fait de vivre au Québec aura un effet positif sur la variable dépendante, alors que ce sera plutôt l'inverse pour les Maritimes.

**Figure 3.2 :** *Score de mathématiques de l'enfant par région*

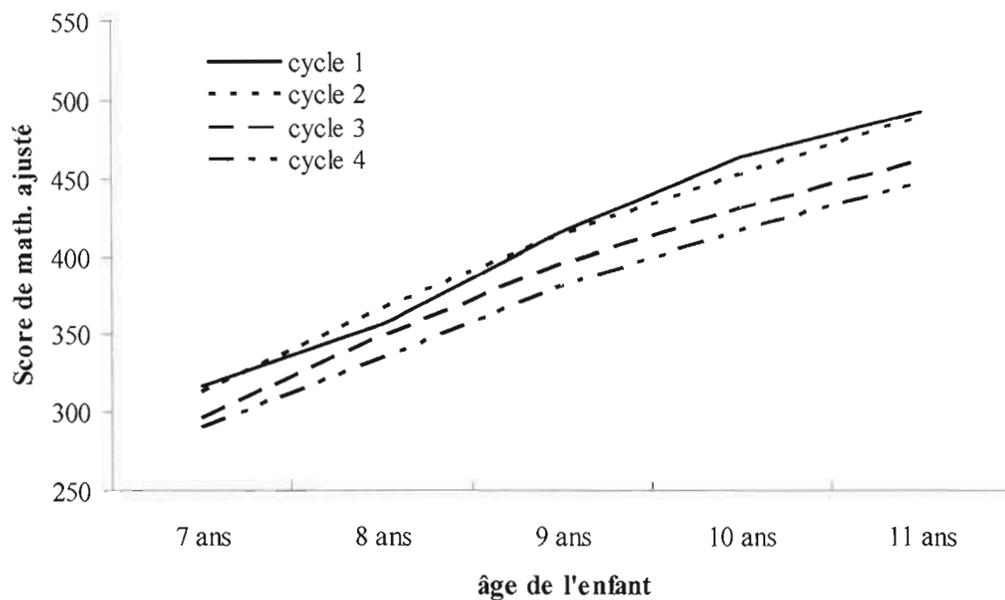


Source : Calculs de l'auteur à partir des données pondérées de l'ELNEJ

Comme il a été mentionné un peu plus haut dans le texte, les rendements de l'éducation ne sont pas constants. Au contraire, Belzil et Hansen (2002) ont trouvé que la courbe reliant le nombre d'années d'éducation et le revenu d'emploi était croissante et convexe, de sorte qu'il était plutôt question d'effets non linéaires. D'une certaine manière, on peut établir une relation semblable entre l'âge de l'enfant et sa performance à des tests cognitifs puisque le niveau d'éducation de l'enfant croît avec son âge et son score en

mathématiques est positivement corrélé avec son rendement futur sur le marché du travail<sup>46</sup>. Donc, on pourrait s'attendre à ce que la courbe du graphique suivant soit également croissante et convexe. Or, c'est plus ou moins le cas, car celle-ci est croissante, mais peu ou pas du tout convexe. Toutefois, l'intervalle de temps est relativement court et il n'est pas le même que celui analysé par Jencks et Phillips. En effet, ceux-ci s'attardaient davantage à des jeunes plus âgés, qui complétaient leurs études secondaires.

**Figure 3.3 :** *Évolution du score de mathématiques de l'enfant en fonction de son âge*



Source : Calculs de l'auteur à partir des données pondérées de l'ELNEJ

Pour conclure, cette section a permis de cibler les éléments essentiels qui décrivent, brièvement, les caractéristiques des enfants de ce sous-échantillon spécifique à ma recherche. Par contre, si on veut faire une analyse plus en profondeur, il ne faut pas s'en tenir qu'à une statistique descriptive, mais on doit plutôt tenter de déterminer l'impact individuel de chacune des variables d'intérêt sur le développement cognitif des enfants compte tenu de la présence d'autres variables. Cela se fait par différents modèles de régression.

<sup>46</sup> Jencks et Phillips, (1999).

## CHAPITRE IV

### ESTIMATIONS ET RÉSULTATS

Le principal objectif de cette étude était de mesurer la contribution de la famille versus celle du milieu scolaire quant au développement cognitif de l'enfant. Les différentes spécifications économétriques élaborées plutôt peuvent s'avérer très efficaces pour ce qui est de la comparaison de l'impact de ces deux types d'inputs. En effet, l'analyse des coefficients des régresseurs permet d'estimer, *ceteris paribus*, l'effet individuel sur le score en mathématiques d'une variation de chacune de ces variables explicatives. Dès lors, nous sommes en mesure de déterminer l'ampleur, le sens de la causalité et le niveau de significativité de chacun des inputs et des variables de contrôle du modèle. Étant donné que toutes les spécifications étudiées sont de formes linéaires<sup>47</sup>, les coefficients agissent à titre de pente de la fonction de  $T_{it}$  définie par  $HI_{it}$ ,  $SI_{it}$  ou  $X'_{it}$ . Par exemple, le paramètre associé à un des inputs scolaires pourrait estimer les deux éléments suivants :

- $\Delta T_{it} = \beta_{SI} \Delta SI_{it} \Rightarrow$  variation du score en mathématiques provoquée par une variation de la quantité d'inputs scolaire, (1.24)

- $\beta_{SI} = \frac{\Delta T_{it}}{\Delta SI_{it}} \Rightarrow$  pente de la fonction. (1.25)

Cette section interprète les coefficients de la fonction de production de résultats cognitifs et ce, pour chacune des trois spécifications, soit le modèle contemporain, le modèle à valeur ajoutée et le modèle cumulatif. Enfin, la dernière sous-section se concentre davantage sur les élasticités des inputs. Ceci permet d'obtenir des variations en pourcentage et par conséquent, d'avoir une idée plus précise de l'amplitude des effets propres à chacune des variables indépendantes.

---

<sup>47</sup> La spécification est linéaire par rapport aux paramètres, mais les variables, elles, peuvent être non linéaires.

#### 4.1 LA SPÉCIFICATION CONTEMPORAINE

Avant de commencer, il faut se rappeler que cette spécification consiste en un MCO pour chacun des quatre cycles de l'enquête. Le tableau 4.1 montre clairement que les résultats diffèrent d'un cycle à l'autre. Autrement dit, les inputs n'ont pas le même impact en 1994 (cycle 1) qu'en 2000 (cycle 4). Il est difficile de comprendre pourquoi la contribution des inputs change autant à travers le temps. Malheureusement, ce type de spécification n'est pas en mesure de fournir une réponse à cette interrogation. De ce fait, je peux seulement émettre quelques hypothèses pour tenter d'y remédier. D'une part, certaines décisions politiques ou événements majeurs peuvent influencer les choix des parents au fil du temps. Ces éléments ne sont pas inclus dans les données et nécessitent une analyse de nature dynamique. D'autre part, certains facteurs inobservables ou certaines erreurs de mesure peuvent perturber les résultats, de sorte que ces imprécisions biaisent l'impact des variables explicatives du modèle. Mises à part ces petites imperfections, la spécification contemporaine fait un portrait biennal de l'impact de l'environnement familial et scolaire sur le développement cognitif de l'enfant.

##### *Les inputs familiaux*

Tout d'abord, le score de fonctionnement familial (SCOREFAM), le score d'inefficience parentale (INEFFECTIVESCORE) et le degré de sévérité des parents (AVERSIVESCORE) sont les inputs familiaux les moins influents. En effet, ceux-ci sont rarement statistiquement significatifs<sup>48</sup>, mais quand ils le sont, leur impact est d'une amplitude semblable aux autres inputs familiaux. Toutefois, deux d'entre eux sont de signe attendu. Par exemple, un dysfonctionnement familial plus élevé a un effet négatif sur le résultat de l'enfant au test de mathématiques. Parallèlement, plus les parents agissent de façon inefficace avec leur enfant, moins le score de l'enfant sera élevé. Contrairement à ce que certains pourraient croire, le degré de sévérité des parents aurait une influence négative sur la performance des jeunes en mathématiques. Ceci semble contre-intuitif, mais s'explique

---

<sup>48</sup> Dorénavant, le terme *significatif* sera utilisé pour désigner qu'une variable est statistiquement significative.

peut-être par un abus de pouvoir parental qui pourrait occasionner un sentiment de répression chez les enfants<sup>49</sup> et par le fait même, créer des troubles de comportement ou d'apprentissage.

Ensuite, le degré de cohérence des parents est significatif à un niveau de confiance de 99 % lors du premier cycle de l'enquête. De plus, son amplitude est nettement supérieure à celle des inputs précédents. Par exemple, une augmentation d'un point du CONSISTENCYSCORE engendrerait une augmentation de 1,21 points au test de mathématiques, ce qui représente un effet au moins deux fois plus important que les inputs précédents. En d'autres termes, les parents se doivent d'être rigoureux dans leurs interventions et éviter de revenir sur leurs décisions. Sinon, le jeune s'apercevra de ce manque de rigueur et pourra en tirer profit, s'en suivra une perte de contrôle pour les parents. Il faut cependant noter que cette variable n'est pas significative pour les cycles 2 et 3 de l'ELNEJ.

Enfin, le tableau 4.1 nous montre clairement que la variable INTERACTIONSORE est l'input familial le plus significatif. En effet, celle-ci l'est pour trois des quatre cycles de l'ELNEJ et à des niveaux de confiance de 95 % et 99 %. Or, le signe du coefficient n'est pas celui attendu. Ainsi, on pourrait s'attendre à ce qu'une augmentation des interactions positives entre les parents et leur enfant ait un impact positif sur le résultat au test de mathématiques, mais les résultats de cette régression ne vont pas dans ce sens. Un accroissement d'un point du score d'interactions provoquerait une diminution du MATHSCORE de l'ordre de 1,18 à 1,75 points. Donc, un impact négatif assez considérable. Par contre, ce résultat contre-intuitif peut être expliqué par un biais d'endogénéité. L'habileté innée que possède l'enfant n'est pas vraiment quantifiable, de sorte qu'il est difficile de l'incorporer dans une analyse empirique. Cependant les parents sont en mesure d'observer le niveau d'habileté de leur enfant. D'un côté, s'il est élevé, ces derniers pourraient être tentés d'investir plus dans l'éducation de l'enfant, sachant que ce sera profitable dans le futur. D'une autre côté, si les parents s'aperçoivent que leur enfant n'est pas habile dès la naissance, ils voudront passer davantage de temps avec lui afin de pallier pour le manque d'aptitudes

---

<sup>49</sup> Cette explication n'est qu'une hypothèse subjective, étant donnée que la psychologie ne fait pas partie des compétences de l'auteur.

intellectuelles. C'est probablement cela qui explique le sens inattendu de la causalité entre le score en mathématiques et ce type d'input familial. Autrement dit, les enfants ayant peu d'habileté à la naissance seraient faibles en mathématiques et nécessiteraient plus d'attention de leurs parents.

### *Les inputs scolaires*

Premièrement, contrairement à ce que dit Krueger<sup>50</sup>, mes résultats ne montrent pas qu'une réduction de la taille des classes a un effet bénéfique sur la performance des jeunes en mathématiques. Au contraire, l'effet marginal<sup>51</sup> de l'augmentation de la taille d'une classe varie entre 0,67 et 1,36. Cet impact positif converge plutôt vers les conclusions des auteurs qui critiquaient les fondements théoriques et les résultats empiriques de Krueger. Cet input n'est pas significatif à une seule reprise, soit au cycle 2.

Deuxièmement, la catégorie d'inputs scolaires construits à partir de scores dérivés directement de l'ELNEJ est rarement significative, à l'exception d'un qui est seulement disponible lors des deux premiers cycles de l'enquête. Il s'agit de la performance pédagogique de l'enseignant de l'enfant (PROFACADSCORE). En effet, le tableau 4.1 illustre bien le fait que cet input scolaire surpasse aisément tous les inputs familiaux en terme d'amplitude. Il est dommage que ce score soit dorénavant exclu de l'enquête, puisque il s'avérait un élément indispensable dans l'analyse du développement de l'enfant. Pour ce qui est des autres scores, seul celui concernant le degré d'implication des professeurs quant à la réussite de leur établissement scolaire est significatif à plus d'une reprise et ce, seulement lors du deuxième et troisième cycle. Toutefois, le signe du coefficient ne va pas dans le sens attendu au cycle 2, de sorte que plus les enseignants s'impliquent dans l'école, moins l'enfant performera en mathématiques. Au cycle 1, le fait qu'une école ait un bon climat de discipline favorise la réussite académique de ses élèves.

---

<sup>50</sup> Krueger A., (2003).

<sup>51</sup> L'effet marginal est la variation d'un élève, cette variable étant de nature discrète.

Troisièmement, la quantité de devoirs que le professeur donne à l'enfant semble avoir un effet bénéfique sur la production de résultats cognitifs. En effet, dans presque 50 % des cas, les variables MIDHW (entre 15 et 45 minutes de devoirs par soir) ET HIGHHW (plus de 45 minutes de devoirs par soir) sont significatives à des niveaux de confiance de 90 %, 95 % et 99 %. Rappelons que ces variables dichotomiques se comparent avec la variable de référence LOWHW (moins de 15 minutes de devoirs par soir). Donc, si on regarde les résultats du tableau 4.1, on peut voir qu'un enfant qui changerait de catégorie, soit de LOWHW à MIDHW, verrait son score en mathématiques bondir de plus de 8 points au cycle 1 et de plus de 9 points au cycle 3. Parallèlement, un enfant qui passerait de la catégorie LOWHW à HIGHHW au cycle 3 aurait un gain encore plus considérable, soit plus de 11 points. Tout compte fait, cet input est de loin le plus important quant à l'accroissement des aptitudes intellectuelles du jeune. Autrement dit, un enfant qui veut performer en mathématiques se doit d'être jumelé avec un professeur qui lui demande un minimum d'effort à la maison, soit au moins 15 minutes de devoirs à chaque soir.

Enfin, la dernière catégorie d'inputs scolaires concerne le niveau d'expérience du professeur. Les variables EXPPROF et EXPPRONIV représentent respectivement le niveau d'expérience globale de l'enseignant et son expérience dans le niveau académique de l'enfant. Ces termes ont été mis au carré pour capter les effets non linéaires de l'expérience. Ce genre d'inputs scolaires ne semble pas être très influent, puisqu'il n'est significatif que dans seulement 25 % des cas. De plus, c'est uniquement au cycle 1 qu'une de ces deux variables et son carré sont significatifs. En effet, une augmentation d'une unité de la variable EXPPROFNIV et EXPPRONIV<sup>2</sup>, soit 100 mois d'expérience de plus, aurait un effet négatif d'environ 7 points<sup>52</sup> sur la performance de l'enfant en mathématiques. Encore une fois, ce résultat est contre-intuitif, puisque l'expérience devrait être un facteur indispensable pour l'apprentissage de l'enfant. L'interprétation des autres paramètres mesurant la contribution de l'expérience s'avère difficile, étant donné que ces variables et leur carré ne sont pas simultanément significatifs.

---

<sup>52</sup> Dans un modèle où une variable et son carré sont inclus, l'effet de cette variable se calcule en additionnant le coefficient de la variable avec le double du coefficient associé au carré de celle-ci.

### *Les variables de contrôle*

Avant tout, il est utile de mentionner que les résultats des régressions transversales qui concernent les variables de contrôle se retrouvent en annexe dans le tableau B.1. L'inclusion de cette catégorie de variables avait pour objectif d'éviter un biais d'omission, alors leur analyse sera plutôt brève, étant donné qu'elles ne représentent pas l'intérêt principal de cette étude. Toutefois, l'impact de celles-ci est parfois beaucoup plus important que celui des inputs familiaux et scolaires.

En premier lieu, le fait d'être une fille ou un garçon ne semble pas être un avantage. En effet, cette variable n'est significative qu'à une seule reprise (cycle 2) et le paramètre d'estimation est de -4,992. Ainsi, durant cette période, les filles sont désavantagées par rapport aux garçons. Ensuite, le fait d'avoir des difficultés d'apprentissage nécessitant une aide pédagogique a un effet très négatif sur la performance de l'enfant. Ainsi, la variable *DISABLE* est très significative et son coefficient varie entre -25,5 et -49,33. Comme on a pu le voir dans la figure 3.3, la performance académique croît avec l'âge. C'est probablement pour cette raison que l'effet net de l'âge de l'enfant est fortement positif et significatif et ce, pour tous les cycles analysés.

En second lieu, les caractéristiques de la famille ne semblent pas avoir d'effets très considérables. Cette catégorie de variables est rarement significative. C'est le cas pour la variable *SIBLINGS*. Par contre, pour ce qui est du troisième cycle de l'enquête, le statut parental semble avoir une certaine influence sur le score en mathématiques. Par exemple, un enfant vivant avec ses deux parents biologiques par rapport à la catégorie de base (familles reconstituées) voit son score augmenter de plus de 16 points. Or, l'impact d'une famille monoparentale est négligeable, car la variable n'est jamais significative.

Ensuite, le tableau B.1 illustre également le fait que l'impact des variables de nature géographique est plus ou moins convaincant. Toutefois, les variables de provinces le sont un peu plus que celles définissant la taille de la région où habite l'enfant. Notamment, le fait d'habiter dans la province du Québec, de la Colombie-Britannique ou dans une des provinces

des Prairies est avantageux par rapport à habiter dans l'Ontario. Cependant, un enfant vivant dans les Maritimes ne semble pas avoir d'avantages nets puisque le coefficient est uniquement significatif au cycle 4 et il est négatif. À propos de la taille des régions, celle-ci ne mène à aucune conclusion intéressante, car elle est, la plupart du temps, non significative.

En dernier lieu, les variables concernant les caractéristiques de la mère de l'enfant contribuent de façon considérable au développement cognitif des jeunes. Par exemple, l'augmentation d'une seule année d'éducation de la mère fait augmenter le score de mathématiques d'environ 3 points. Il en va de même pour l'âge de la mère à la naissance de son enfant. Notons cependant que l'amplitude et la significativité ne sont pas aussi grandes que celles de la variable précédente. En revanche, le revenu familial semble avoir un effet positif nettement plus important. En effet, une augmentation d'un pourcent du logarithme du revenu réel permanent crée une hausse d'environ 6 points du score en mathématiques. Ce résultat soulève donc un problème d'inégalité des chances. C'est-à-dire qu'un enfant vivant dans une famille pauvre a de fortes chances de demeurer dans la même situation que ses parents lorsqu'il fera la transition vers le monde adulte et ce, *ceteris paribus*. Ceci peut justifier, en quelque sorte, l'intervention du gouvernement canadien quant à la lutte à la pauvreté chez les enfants. Enfin, être né d'une mère immigrante n'a pas réellement d'impact sur la performance, puisque cet élément est seulement significatif au cycle 3 et à un niveau de confiance de 90 %. Néanmoins, durant ce cycle, cette variable a une influence positive non négligeable par rapport à un enfant ayant une mère canadienne.

Somme toute, il faut rester vigilant quant à l'interprétation de ces résultats, puisque cette spécification risque fortement d'être biaisée étant donné qu'elle ne peut pas tenir compte de facteurs inobservables comme l'habileté innée de l'enfant. Voyons si les prochaines spécifications sont en mesure d'enrayer ce problème.

Tableau 4.1 : Régressions transversales pour les cycles 1 à 4 (inputs familiaux et scolaires)

	Variable dépendante : mathscore			
	cycle 1	cycle 2	cycle 3	cycle 4
<i>Inputs familiaux</i>				
scorefam	-0,534 (0.293)*	-0,539 (0.332)	-0,427 (0.341)	0,371 (0.392)
interactionscore	-1,178 (0.539)**	-1,815 (0.590)***	-1,747 (0.595)***	-0,671 (0.75)
ineffectivescore	0,047 (0.39)	0,034 (0.483)	0,236 (0.534)	-1,308 (0.552)**
consistencyscore	1,206 (0.408)***	0,130 (0.446)	-0,493 (0.505)	1,077 (0.617)*
aversivescore	-0,108 (0.899)	0,521 (0.803)	-1,655 (0.927)*	0,668 (1.204)
<i>Inputs scolaires</i>				
classize	1,229 (0.360)***	0,231 (0.312)	0,665 (0.323)**	1,362 (0.407)***
profacadscore	2,214 (0.750)***	3,244 (0.768)***	/	/
participschool	-0,167 (0.463)	-0,906 (0.491)*	0,860 (0.509)*	-0,121 (0.627)
supportschool	0,502 (0.488)	0,585 (0.495)	-0,820 (0.562)	0,377 (0.695)
disciplineschool	0,939 (0.565)*	0,128 (0.635)	0,425 (0.641)	-0,732 (0.647)
midhw	8,378 (3.109)***	-5,827 (4.177)	9,092 (4.871)*	0,940 (5.233)
highhw	7,860 (5.733)	2,661 (4.272)	11,975 (5.841)**	-3,533 (6.566)
expprof	6,932 (5.007)	-7,793 (4.926)	-8,835 (5.674)	2,612 (6.316)
expprof2	-0,866 (1.138)	2,680 (1.043)**	2,367 (1.318)*	-1,216 (1.433)
expprofniv	-15,417 (5.832)***	0,241 (3.585)	7,183 (6.363)	6,005 (7.817)
expprofniv2	4,198 (1.778)**	0,454 (0.882)	-0,628 (1.825)	-1,244 (2.497)
Constant	-383,995 (72.032)***	-362,175 (78.041)***	-448,731 (86.219)***	-269,068 (94.018)***
Observations	3816	3474	2648	2132
R-carré	0.62	0.62	0.62	0.62

Valeur absolue des écarts-types robustes entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données pondérées de l'ELNEJ.

#### 4.2 LE MODÈLE À VALEUR AJOUTÉE

Ce type de spécification a pour mission de combler pour le manque de données sur l'historique des inputs familiaux et scolaires. En fait, l'introduction du score de mathématiques retardé capturerait l'impact des inputs retardés omis de la régression. Quoiqu'on en dise, cette hypothèse est forte et par conséquent, on doit la tester deux fois plutôt qu'une afin de faire une vérification rigoureuse. Dans un premier temps, on fait une régression en incluant le score retardé pour voir si le coefficient de celui-ci est significatif. Ensuite, on fait une deuxième régression, cette fois-ci, en incluant les inputs retardés disponibles et on observe leur statistique de test dans le but de déterminer s'ils sont significatifs ou non. C'est cette procédure qui justifie l'utilisation ou le rejet du modèle à valeur ajoutée.

Le tableau 4.2 nous donne les résultats des modèles à valeur ajoutée à un retard et à deux retards<sup>53</sup>. Delà, il est évident de remarquer que le score ( $T_{i,t-1}$ ) passe la première étape de vérification. En effet, le paramètre d'estimation est significatif à un niveau de confiance de 99 %. Toutefois, son amplitude est très faible, de sorte qu'une augmentation d'un point au test de l'année précédente résulte en une augmentation de moins d'un demi point pour ce qui est de l'année courante. Dans un autre ordre d'idées, l'introduction de cette nouvelle variable vient modifier les résultats obtenus pour le cycle 4 avec la spécification contemporaine. Certains inputs familiaux qui n'étaient pas significatifs le deviennent soudainement. C'est le cas pour la variable INTERACTIONSCORE. À l'inverse, la taille de la classe était significative à 99 % avec la spécification précédente, alors que dans le modèle à valeur ajoutée à un retard, elle n'est tout simplement plus significative. C'est plutôt la quantité de devoirs que l'enseignant donne à l'enfant (MIDHW) qui devient significatif, mais, curieusement, ceci influence négativement le résultat de l'enfant. En outre, la plupart des variables de contrôle deviennent non significatives suite à l'introduction de ce score retardé<sup>54</sup>.

<sup>53</sup> Les régressions ont été effectuées pour le cycle 4, à partir du fichier de données longitudinales.

<sup>54</sup> La présentation des coefficients des variables de contrôle des modèles à valeur ajoutée se retrouve en annexe B, dans le tableau B.2.

**Tableau 4.2 : Régressions pour les modèles à valeur ajoutée à un retard et à deux retards  
(inputs familiaux et scolaires)**

	Variable dépendante : mathscore contemporain	
	Modèle à un retard	Modèle à deux retards
<i>Inputs familiaux</i>		
scorefam	0,611 (0.509)	-0,149 (1.048)
interactionscore	-2,594 (1.096)**	0,938 (1.987)
ineffectivescore	-0,537 (0.805)	-2,897 (1.470)*
consistencyscore	2,840 (0.875)***	0,294 (1.617)
aversivescore	-0,342 (1.688)	7,426 (3.699)**
<i>Inputs scolaires</i>		
classize	0,915 (0.597)	1,978 (1.369)
participschool	0,377 (0.758)	-1,081 (1.682)
supportschool	-1,468 (0.818)*	3,005 (1.513)**
disciplineschool	0,431 (0.961)	-0,638 (1.533)
midhw	-26,320 (14.242)*	-55,286 (23.594)**
highhw	-21,614 (14.604)	-61,152 (23.710)**
expprof	9,234 (9.417)	51,433 (16.943)***
expprof2	-2,889 (2.082)	-12,248 (3.819)***
expprofniv	1,669 (9.976)	-31,014 (18.518)*
expprofniv2	2,488 (3.027)	15,798 (5.747)***
mathscore en t-1	0,443 (0.050)***	0,431 (0.110)***
mathscore en t-2	/	0,184 (0.103)*
Constant	190,899 (485.858)	249,31 (138.515)*
Observations	699	145
R-Carré	0.48	0.62

Valeur absolue des écarts-types robustes entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données pondérées de l'ELNEJ.

Dans le désir d'aller plus loin dans l'analyse de ce type de spécification, j'ai tenté de voir l'effet de l'introduction d'un score retardé de deux années, soit  $T_{i,t-2}$ . L'ajout de cette variable est, pour l'instant, approprié, car on peut rejeter l'hypothèse nulle ( $\beta_{i,t-2} = 0$ ). Par contre, le nombre d'observations de l'échantillon n'est pas très satisfaisant. En effet, ce sous-échantillon longitudinal ne compte que 145 enfants ayant passé un test de mathématiques à trois reprises, soit au cycle 4 (année courante), au cycle 3 (1<sup>er</sup> retard) et au cycle 2 (2<sup>ème</sup> retard). Bref, le modèle à valeur ajoutée à deux retards est plus ou moins en mesure d'apporter une quelconque amélioration au modèle contemporain, compte tenu du fait que la quantité limitée d'observations qu'il possède diminue énormément la fiabilité de ses résultats.

Eu égard à ce qui précède, si on veut se pencher en faveur de ce type de modèle, il faut s'assurer qu'il passe avec succès à la deuxième vérification. Le tableau qui suit nous fournit les résultats nécessaires au verdict final. Premièrement, à ce qui a trait au modèle à un retard, il réussit à passer la deuxième étape. Autrement dit, son utilisation est appropriée, car aucun inputs retardés n'est significatif. Donc, le score retardé ( $T_{i,t-1}$ ) capterait l'historique des inputs familiaux et scolaires de l'enfant et par conséquent, il constituerait l'argument principal de la fonction de production intertemporelle de résultats cognitifs. Deuxièmement, le scénario est différent pour ce qui est du modèle à deux retards. On doit rejeter celui-ci, mais pas de façon catégorique. En effet, seulement deux des inputs retardés est significatif, soit le niveau d'implication des professeurs dans l'école de l'enfant (PARTICIP SCHOOL) et le support administratif sur lequel peuvent compter les professeurs (SUPPORT SCHOOL). Toutefois, en incluant ces nouvelles variables retardées, le  $T_{i,t-2}$  n'est désormais plus significatif. Ceci vient confirmer que l'utilisation de ce modèle n'est pas justifiée.

En somme, le modèle à valeur ajoutée à un retard est dans ce cas-ci adéquat, de sorte qu'il permet d'expliquer une partie de la variation du score cognitif. Néanmoins, cet outil économétrique souffre parfois d'un biais d'endogénéité, car les inputs contemporains peuvent être fonction du score de mathématiques retardé. En tant qu'analyste, on se doit d'être conscient des limites qu'entraîne ce genre de spécification.

**Tableau 4.3 : Tests des modèles à valeur ajoutée à un retard et à deux retards**

	Variable dépendante: mathscore contemporain	
	Modèle à un retard	Modèle à deux retards
scorefam en t-1	-0,475 (0.520)	1,216 (1.489)
interactionscore en t-1	0,701 (1.100)	2,401 (2.655)
ineffectivescore en t-1	1,204 (0.867)	0,362 (2.285)
consistencyscore en t-1	-0,103 (0.938)	-1,984 (2.128)
aversivescore en t-1	-2,503 (1.648)	-2,349 (3.705)
classize en t-1	-0,187 (0.572)	1,450 (1.738)
participschool en t-1	-1,275 (0.856)	-2,981 (1.611)*
supportschool en t-1	1,264 (1.069)	3,112 (1.828)*
disciplineschool en t-1	0,986 (0.962)	1,560 (1.910)
midhw en t-1	4,480 (6.699)	-15,000 (19.292)
highhw en t-1	4,384 (8.093)	-6,147 (16.252)
expprof en t-1	-8,118 (9.743)	-15,099 (23.418)
expprof2 en t-1	1,946 (2.214)	2,459 (5.330)
expprofniv en t-1	10,005 (9.909)	12,612 (26.045)
expprofniv2 en t-1	-4,521 (2.908)	-3,637 (8.652)
mathscore en t-1	0,444 (0.050)***	0,360 (0.143)**
mathscore en t-2	/	0,199 (0.134)
Constant	230,366 (463.704)	166,987 (145.326)
Observations	699	145
R-Carré	0.49	0.63

Valeur absolue des écarts-types robustes entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données pondérées de l'ELNEJ.

### 4.3 LA SPÉCIFICATION CUMULATIVE

Comme je l'ai mentionné antérieurement, l'aspect longitudinal de l'ELNEJ permet d'utiliser des méthodes d'analyse pour données en panels. Les paragraphes qui suivent discutent des résultats obtenus en utilisant des modèles avec effets fixes (FE) et aléatoires (RE). Ces deux concepts économétriques ont pour objectif d'éliminer l'effet des variables qui ne varient pas dans le temps. Ceci représente un énorme avantage, puisqu'il est maintenant possible de contrôler pour l'habileté inobservée que possèdent les enfants dès leur naissance. Étant donné que cette caractéristique est innée, elle reste constante et ce, même si l'enfant grandit. Dès lors, il est fort probable que l'hypothèse d'exogénéité contemporaine (ou stricte si on est présence d'effets aléatoires) soit respectée. Si tel est le cas, on vient d'enrayer une partie importante de l'endogénéité causée par des facteurs inobservables qui influencent le développement cognitif des jeunes. Autrement dit, les résultats provenant de ces régressions sont beaucoup plus fiables et représentent davantage la réalité.

Avant de faire l'interprétation des nombreux résultats obtenus par ces deux méthodes d'analyse, il est primordial de faire un test de Hausman afin d'évaluer quelle procédure est la plus appropriée. Le but de ce test est de comparer la matrice de variances-covariances des deux estimateurs afin de déterminer si leurs coefficients sont statistiquement différents. Bref, le rejet de l'hypothèse nulle (les coefficients ne sont pas différents) implique que le modèle avec effets aléatoires n'est pas adéquat et qu'on doit plutôt opter pour l'utilisation du modèle avec effets fixes. Les résultats de ce test sont présentés en annexe, au tableau B-3. La statistique des tests nous montre clairement que l'hypothèse nulle est rejetée, de sorte qu'il est nettement préférable de se servir du modèle FE comme outil d'analyse. Ce résultat était tout de même assez prévisible, étant donné que le cadre de cette analyse est de type non expérimental, ce qui rend l'hypothèse d'exogénéité stricte plus ou moins crédible. En raison de ce qui précède, j'interpréterai uniquement les coefficients du modèle avec effets fixes. Malgré tout, les résultats du modèle avec effets aléatoires sont présentés en annexe, au tableau B-4.

Premièrement, le tableau 4.4 montre que lorsqu'on utilise le modèle avec effets fixes, les inputs familiaux n'ont presque plus d'influence sur la performance des enfants en mathématiques. En effet, les inputs qui avaient un impact lors des régressions transversales ne semblent plus jouer un rôle important dans le développement intellectuel des jeunes. Par exemple, le niveau d'interactions entre les parents et leur enfant avait un effet négatif sur le MATHSCORE. Ceci semblait, à première vue, être contre intuitif, mais s'expliquait par le problème d'endogénéité relié à l'omission de certains facteurs clés inobservables. Désormais, le coefficient de la variable INTERACTIONSCORE est maintenant positif. Par contre, il n'est pas significatif, ce qui me contraint à ne pas pouvoir l'interpréter. Le scénario se répète pour l'input qui mesure le degré de cohérence des parents. Celui-ci demeure positif, mais la statistique de test montre qu'on ne peut pas rejeter l'hypothèse nulle. Toutefois, le niveau d'hostilité des parents ou leur inefficacité dans leur façon d'éduquer leur enfant semble avoir un effet négatif sur la variable dépendante. Une augmentation d'un point de la variable INEFFECTIVESCORE fait diminuer le MATHSCORE de plus d'un demi point. Malgré cela, ces résultats vont légèrement à l'encontre de ce que Todd et Wolpin<sup>55</sup> avaient obtenus dans leur étude, car l'apport des inputs familiaux ne semble pas être aussi prédominante.

Deuxièmement, l'influence de l'environnement scolaire de l'enfant sur sa performance en mathématiques n'est pas plus convaincante. À vrai dire, un seul input affecte le score mathématique de l'enfant, soit le score mesurant le niveau d'implication du professeur de l'enfant dans le bon fonctionnement de l'établissement scolaire. Cet input a une influence positive sur le développement du jeune, mais l'apport est tout de même minime. Si la variable PARTICIPSCHOOL augmente d'une unité, la variable dépendante MATHSCORE s'accroît seulement d'un peu plus d'un demi point (0,622). Cet effet est très faible si on le compare avec l'effet de la quantité de devoirs que l'enseignant donne à ses élèves. À vrai dire, un enfant dont la variable LOWHW changerait pour MIDHW verrait son score de mathématiques majorer de presque 2 points. Or, contrairement à ce que nous retrouvons dans les régressions transversales, on ne peut plus dire que le fait de donner une quantité modérée de devoirs à chaque soir, dans ce cas-ci entre 15 et 45 minutes par soir, procure à l'enfant un avantage considérable par rapport à ceux qui n'en font que 15 minutes

---

<sup>55</sup> Todd & Wolpin, (2004).

et moins par soir, car cette variable n'est plus significative. Enfin, les autres inputs scolaires, soient les scores relatifs à l'école dérivés directement dans l'ELNEJ, la taille de la classe de l'enfant et l'expérience du professeur, ne sont pas significatifs et par conséquent, il m'est impossible de mesurer leur contribution quant au développement cognitif des jeunes.

La partie de droite du tableau 4.4 présente les coefficients des variables de contrôle. Contrairement à la spécification contemporaine, certaines de ces variables ont été retirées de la régression pour différentes raisons. D'une part, certaines d'entre elles ne varient pas ou très peu dans le temps, il est donc inutile de les inclure, puisque le modèle FE va les éliminer de toute manière. C'est le cas pour les variables dichotomiques suivantes : FEMALE, DISABLE, PMKAGEBIRTH et IMMIGRANT. Évidemment, car il est pratiquement impossible de changer de statut (ex. FEMALE devient MALE). D'autre part, certaines variables étaient non significatives<sup>56</sup> à un point tel que les exclure de la spécification ne modifiait pas les coefficients des autres variables incluses. Notamment, le nombre de frères et sœurs que possède l'enfant (SIBLINGS) et les variables définissant le statut parental en sont deux bons exemples. Néanmoins, quelques éléments du vecteur  $X_{it}'$  sont significatifs et ont un effet non négligeable sur la performance de l'enfant. Bien sûr, l'âge de l'enfant est la variable qui a l'effet net le plus important. Ceci confirme les propos de Belzil et Hansen<sup>57</sup> concernant les rendements non linéaires de l'éducation. Étonnamment, le fait de déménager dans une des provinces des Maritimes semble avoir un effet favorable (51,176) sur le score en mathématiques. Ceci est contraire à ce qui ressortait de la spécification contemporaine, où l'impact était plutôt négatif. Je ne vois aucune explication évidente qui pourrait expliquer ce résultat. À l'opposée, l'impact positif qu'amène le fait de déménager dans une grande ville est plus facilement justifiable. En effet, l'enfant peut bénéficier de plus d'infrastructures et de services favorables à son développement. De plus, une grande ville possède un plus grand choix d'établissements scolaires et un plus grand éventail d'activités parascolaires. Tous ces éléments sont susceptibles d'améliorer la performance académique ou intellectuelle de l'enfant. Or, cette variable n'est significative qu'à des niveaux de significativité inférieurs à 90 %, ce que je considère comme un seuil entraînant un niveau de précision insatisfaisant.

---

<sup>56</sup> Je considère qu'une variable est amplement non significative si sa statistique  $t$  est inférieure à 0,5.

<sup>57</sup> Belzil & Hansen, (2002).

Enfin, les variables dichotomiques de cycle sont extrêmement significatives (99% de niveau de confiance) et leur amplitude est quasiment démesurée. D'où la nécessité, voir l'obligation, d'introduire cette catégorie de variables dans un modèle avec effets fixes.

En terminant, même si cette spécification est supposée nous fournir des résultats moins biaisés que ceux des autres modèles, il n'en demeure pas moins que nous nous devons d'être vigilants avant de conclure quoi que ce soit. En effet, même si l'ELNEJ contient une grande quantité d'information sur l'environnement de l'enfant, les inputs familiaux construits à partir des scores dérivés directement par les enquêteurs de Statistique Canada dévient légèrement de la définition exacte d'un input familial. C'est probablement pour cette raison que le modèle avec effets fixes sous-estime l'apport de ce type d'inputs quant au développement cognitif des enfants. Si j'avais été en mesure de récolter de l'information plus précise sur l'environnement familial pour le sous-échantillon défini précédemment, il est fort probable que les coefficients auraient été différents. Par ailleurs, dans un modèle avec effets fixes, pour que les coefficients d'intérêt soient significatifs, il faut être en mesure d'observer une certaine variation intertemporelle au niveau de leur régresseur. Or, comme nous l'avons vu au chapitre précédent, la plupart des inputs familiaux et scolaires sont fixes d'un cycle à l'autre. Cette faible variation explique en partie pourquoi la majorité des coefficients de ces inputs ne sont pas significatifs. En dépit de tout cela, l'information relative à l'environnement scolaire m'a permis de conclure que l'école avait un rôle à jouer dans le développement cognitif des jeunes. C'est à ce niveau que mes résultats diffèrent de ceux de Todd et Wolpin et ceci s'explique en partie grâce à la qualité supérieure des données concernant le milieu scolaire de l'enfant.

#### 4.4 LES ÉLASTICITÉS

Le tableau 4.5 nous montre clairement que la plupart des inputs de cette fonction de production sont essentiellement inélastiques. En effet, tous les paramètres d'élasticité se rapprochent davantage de 0 que de 1 et ce, pour les deux types de spécification. En d'autres termes, une variation de la quantité d'inputs engendre très peu de variation du score cognitif. Par exemple, au cycle 2, une augmentation de 10 % du score relatif à la qualité

d'enseignement du professeur de l'enfant provoque seulement une augmentation de 1,0 % du MATHSCORE, ce qui est extrêmement petit. Donc, si le prix de ces inputs est très élevé, il devient alors primordial de connaître le rendement réel futur de ceux-ci avant de faire des investissements qui peuvent s'avérer non productifs. Bref, cet input représente l'élément qui semble être le plus « élastique ». Vient ensuite la taille de la classe, celle-ci atteint une élasticité de 0,092 au cycle 4. Quoi qu'il en soit, d'après les résultats, il est impossible de déterminer si les inputs familiaux sont plus élastiques que les inputs scolaires.

**Tableau 4.4 : Résultats du modèle avec effets fixes**

Variable indépendante: mathscore contemporain			
inputs		variables de contrôle	
<i>inputs familiaux</i>		childage	85,261
scorefam	-0,237 (0.202)	childage2	(7.970)*** -4,625
interactionscore	0,007 (0.387)	maritimes	(0.372)*** 51,176
ineffectivescore	-0,537 (0.321)*	quebec	(18.842)*** 32,781
consistencyscore	0,421 (0.348)	prairies	(42.308) 31,671
aversivescore	0,657 (0.577)	bc	(19.390) 36,051
<i>inputs scolaires</i>		bigurban	(24.884) 15,182
classize	0,216 (0.202)	midurban	(10.046) -4,056
participschool	0,622 (0.273)**	smallurban	(5.877) 1,752
supportschool	-0,381 (0.291)	pmkeduc	(4.522) 1,191
disciplineschool	-0,230 (0.334)	bio	(1.583) -5,361
midhw	-1,761 (2.044)	monop	(6.509) -3,833
highhw	0,375 (2.351)	realincome	(5.411) -1,883
expprof	0,174 (0.659)	cycle 2	(2.825) 84,451
expprof2	0,915 (2.884)	cycle 3	(8.460)*** 145,051
expprofniv	2,333 (1.996)	cycle4	(16.336)*** 216,940
expprofniv2	-0,331 (0.448)	Constant	(24.718)*** -407,955
Observations	12070		(55.537)***
Nombre de childid	9077		
R-carré	0.71		

Valeur absolue des écarts-types entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données de l'ELNEJ.

**Tableau 4.5 : Élasticités des inputs familiaux et scolaires pour les spécifications contemporaine et cumulative**

	Variable dépendante : <i>mathscore</i>				
	cycle 1	cycle 2	cycle 3	cycle 4	FE
<i>inputs familiaux</i>					
scorefam	<b>-0,010</b>	-0,011	-0,009	0,008	-0,005
interactionscore	<b>-0,033</b>	<b>-0,053</b>	<b>-0,053</b>	-0,022	0,002
ineffectivescore	0,001	0,001	0,005	<b>0,029</b>	<b>-0,012</b>
consistencyscore	<b>0,044</b>	-0,005	-0,019	<b>0,044</b>	0,016
aversivescore	-0,003	0,011	<b>-0,037</b>	0,015	0,014
<i>inputs scolaires</i>					
classize	<b>0,074</b>	0,014	<b>0,042</b>	<b>0,089</b>	0,013
expprof	0,035	-0,040	-0,045	0,013	0,005
expprof2	-0,001	<b>0,038</b>	<b>0,031</b>	-0,016	0,002
expprofniv2	<b>0,014</b>	0,000	-0,019	-0,003	-0,001
expprofniv	<b>-0,032</b>	0,001	0,015	0,010	0,005
midhw	<b>0,012</b>	-0,003	<b>0,012</b>	0,001	-0,002
highhw	0,002	0,004	<b>0,008</b>	-0,003	0,000
profacadscore	<b>0,065</b>	<b>0,100</b>	/	/	/
participschool	-0,007	<b>-0,042</b>	<b>0,044</b>	-0,006	<b>0,031</b>
supportschool	0,017	0,020	-0,031	0,014	-0,014
disciplineschool	<b>0,024</b>	0,003	0,012	-0,021	-0,006
Observations	3 816	3 474	2 648	2 132	12 070

Les élasticités en caractère gras sont associées à des coefficients de régression significatifs

Source : calculs de l'auteur à partir des données pondérées et non pondérées de l'ELNEJ.

## CONCLUSION

Dans un premier temps, le principal objectif de cette étude était de mesurer l'impact de l'environnement familial et scolaire sur le développement cognitif de l'enfant et ce, dans une perspective bilatérale. En effet, une seule étude empirique (Todd et Wolpin (2003)) a su faire l'analyse simultanée de l'apport des inputs scolaires et familiaux sur l'évolution cognitive des jeunes. Dans un deuxième temps, après avoir quantifié leurs effets, il fallait déterminer lequel de ces deux facteurs influençait le plus le score de l'enfant au test de mathématiques. Ainsi, les résultats de ces estimations pourraient orienter efficacement les décisions que les politiciens auront à prendre quant à la mise en oeuvre de diverses politiques sociales, notamment celles relatives à l'éducation ou la famille.

Plusieurs auteurs se sont penchés sur la question, mais les divergences dans leurs méthodes d'analyse font en sorte qu'il est difficile d'avoir des résultats qui convergent dans la même direction et de conclure uniformément. Todd et Wolpin sont les seuls qui ont tenté de combiner les deux types d'inputs et en sont arrivés à conclure que les inputs familiaux avaient une influence amplement supérieure aux inputs scolaires. Par contre, les données qu'ils détenaient sur les facteurs d'environnement relatifs aux établissements scolaires manquaient de précision. Ceci a peut-être eu pour effet de diminuer indirectement l'impact de cette catégorie d'inputs ou tout simplement de surestimer la contribution familiale. Par ailleurs, Krueger défend fortement l'importance des ressources académiques quant à la réussite scolaire des enfants. Dans son étude, il met l'accent sur le rôle que l'effectif des classes joue au sein du développement des capacités intellectuelles des jeunes du primaire et du secondaire. Il mentionne que plus la taille de classe est petite, plus élevés sont les résultats des élèves à des tests de mathématiques et de vocabulaire. Enfin, l'étude de Jencks et Phillips justifie le choix de la variable dépendante de mon modèle. En effet, ces derniers démontrent que les résultats qu'un enfant obtient à un test de mathématiques influencent grandement son accomplissement scolaire et le revenu futur qu'il gagnera lors de son entrée sur le marché du travail.

La base de données de l'ELNEJ a plusieurs avantages. Premièrement, l'abondance de questions posées aux enfants, aux parents, aux enseignants et aux directeurs permet d'obtenir une quantité importante d'information sur l'environnement familial et scolaire de l'enfant. Ceci fait en sorte que les différentes estimations que j'effectue sont précises et moins susceptibles d'être biaisées. Deuxièmement, la structure longitudinale de cette enquête permet de suivre un enfant à travers le temps, ce qui permet d'apporter un aspect dynamique à l'analyse. Cette enquête a également quelques faiblesses. Tout d'abord, mes sous-échantillons d'analyse sont au prise avec de l'attrition, de sorte que seulement une minorité d'enfants sont suivis lors des quatre premiers cycles de l'enquête. De plus certaines questions d'enquête, très intéressantes pour l'analyse de ce problème, ne s'adressent pas directement aux enfants âgés entre 6 et 11 ans. Enfin, les questions se rattachant aux inputs scolaires ont un taux de non-réponse non négligeable. Ceci a pour effet de diminuer la précision des résultats, car j'ai dû avoir recours à l'imputation à quelques reprises.

L'estimation des paramètres de la fonction de production de résultats cognitifs a été réalisée à l'aide de trois types de modèles de régressions économétriques. Le premier était la spécification contemporaine. Il s'agit d'un MCO qui capte l'effet des inputs familiaux et scolaires courants sur le score en mathématiques de la même année et ce, pour les cycles 1 à 4. Les résultats qui en ressortent sont, premièrement, que les inputs familiaux, mesurés par les pratiques parentales, semblent avoir peu d'effets, sauf pour ce qui est du degré de cohérence des parents et du niveau d'interactions entre ces derniers et leur enfant. Toutefois, les résultats démontrent qu'une interaction accrue entre les parents et le jeune aurait une influence négative sur son développement. Ceci étant plutôt contre-intuitif, on peut facilement supposer que cette spécification souffre d'un biais lié à un problème d'endogénéité. Deuxièmement, la taille des classes et la quantité de devoirs que l'enseignant donne à ses élèves semblent être les facteurs d'environnement scolaire les plus influents. Une augmentation de ces deux inputs scolaires fait bondir le score au test de mathématiques de façon plus importante que les inputs familiaux. Même si cette variable n'est présente que dans les deux premiers cycles de l'enquête, il ne faut pas négliger l'impact positif de la performance académique du professeur de l'enfant. Enfin, pour ce qui est des variables de

contrôle, plusieurs d'entre elles ont un impact considérable sur la performance de l'enfant, ce qui justifie leur inclusion dans le modèle. D'un côté, l'âge de l'enfant, le fait qu'il habite dans la province de Québec, le revenu familial et l'éducation de sa mère a un effet positif sur son développement. D'un autre côté, le fait qu'il nécessite une aide académique supplémentaire produit plutôt l'effet contraire. Le deuxième type de spécification est le modèle à valeur ajoutée. Il intègre un ou deux retards de la variable dépendante dans la spécification contemporaine afin de combler pour le manque d'inputs retardés. Or, comme on a pu le constater auparavant, c'est uniquement le modèle à un retard qui est adéquat. En effet, même si le coefficient de la variable  $T_{i,t-2}$  du modèle à deux retards est significatif, celui-ci s'avère inapproprié puisque l'ajout de quelques inputs retardés statistiquement significatifs vient désapprouver son utilisation. Autrement dit, cette spécification n'apporte aucune amélioration à la précédente. Le dernier modèle est la spécification cumulative. Celui-ci permet d'enrayer une grande partie de l'endogénéité grâce à l'élimination de l'habileté innée et constante que possède l'enfant. Ce type de régression longitudinale s'effectue à l'aide du modèle avec effets fixes ou aléatoires. Toutefois, le test de Hausman rejette fortement l'utilisation du second, de sorte que l'analyse dynamique a été faite avec l'aide des effets fixes. Les principaux résultats qui en ressortent sont que les inputs familiaux ne semblent pas avoir une influence très significative sur la performance des jeunes en mathématiques. Il n'y a que le niveau d'hostilité des parents qui a un effet négatif sur la variable dépendante du modèle. Parallèlement, seulement qu'un des inputs scolaires a impact significatif et positif, soit le niveau d'implication des enseignants dans leur établissement scolaire. Il faut cependant faire une petite mise en garde quant à l'interprétation de ces résultats, puisque les inputs familiaux utilisés dans les régressions précédentes diffèrent légèrement de ceux définis dans la littérature, ce qui a peut-être pour effet de sous-estimer leur apport dans le développement cognitif des enfants. De plus, la faible variation intertemporelle de certains inputs familiaux et scolaires au niveau individuel fait sorte que la plupart des coefficients de ces inputs ne sont pas statistiquement significatifs. Il serait donc intéressant de voir comment les résultats pourraient changer si j'utilisais des données de niveau familial plutôt que de niveau individuel. Je serais en mesure d'observer plus de variation et par conséquent, d'obtenir des coefficients nettement plus significatifs. Enfin, après avoir calculés l'élasticité de chacun des inputs de la fonction de production, j'en vient à conclure que ceux-ci sont plutôt inélastiques.

En somme, les conclusions tirées de cette étude ne sont pas à négliger. Le développement des enfants est un facteur essentiel à la réussite économique d'une nation. En effet, ces jeunes élèves représentent la main-d'œuvre des prochaines décennies. Si ces derniers n'arrivent pas à atteindre un certain niveau de compétences intellectuelles, ils seront peu productifs et par le fait même, nullement compétitifs sur le marché du travail international. Si les autorités gouvernementales n'interviennent pas au niveau de leur éducation, scolaire ou familial, de graves conséquences économiques et sociales peuvent survenir. Dès lors, les politiciens doivent être plus consciencieux lorsqu'ils allouent des sommes importantes dans leurs budgets aux secteurs académique et familial. Ceux-ci doivent s'assurer qu'ils attribuent des montants qui généreront des résultats efficaces et équitables, autant que possible. Par ailleurs, ils doivent tenir compte de toutes les répercussions économiques et sociales des actions qu'ils entreprennent. Par exemple, une mesure qui inciterait un des parents à rester à la maison afin de passer davantage de temps avec son enfant pourrait affecter le taux d'emploi et par conséquent, la production intérieure brute du pays. Finalement, il serait intéressant de voir quel serait l'impact de l'environnement familial, mais lorsque l'enfant est en bas âge, soit durant la période 0-5 ans. Autrement dit, on chercherait à savoir si l'éducation que les parents donnent à leurs enfants lorsqu'ils sont encore des bébés influence leur développement intellectuel lors de la transition à l'école primaire. Pour l'instant, il est difficile de prévoir si cette analyse est réalisable, car le nombre d'observations relatifs à un échantillon de la sorte est très petit, étant donné que la cohorte longitudinale de départ est au prise avec de l'attrition et qu'à partir du cycle 2, l'ELNEJ s'intéresse davantage à des enfants entre 0 et 5 ans. Quoi qu'il en soit, il serait profitable de connaître l'effet de cette nouvelle catégorie d'inputs dans la fonction de production de résultats cognitifs, car ceci nous permettrait d'intervenir très tôt dans le processus de développement de l'enfance.

## BIBLIOGRAPHIE

Belzil C. & Hansen J., "*Unobserved ability and the return to schooling*", Institute for the Study of Labour, Berkley, IZA DP No. 508, mai 2002.

Card D., "*The Causal Effect of Education on Earnings*", Center for Labour Economics, University of California, Berkley, Working Paper No. 2, mai 1998.

Coleman James, "*Social Capital in the Creation of Human Capital*", American Journal of Sociology, 1994.

Currie J. & Thomas D., "*Early Test Scores, Socioeconomic Status and Future Outcomes*", NBER, Working Paper No. 6943, February 1999.

Division de la statistique du revenu, "*Les seuils de faible revenu de 2005 et les mesures de faible revenu de 2004*", Document de recherche, Statistique Canada, avril 2006.

Giles P., "*Mesure de faible revenu au Canada*", Document de recherche, Statistique Canada, décembre 2004.

GREENE, H. William . "*Econometric Analysis*", New York University, Third Edition, 1075 pages, 1997.

Hanushek E.A., "*The Evidence on Class Size*", Occasional Paper Number 98-1, W. Allen Wallis, Institute of Political Economy, University of Rochester, Rochester, February 1998.

Jencks C. et Phillips M., "*Aptitude or achievement: why do test scores predict educational attainment and earnings?*", Earning and learning : how school matters, Brookings Institution Press, pp.15-47, 1999.

Krueger A., "*Economic considerations and class size*", The Economic Journal, vol. 113 (February), F34-63, 2003.

Lazear E.P., "*Educational Production*", Quarterly Journal of Economics, vol. 116(3), pp.777-803, 2001.

Leibowitz A., "*Home Investments in Children*", Journal of Political Economy, vol. 82(2), Pt. I, pp. S111-31, avril 1974.

Merrigan P., "*Analyse statistique des données en panels et évaluation des politiques publiques*", Recueil de notes de cours, CIQSS, été 2005.

Phipps S. & Lethbridge L., "*Le revenu et les résultats des enfants*", Document de recherche, Statistique Canada, mai 2006.

Todd P. et Wolpin K., "*On the specification and estimation of the production function for cognitive achievement*", The Economic Journal, vol. 113 (February), F3-33, 2003.

Todd P. et Wolpin K., "*The production of cognitive achievement in children: home, school and racial test score gaps*", Department of Economics, University of Pennsylvania, april 26, 2004.

Varian H.R., "*Analyse microéconomique*", De Boeck Université, 1995.

Weiss A., "*A Sorting-cum-Learning Model of Education*", The Journal of Political Economy, Vol. 91, No.3, juin 1983.

Wooldridge J. M., "*Economic analysis of cross section and panel data*", Cambridge, Mass.: MIT Press, 2002.

## ANNEXE A

### SCHÉMATISATION DE L'ELNEJ

Figure A-1 : Contenu de l'ELNEJ

<p><b>Parents/autres membres de la famille</b> <i>caractéristiques démographiques</i></p> <ul style="list-style-type: none"><li>- nom</li><li>- date de naissance*</li><li>- sexe*</li><li>- état matrimonial</li><li>- lien entre tous les membres du ménage</li><li>- pays de naissance/citoyenneté/immigration</li><li>- origine ethnique</li><li>- langue</li><li>- religion</li></ul> <p><i>scolarité</i></p> <ul style="list-style-type: none"><li>- plus haut niveau atteint</li><li>- diplôme</li><li>- cours suivi actuellement</li></ul> <p><i>marché du travail</i></p> <ul style="list-style-type: none"><li>- activité principal</li><li>- emploi</li><li>- nombre d'heures de travail; le salaire horaire</li><li>- travail par quart/travail les fins de semaine</li></ul> <p><i>revenu</i></p> <ul style="list-style-type: none"><li>- sources (ménage)</li><li>- montant (ménage, répondant)</li></ul> <p><i>santé physique</i></p> <ul style="list-style-type: none"><li>- limitation des activités</li><li>- problèmes de santé chroniques</li><li>- état de santé général</li><li>- usage du tabac</li><li>- consommation d'alcool</li><li>- antécédents de la mère</li></ul> <p><i>santé mentale (répondant)</i></p> <ul style="list-style-type: none"><li>- dépression</li></ul>	<p><b>Famille</b> <i>fonctionnement de la famille</i> <i>satisfaction conjugale</i> <i>soutien social (répondant)</i></p> <ul style="list-style-type: none"><li>- famille/amis</li><li>- autres</li></ul> <p><b>Logement</b></p> <ul style="list-style-type: none"><li>- propriétaire</li><li>- subventionné</li><li>- conditions</li><li>- nombre de chambres à coucher</li></ul> <p><b>Collectivité</b> <i>quartier</i></p> <ul style="list-style-type: none"><li>- satisfaction</li><li>- sécurité</li><li>- cohésion sociale</li><li>- problèmes</li><li>- bénévolat</li></ul> <p><i>observation du quartier (par l'interviewer)</i></p> <ul style="list-style-type: none"><li>- problèmes</li><li>- aménagement</li><li>- état des immeubles</li></ul> <p><b>École</b> <i>enseignant</i></p> <ul style="list-style-type: none"><li>- rendement de l'enfant</li><li>- comportement de l'enfant</li><li>- méthodes d'enseignement</li><li>- participation des parents</li><li>- caractéristiques démographiques de l'enseignant</li></ul> <p><i>directeur</i></p> <ul style="list-style-type: none"><li>- élèves à l'école</li><li>- problèmes de discipline</li><li>- participation des parents</li><li>- caractéristiques démographiques du directeur</li></ul> <p><i>Test d'habileté en mathématique</i></p>
--	--

\* recueilli au sujet de tous les membres de la famille, les autres variables de cette case sont recueillies pour les parents seulement

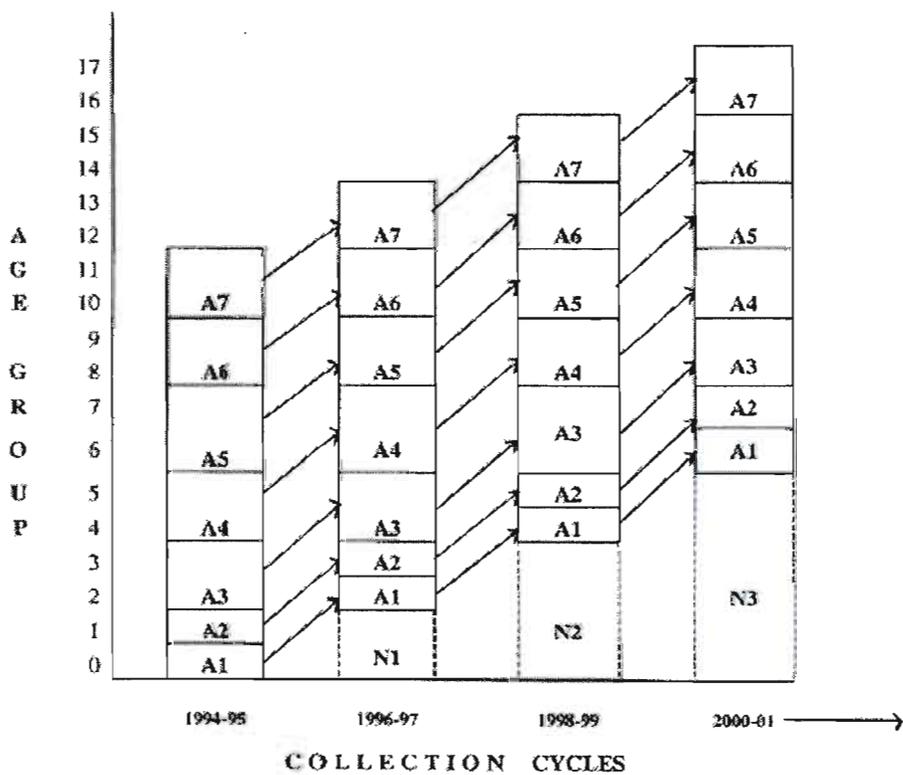
<p><b>Enfants</b>  <i>caractéristiques démographiques</i>  <i>état de santé</i></p> <ul style="list-style-type: none"> <li>- état de santé général</li> <li>- taille et poids</li> <li>- niveau d'activité**</li> <li>- état de santé**</li> <li>- blessures</li> <li>- problèmes de santé chroniques/limitation des activités</li> <li>- utilisation des services de santé</li> <li>- usage de médicaments</li> <li>- événements traumatisants**</li> </ul> <p><i>renseignements périnataux**</i></p> <ul style="list-style-type: none"> <li>- santé de la mère avant la naissance du bébé</li> <li>- détails sur l'accouchement</li> <li>- santé du bébé et de la mère après la naissance</li> <li>- allaitement maternel</li> </ul> <p><i>tempérament**</i></p> <p><i>scolarité**</i></p> <ul style="list-style-type: none"> <li>- niveau</li> <li>- genre d'école/langue</li> <li>- absentéisme</li> <li>- problèmes de comportement</li> <li>- attitudes et attentes quant au rendement</li> <li>- éducation spécialisée</li> <li>- déménagements</li> </ul> <p><i>alphabétisation**</i></p> <ul style="list-style-type: none"> <li>- lecture</li> <li>- encouragements à écrire</li> <li>- devoirs</li> </ul>
---

<p><b>Enfants - suite</b>  <i>activités**</i></p> <ul style="list-style-type: none"> <li>- préscolaires</li> <li>- activités après l'école</li> <li>- télévision et jeux vidéos</li> <li>- responsabilités</li> </ul> <p><i>comportement**</i></p> <ul style="list-style-type: none"> <li>- habitudes de sommeil et alimentation</li> <li>- sentiments/actions</li> <li>- comportement difficile</li> </ul> <p><i>développement moteur et social**</i></p> <p><i>relations**</i></p> <p><i>rôle parental**</i></p> <ul style="list-style-type: none"> <li>- interaction positive</li> <li>- techniques parentales</li> <li>- soins de base</li> </ul> <p><i>famille et garde légale des enfants</i></p> <ul style="list-style-type: none"> <li>- garde légale de l'enfant</li> <li>- unions des parents antérieures et (ou) subséquentes</li> <li>- séparations</li> <li>- membres de la famille vivant à l'extérieur de la maison</li> </ul> <p><i>garde des enfants</i></p> <ul style="list-style-type: none"> <li>- types</li> <li>- heures</li> <li>- mode de garde pendant l'été</li> </ul> <p><i>mois compris**</i></p>
---

<p><i>Enfants âgés de 10 et 11 ans</i></p> <p><i>amis et famille</i></p> <p><i>école</i></p> <p><i>devoirs</i></p> <p><i>sentiments et comportement</i></p> <p><i>puberté</i></p> <p><i>tabac, alcool et drogues</i></p> <p><i>activités</i></p> <p><i>estime de soi</i></p>
--

\*\* varie selon l'âge de l'enfant

Figure A-2: Organisation de l'ELNEJ



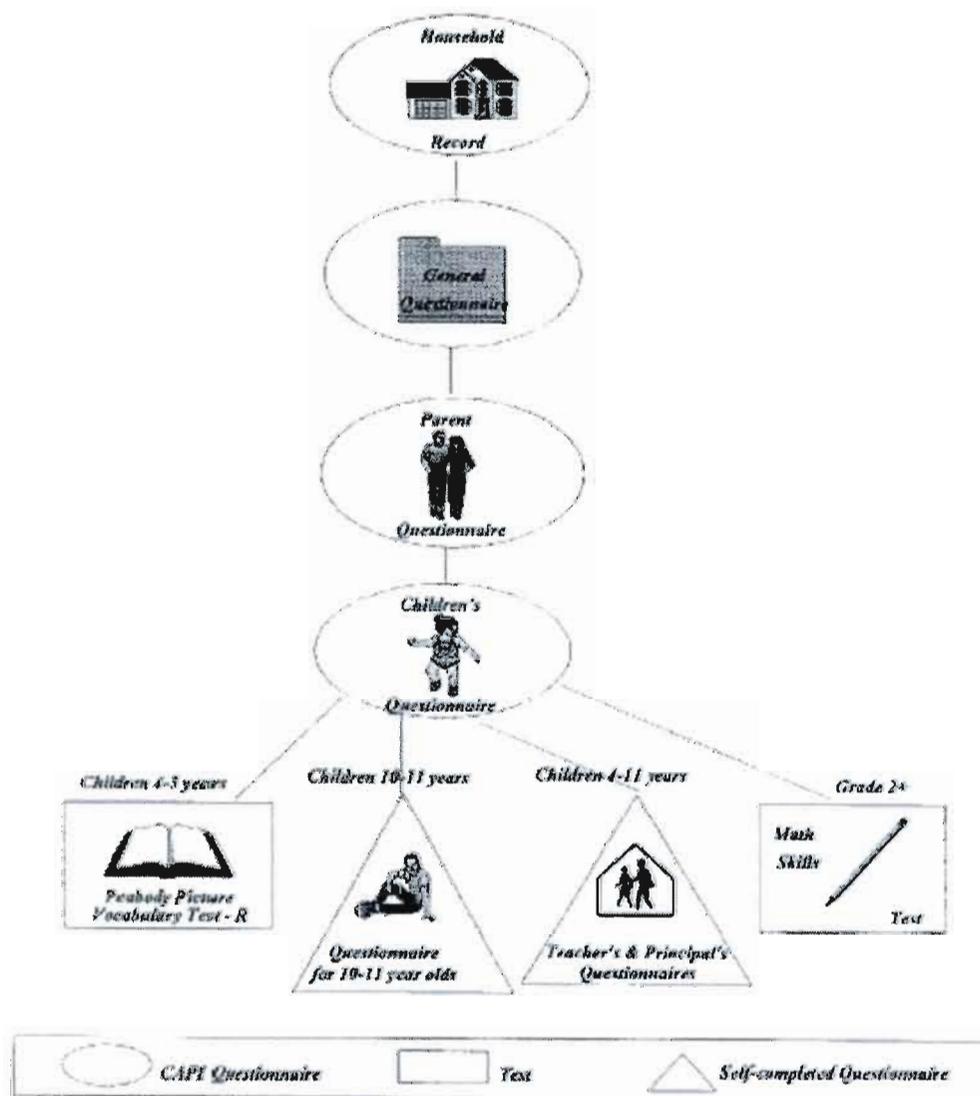
Source: NLSCY, Overview of Survey Instruments for 1994-1995 Data Collection Cycle 1, figure 1.

**Figure A-3:** Répartition de l'échantillon entre les différentes provinces et territoires du Canada



Source: NLSCY, Overview of Survey Instruments for 1994-1995 Data Collection Cycle 1, figure 2.

Figure A-4: Structure et composition de l'ELNEJ



Source: NLSCY, Overview of Survey Instruments for 1994-1995 Data Collection Cycle 1, figure 6.

## ANNEXE B

## AUTRES RÉSULTATS

**Tableau B-1 : Régressions transversales pour les cycles 1 à 4 (variables de contrôle)**

	Variable dépendante: mathscore			
	<i>cycle 1</i>	<i>cycle 2</i>	<i>cycle 3</i>	<i>cycle 4</i>
female	-0,173 (2.608)	-4,992 (2.830)*	1,419 (3.026)	-2,016 (3.345)
disable	-49,330 (5.221)***	-34,123 (5.418)***	-32,019 (6.236)***	-25,501 (3.913)***
childage	88,512 (13.865)***	97,432 (15.235)***	116,947 (17.040)***	95,149 (17.781)***
childage2	-2,389 (0.768)***	-2,985 (0.852)***	-4,219 (0.949)***	-3,128 (0.991)***
siblings	-0,281 (1.371)	0,549 (1.433)	0,409 (1.870)	-2,191 (1.999)
monop	-5,743 (5.727)	6,776 (6.514)	9,049 (5.785)	-9,298 (7.660)
bio	-3,332 (4.806)	7,383 (5.700)	16,715 (4.964)***	-0,699 (5.795)
maritimes	6,352 (3.403)*	1,502 (4.030)	-3,399 (3.820)	-12,345 (4.513)***
quebec	24,505 (4.662)***	24,969 (4.676)***	15,311 (5.157)***	26,007 (5.795)***
prairies	10,191 (3.326)***	8,396 (3.530)**	-0,097 (4.033)	-8,520 (4.552)*
bc	26,492 (4.558)***	12,905 (5.146)**	2,272 (5.838)	-1,581 (6.198)
bigurban	-1,716 (3.663)	0,852 (4.177)	1,241 (4.390)	0,342 (4.875)
midurban	-1,410 (3.030)	-3,003 (3.521)	-10,323 (3.847)***	-2,466 (4.486)
smallurban	4,285 (3.120)	-5,195 (3.556)	-7,403 (4.060)*	-8,388 (4.805)*
pmkagebirth	0,847 (0.372)**	0,344 (0.340)	0,523 (0.336)	-0,433 (0.428)
pmkeduc	2,404 (0.720)***	3,128 (0.763)***	3,555 (1.105)***	3,419 (1.291)***
realavgincome	5,988 (3.015)**	6,234 (3.065)**	6,715 (2.995)**	-0,278 (4.668)
immigrant	2,624 (4.478)	3,093 (4.936)	10,967 (5.850)*	-2,725 (6.870)
Constant	-383,995 (72.032)***	-362,175 (78.041)***	-448,731 (86.219)***	-269,068 (94.018)***
Observations	3816	3474	2648	2132
R-carré	0.62	0.62	0.62	0.62

Valeur absolue des écarts-types entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données pondérées de l'ELNEJ.

**Tableau B-2 : Régressions des modèles à valeur ajoutée à un retard et à deux retards**  
(variables de contrôle)

	Variable dépendante: mathscore contemporain	
	Modèle à un retard	Modèle à deux retards
female	-4,858 (5.084)	-12,054 (9.418)
disable	-26,402 (5.266)***	-35,698 (11.223)***
childage	-0,693 (98.63)	/
childage2	0,449 (4.904)	/
siblings	1,798 (3.056)	1,088 (6.349)
monop	-3,078 (9.784)	7,264 (18.762)
bio	-2,292 (6.548)	-16,252 (15.898)
maritimes	-8,057 (6.61)	-7,747 (12.491)
quebec	13,918 (7.196)*	17,314 (15.418)
prairies	-4,851 (6.205)	-13,979 (14.326)
bc	-3,673 (9.18)	-18,673 (21.441)
bigurban	-4,289 (8.03)	12,496 (15.416)
midurban	-4,826 (7.578)	-7,338 (13.229)
smallurban	-10,471 (7.936)	-5,088 (13.9)
pmkagebirth	0,492 (0.507)	0,796 (1.45)
pmkeduc	2,367 (1.66)	2,992 (2.388)
realavgincome	-0,083 (5.881)	-13,067 (11.011)
immigrant	-7,565 (9.091)	-23,239 (33.42)
Constant	190,899 (485.858)	249,310 (138.515)*
Observations	699	145
R-Carré	0.48	0.62

Valeur absolue des écarts-types robustes entre parenthèses  
\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données de l'ELNEJ.

Tableau B.3 : Test de Hausman (comparaison entre FE et RE)

Variables indépendantes	Variable dépendante : mathscore			
	Coefficients		Différence	Écart-types
	(b) FE	(B) RE	(b-B)	$\sqrt{\text{diag}(V_b - V_B)}$
cycle2	84,451	-3,35	87,799	8,352
cycle3	145,052	-24,35	169,399	16,286
cycle4	216,940	-34,77	251,711	24,674
childage	85,261	104,10	-18,842	6,337
childage2	-4,625	-3,39	-1,237	0,256
bio	-5,361	5,72	-11,077	6,242
monop	-3,833	3,68	-7,509	4,905
maritimes	51,176	0,57	50,605	18,774
quebec	32,781	23,11	9,666	42,271
prairies	31,671	2,04	29,633	19,330
bc	36,051	-1,01	28,461	24,793
bigurban	15,182	2,79	12,390	9,892
midurban	-4,056	-2,72	-1,339	5,703
smallurban	1,752	0,44	1,314	4,288
pmkeduc	1,191	2,93	-1,740	1,555
interactionscore	0,007	-0,68	0,686	0,332
scorefam	-0,237	-0,29	0,050	0,174
ineffectivescore	-0,537	-0,55	0,012	0,279
consistencyscore	0,421	0,41	0,008	0,307
aversivescore	0,657	0,36	0,296	0,493
classize	0,216	0,34	-0,120	0,166
participschool	0,622	0,36	0,265	0,217
supportschool	-0,381	-0,14	-0,243	0,232
highhw	-0,230	0,07	-0,301	0,267
expprof2	0,375	0,40	-0,230	1,801
expprof	0,915	-0,52	-4,565	2,248
expprofniv	2,333	4,94	3,343	1,225
expprofniv2	-0,331	0,62	-0,949	1,632
midhw	-1,761	1,94	-3,697	/
disciplinool	0,174	7,59	1,433	0,509

b = consistant sous  $H_0$  et  $H_a$ ; obtenu par FE

B = non consistant sous  $H_a$ , efficient sous  $H_0$ ; obtenu par RE

Test:  $H_0$ : la différence entre les coefficients n'est pas systématique

$$\chi^2(26) = (b-B)'[(V_b - V_B)^{-1}](b-B)$$

$$= 150.39$$

$$\text{Prob} > \chi^2 = 0.0000$$

Source : calculs de l'auteur à partir des données de l'ELNEJ.

Tableau B.4 : Résultats du modèle avec effets aléatoires

	Variable dépendante: <i>mathscore</i>
<i>inputs familiaux</i>	
scorefam	-0.679 (0.198)***
interactionscore	-0.287 (0.103)***
ineffectivescore	-0.549 (0.159)***
consistencyscore	0.413 (0.164)**
aversivescore	0.361 (0.301)
<i>inputs scolaires</i>	
classize	0.336 (0.115)***
participschool	0.357 (0.166)**
supportschool	-0.139 (0.177)
disciplineschool	0.071 (0.201)
midhw	1.936 (1.230)
highhw	4.941 (1.511)***
expprof	-0.517 (1.806)
expprof2	0.403 (0.419)
expprofniv	-1.011 (1.577)
expprofniv2	0.618 (0.454)
Constant	-398.992 (24.515)***
Observations	12070
Nombre de childid	9077

Valeur absolue des écarts-types entre parenthèses

\* significatif à 10%; \*\* significatif à 5%; \*\*\* significatif à 1%

Source : calculs de l'auteur à partir des données de l'ELNEJ.