

Université de Bourgogne
Faculté de Lettres et Sciences Humaines
École doctorale LISIT
INSTITUT DE RECHERCHE SUR L'ÉDUCATION
IREDU - CNRS

Texte présenté en vue de l'obtention du titre de
Docteur en Sciences de l'Éducation

Dijon, le 27 juin 2011

Les effets de pairs à la lumière des interactions entre élèves et des dimensions subjectives du vécu scolaire

TOME II - ANNEXES

Rodrigo ROCO FOSSA

Sous la direction de Madame Marie DURU-BELLAT

MEMBRES DU JURY : Mesdames et Messieurs

Pascal BRESSOUX	Professeur à l'Université Pierre Mendès-France - Grenoble II
François DUBET	Professeur à l'Université Victor Segalen - Bordeaux II
Vincent DUPRIEZ	Professeur à l'Université Catholique de Louvain
Marie DURU-BELLAT	Professeur à l'Université de Bourgogne et à l'Institut d'Études politiques de Paris
Bruno SUCHAUT	Professeur à l'Université de Bourgogne

ANNEXES

Liste d'Annexes

- A. Données et Echantillons (p. 7)
- B. Construction des variables principales et complémentaires (p. 13)
- C. Questions et choix méthodologiques (p. 129)
- D. Modélisation des acquis scolaires (p. 175)
- E. Modélisation des principales dimensions subjectives (p. 207)
- F. Modélisation des pratiques d'entraide (p. 223)
- G. Questionnaires utilisés (p. 237)
- H. Bibliographie des annexes (p. 245)

ANNEXE « A »

**DONNÉES et
ÉCHANTILLONS**

Le choix des données

L'un des principaux intérêts et avantages des enquêtes SIMCE est leur caractère exhaustif (i.e. elles recouvrent la quasi totalité des élèves d'un grade donné). A l'origine, celles-ci correspondaient surtout à une comparaison relative entre établissements, à la manière d'un palmarès. Depuis l'année 1996, ces enquêtes communiquent des résultats et des variables désagrégées jusqu'au niveau individu. Toutefois, une faiblesse de ce système d'évaluation est qu'il n'est pas construit en forme de panel ou de pseudo-panel.

Nos principales données d'analyse (bases SIMCE 2004 du Ministère d'Education au Chili), sont synthétisées dans les deux tableaux suivants (A1 et A2). Sauf indication contraire explicitant d'autres sources, toutes les estimations construites proviennent de ces mêmes bases des données.

Tableau A1
Nombre total d'élèves ayant participé à l'évaluation SIMCE 2004,
par discipline et par réponse au questionnaire élèves.

TESTS SIMCE 2004	Nb d'élèves ayant passé les tests (1)	Nb d'élèves ayant passé les tests et qui ont répondu à la fois au questionnaire Élèves (2)	Delta (2) - (1)	Perte (non identifiés) en %
TOTAL	282276(*)	278476	-3800	-1,35
Mathématiques	276365	272670	-3695	-1,34
Langage	275547	274817	-730	-0,26
Sc. de la Nature	276300	275652	-648	-0,23
Sc. Sociales	275731	272037	-3694	-1,34

(*) Lecture : 28 2276 élèves ont passé au moins un des quatre tests, dont 278 476 ont répondu au questionnaire pour élèves. Il existe donc, une perte de 1,35% d'élèves qui ne pourront pas être considérés dès le départ parce que leurs caractéristiques ne peuvent nullement être identifiées. Notons que selon les données officielles du MINEDUC ces élèves correspondent à 97,1% du total d'élèves du pays pour le grade concerné.

Dans le tableau A1, nous pouvons apprécier le nombre total d'élèves qui ont participé à l'évaluation SIMCE 2004. De même, nous y pouvons voir la différence entre ce nombre et le nombre d'élèves pour lesquels il a été possible d'apparier les informations contenues dans le questionnaire aux parents et le questionnaire aux élèves (cf. tableau A2 ; Annexes G). Pour nos analyses, ces sources constituent le socle fondamental d'information pour saisir les

caractéristiques individuelles, puis agrégées, liées à la scolarisation des élèves chiliens. En effet, le haut niveau de réponse aux questionnaires apparentés aux tests et les multiples informations qu'ils contiennent nous ont permis de remplir soigneusement et avec un bon degré de certitude les données manquantes, dont la proportion se situe, en moyenne, à près de 10%.

Tableau A2

Nombre total d'élèves, parents et enseignants selon questionnaire répondu.

Type de Questionnaire SIMCE 2004	Nb d'identifiés ou répondants	Nb d'ITEMS à Répondre
Élèves	278476	116
Parents	262536 (*)	56
Profs. Maths	9356	129
Profs. Langage	9363	113
Profs. Sc. de la Nature	9350	113
Profs. Sc. Sociales	9356	99

(Référence : 9412 classes sur 5612 établissements)

(*) Lecture : 262536 familles/parents ont répondu au questionnaire respectif. Cela correspond à 94,3% de répondants au questionnaire élèves qui constitue la base de notre travail. En outre, sur 9412 classes, le taux de réponse des enseignants a été en moyenne de 99%. Ceci sachant que certaines classes sont assurées par les mêmes enseignants sans que cela soit, hélas, identifiable. Notons que les 5 612 établissements correspondent à 99,8% d'établissement offrant ce niveau en 2004 dans le pays.

Pour ce faire, nous avons en général apparié des informations d'une réponse à l'autre selon le cas (e.g. le niveau de scolarité des parents avec le type d'emploi ou de sécurité sociale ; le genre de l'élève et le caractère non-mixte de l'école ; etc.). Notons toutefois, une perte additionnelle de 0,91% des individus (n=2540) due à l'impossibilité d'identification de la variable genre de l'élève. Toutefois, ayant déclaré d'autres pratiques et caractéristiques, ces élèves seront utilisés à certains moments. De même, ceux qui n'ont pas passé l'un ou plusieurs des tests, mais qui sont dans le questionnaire d'élèves participeront, par exemple, à la construction de mesures agrégées. C'est notamment le cas des variables telles que la taille des classes/école pour lesquelles nous avons mis à contribution toutes les données disponibles.

Enfin, notons que, d'une manière complémentaire et non-exhaustive, nous avons aussi mobilisé d'autres bases des données. Certaines correspondent à des évaluations SIMCE additionnelles (2000 et 2006), tandis que les autres contiennent des données administratives demandées spécialement auprès du Ministère de l'Éducation et du Ministère de Planification,

ainsi que du Programme des Nations Unies pour le Développement Humain (PNUD). De cette manière, nous avons construit des variables liées aux établissements et aux communes pour tester ensuite la pertinence de nos remplissages des données manquantes (e.g. scolarité des parents), ou encore, pour décider de la pertinence de certaines variables. Plus précisément, nous avons considéré les résultats de l'enquête nationale des ménages effectuée en 2003, puis en 2006. Appliquée depuis la fin des années 80, cette enquête, dénommée CASEN (*CAracterizaciónSocioEconómicaNacional*), est le principal instrument de pilotage de l'évolution socioéconomique du pays. A partir des bases CASEN et PNUD, nous avons mis à contribution quelques variables spécifiques désagrégées par commune, notamment les niveaux de scolarisation et de revenu des familles, ainsi que la taille des groupes familiaux et la prévalence de la préscolarisation.

Les échantillons de travail

D'après la littérature révisée et d'une vision succincte portée tant sur la taille des cours (i.e. classes) que sur le nombre des classes offertes par les écoles et l'emplacement géographique de celles-ci (cf.annexes plus loin), nous avons décidé de prendre quelques décisions sur nos échantillons de travail. En effet, si notre objectif est de tester l'existence et l'ampleur des effets de pairs à l'égard des interactions entre élèves, tout en prenant compte des phénomènes de composition de la classe, il semble pertinent d'établir un seuil par rapport à leur taille. Ceci avec l'objectif d'éliminer des biais potentiels liés aux groupes trop restreints. De plus, nous savons que certains phénomènes comme la concurrence entre établissements ou le choix des familles sont fortement urbains, et qu'ils déterminent la conformation de groupes de pairs (cf. chapitres n°1, n°2 et n°6).

Le tableau A3 ci-dessus nous montre trois échantillons différents possibles selon que l'on discrimine la zone (urbaine) et la taille des cours. Ainsi, notre échantillon de travail prioritaire correspond à 85,5% du total d'élèves. Ceci après élimination de ceux situés en zone rurale et de ceux placés dans des cours de taille inférieure à 20. Le nombre des classes utilisables sur le total diminue de 25% tandis que celui des établissements le fait de 36%. Nous allons donc travailler davantage avec un échantillon général (n=236 133), puis des échantillons par discipline. Toutefois, nous allons souvent communiquer des estimations exploratoires ou de construction de variables faites sur l'ensemble de la population disponible, tout en essayant de faire noter les différences entre échantillons, si celles-ci existent.

Tableau A3

	TOUS			CLASSES ≥ 10				CLASSES ≥ 20			
	TOUS (U+R)	URBAIN	%	Sous-Total U+R	% du total	Sous-Total Urbain	% du total	Sous-Total U+R	% du total	Sous-Total Urbain	% du total
<i>MATHS</i>	272554	241245	89	269417	99	240592	88	253388	93	233262	86
<i>LANGAGE</i>	274695	243144	89	271548	99	242489 (*)	88	255384	93	235094	86
<i>Scs. NAT</i>	275446	243847	89	272280	99	243183	88	256062	93	235744	86
<i>Scs. SOC</i>	271911	240710	89	268795	99	240062	88	252833	93	232771	86
<i>Élèves (**)</i>	270130	239038	88	267035	99	238432	88	251164	93	231200	86
<i>classes</i>	9412	7662	81	8910	95	7554	80	7791	83	7061	75
<i>Étabs</i>	5612	4033	72	5112	91	3926	70	4146	74	3565	64

Source : élaboration de l'auteur. Lecture (*) : en langage, il existe un total de 242 489 individus évalués, situés dans des classes en zone urbaine et ayant 10 élèves ou plus. Ceux-ci représentent 88% du total d'élèves évalués pour cette discipline dans cette année et niveau. Ils sont groupés dans 7 554 classes et 3 926 établissements urbains.

NOTE (**): les totaux d'élèves correspondent à chaque fois, à ceux qui ont passé tous les quatre tests. Toutes les proportions sont estimées par rapport à la colonne n°1 (« Tous »). Les données manquantes pour la variable genre ont été éliminées.

ANNEXE « B »

**CONSTRUCTION DES VARIABLES
PRINCIPALES ET COMPLÉMENTAIRES**

Description des variables utilisées dans les différentes analyses réalisées (cf. Encadré n°5)

Cette partie des annexes cherche à décrire les variables qui seront mentionnées à plusieurs reprises, analysées ou simplement incluses dans les modèles construits. Certaines d'entre elles ont été explicitées dans leur construction et caractéristiques principales dans la thèse elle-même ; d'autres non. Ici, nous allons, d'abord, définir d'une manière générale les variables utilisées. Puis, nous allons regarder par groupes de variables le comportement de ces facteurs à la lumière de facteurs importants ainsi que de notre principale variable contextuelle (i.e. le type d'école). Dans certains cas, cela permettra d'expliciter les options prises sur la manière de construire les variables, notamment conforme aux choix méthodologiques (voir plus loin). Pour certains cas aussi, nous allons illustrer des variables non prises directement dans nos analyses mais qui ont servi à élargir et à approfondir la vision d'ensemble sur nos données et leurs possibilités.

Variables de niveau élève

Aux principales variables cibles (i.e. déclarations sur les manières d'étudier et d'interagir, bien-être et confiance en soi mais aussi : appui des parents, rapports aux enseignants, attitudes envers les disciplines, fréquence d'étude déclarée, attentes de scolarité future ; cf. Chapitres N°7, N°8 et N°9), nous ajoutons d'autres variables sociodémographiques dont leur importance a été soulignée constamment dans la littérature. La liste de variables à intégrer dans les analyses au niveau individuel est donc constituée comme suit :

1) Caractéristiques socio-scolaires de l'élève

- **Genre** de l'élève, variable binaire prenant les valeurs de 1 si c'est une fille et 0 pour les garçons.
- **Redoublement**, variable binaire dont 1 indique que l'élève a redoublé au moins une fois auparavant 2004 et 0 si non.
- Avoir passé ou non par le **préscolaire**, variable binaire prenant les valeurs de 1 si les parents de l'élève déclarent qu'il est allé en « pré-kinder » et 0 sinon.
- **Type de famille**, variable à l'origine construite en quatre catégories (i.e. mère et père vivant avec l'enfant, mère ou père seuls avec l'enfant, autres cas). Ici, la catégorie « autre cas » sera additionnée aux familles dont le père ou la mère sont seuls avec leur(s) enfant(s), pour les prendre ensemble comme référence (valeur 1) face aux familles en couple (valeur 0) qui recouvrent 70% des effectifs. L'idée c'est de contrôler les probables effets

d'être enlevé par les deux parents ce qui signifie, entre autres, potentiellement plus de soutien et de ressources pour l'élève.

- L'élève **a au moins un frère** (1 si oui et 0 si non). Bien que nous ne connaissons ni la taille de la fratrie ni le rang qui y occupe l'élève, cette variable prétend contrôler en quelque sorte l'effet probable de remplacement et/ou d'addition aux impacts sur les acquis scolaires de la part des interactions et manière d'étudier.
- **Nombre d'années de scolarité de la mère**. Cette variable, censée refléter le niveau socioéconomique du foyer de l'élève est prise ici par consistance avec la même variable au niveau agrégé. Elle sera notre principal indicateur du niveau socioéconomique dont jouit chaque élève.
- **Capital culturel** associé au foyer de l'élève, pris en trois catégories dont les niveaux bas (<11 livres à la maison) et moyen (entre 11 et 50 livres à la maison) seront comparés avec le niveau haut (plus de 50 livres à la maison).

2) Variables rapportées aux pratiques et aux dimensions subjectives du vécu scolaire (cf. chapitres n°8 et n°9), dont :

a) Variables cible :

- Quatre principales **pratiques d'étude et d'interaction déclarées** par discipline et prises de manière binaire conforme aux déclarations redressées des élèves (cf. chapitre n°8). Pour d'autres analyses on travaillera avec la catégorisation générale à quatre modalités, prenant le profil « Solitaire » comme référence.
- Conditions de **niveau bas et moyen de bien-être à l'école** comparées au niveau haut. Eventuellement sera utilisée l'échelle factorielle construite.
- Conditions de **niveau bas et moyen de confiance en soi scolaire** comparées au niveau haut. Eventuellement sera utilisée l'échelle factorielle choisie.

b) Variables générales et/ou de contrôle :

- **Catégories de parents** d'après le classement proposé à quatre modalités. La modalité « parents préoccupés » (i.e. celle qui cumule de plus hauts niveaux pour les différentes propositions d'origine) servant de référence pour les autres (« aidants », « exigeants » et « souples »). Alternativement sera utilisée l'échelle factorielle choisie dans le cas de l'analyse des dimensions subjectives et de pratiques d'étude.
- **Rapport aux disciplines** pris à trois catégories ordinales, par discipline et en général (i.e. comme rapports aux études d'un point de vue du plaisir que l'on en tire).¹

¹ Ceci sous la déclaration « *Je vais à l'école parce que j'aime bien étudier* » prise à trois niveaux.

- **Rapports aux enseignants** pris, pour les présentes analyses, comme des échelles de type général (i.e. qui ne sont pas rapportées à la moyenne de chaque classe). Cette variable ne sera appliquée que pour les disciplines de langage et mathématiques. Eu égard de la forte corrélation entre ces mesures et afin de simplifier certaines analyses, nous utiliserons une moyenne de ces deux mesures, prise comme « rapport général aux enseignants ».
- **Fréquence d'étude déclarée** en dehors des cours, et prise ici en forme binaire regroupant dans la valeur 0 les options « *je n'étudie jamais* » ou « *... juste en face de contrôles* », puis sur la valeur 1 celles de « *j'étudie quelques jours à la semaine* » et « *j'étudie presque tous les jours* ». Pour l'ensemble (i.e., toutes disciplines confondues), cette mesure sera prise à trois niveaux.
- **Attentes déclarées par l'élève de scolarisation postérieure** en trois niveaux : *université et supérieur technique* comparés aux *autres* (i.e. \leq secondaire scientifique et humaniste). Eventuellement et pour certains modèles, nous allons supprimer l'option *supérieur technique*, laissant la variable en termes binaires (université ou pas), et dans d'autres cas, nous allons ajouter aux déclarations des élèves celles des parents. Ces dernières, en fait, sont soupçonnées de contenir des informations relatives à la manière dont les familles s'occupent de la scolarité de leurs enfants. Si ces deux mesures (attentes des parents et des élèves) se rassemblent beaucoup, nous ne pouvons pas nier, a priori, qu'elles peuvent véhiculer des contenus différents.

c) D'autres mesures prises au niveau individu correspondent à :

- Essai d'approche au **niveau de compétences des élèves**. Cette variable reste à notre avis fort discutable car rien n'autorise à penser que les scores des élèves sur chaque test sont comparables d'une discipline à l'autre et donc, susceptibles d'être mélangés (ce que l'on fait d'ailleurs souvent dans le domaine scolaire). Cela d'autant plus que ces scores reflètent surtout un positionnement de chaque élève à l'échelle nationale. Pourtant, nous n'avons aucun choix alternatif pour combler le besoins d'avoir une mesure proxy spécifique de ce facteur (i.e. une approximation du coefficient intellectuel des élèves), notamment quand on est en face des phénomènes tels que le BFLPE commenté au chapitre N°3. De ce fait, nous prenons le risque de prendre cette approche en additionnant, pour chaque élève, les scores aux quatre tests ici considérés, faisant l'hypothèse que cette mesure s'approche en quelque sorte de cette dimension inconnue pour nous (i.e. le vrai degré de compétences de l'élève). Ainsi, après les avoir additionnés nous avons rapporté ce nouveau score à la moyenne générale en le centrant sur elle avec un écart-type égal à un.² De ce fait, cette

² Consignons qu'une analyse factorielle intégrant les quatre tests donnait exactement quatre dimensions, chacune totalement reliée à une discipline. De même, une AF ne demandant qu'une seule dimension à partir des quatre scores donnait une échelle corrélée à $r=0,99$ avec celle que nous avons finalement choisi de par sa simplicité (i.e.

mesure est censée classer chaque élève sur l'ensemble faisant l'hypothèse que les disciplines sont équivalentes entre elles et qu'elles reflètent chacune une dimension spécifique, ce que nous ne pouvons pas du tout prouver. Au total, nous parions sur la très probable corrélation entre cette mesure, aussi imparfaite qu'elle soit, et le vrai niveau de compétences des élèves. Comme cela a été déjà illustré dans le chapitre précédent, cette variable ne sera utilisée que pour tester son lien avec certaines des dimensions subjectives et soulignant avec prudence ses imperfections d'origine.

- **Raisons déclarées pour aller à l'école**, dont les quatre définies comme plus importantes (i.e. « *aimer étudier* », « *être obligé* », « *tous le font* », etc.). Parmi ces variables, on en utilisera dans quelques analyses multivariées seulement deux (« *être obligé* » et « *avoir un bon travail après* ») les prenant comme des catégories muettes.
- Construction d'un **indice de contrôle du choix d'établissement**. Cet indice correspond à ce qui, dans la littérature économétrique, est connu sous le nom de *double équation d'Heckman* (Heckman, 1979). A différence des autres variables qui seront décrites dans leur construction plus en bas, les démarches et les analyses suivies pour arriver à la construction de cet indice ont été déjà explicitées (cf. annexes méthodologiques).

Variables contextuelles de niveau établissement

A la classification proposée en sept catégories d'école (cf. voir plus loin) nous avons ajouté deux variables. L'une est référée à l'**équipement de l'établissement** (laboratoire de sciences), et a été choisie de par son caractère discriminatoire entre établissements. Si cette variable est censée capturer un certain degré de ressources matérielles de la part des écoles, l'autre variable à intégrer dans les analyses correspond au **taux d'encadrement** (nombre d'élèves par enseignant). Ce taux d'encadrement nous indique aussi, indirectement, la taille des écoles et la quantité des ressources humaines disponibles.

Variables contextuelles associées à la classe

Ces variables, nous les avons divisés en deux grands groupes, d'abord, en nous centrant sur l'enseignant, puis en considérant certaines caractéristiques reliées à la composition de la classe.

1) Du côté des professeurs par classe :

a) Caractéristiques de l'enseignant, considérant :

- Le **nombre d'années d'expérience** déclarées ;

elle reflète indépendamment chaque discipline sans pour autant manquer de faire l'hypothèse, toutefois risquée, que les scores sont équivalents entre eux). Par ailleurs, rappelons que d'autres mesures telles que les pourcentages de réponses correctes par élève ou encore, les réponses déclinées par items n'étaient pas disponibles de par le caractère confidentiel des tests SIMCE.

- Le fait d'**avoir ou pas une formation spécialisée**, soit en éducation, soit dans la discipline enseignée. Ceci à trois niveaux : pas de spécialisation, post-titre (i.e. formation à tonalité professionnelle) et post-grade (i.e. formation à tonalité académique) ;
- Le fait que l'enseignant ait un **autre travail** (hélas, sans pouvoir préciser s'il l'a en tant qu'enseignant ou bien dans d'autres activités) ;

b) Croyances de l'enseignant, considérant :

- Le **niveau de préparation qu'il croit avoir pour enseigner la discipline en cours de 8^{ème} grade**, à trois niveaux (haut et moyen comparés au niveau bas) ;
- Le **niveau dans le système scolaire qu'il pense que leurs élèves atteindront en moyenne**, pour chaque classe et à trois modalités (supérieur technique et université comparés avec tous les autres réponses possibles) ;
- Le **niveau de préoccupation qu'il montre envers la discipline** en cours, appréhendé de manière qualitative, selon qu'il déclare s'adresser ou non aux parents d'élèves sur le sujet. Cette variable est censée traduire aussi l'avis de l'enseignant sur le climat régnant dans la classe ;
- Le **niveau de préoccupation qu'il montre envers les résultats académiques** des élèves du cours, appréhendé de manière qualitative selon qu'il déclare s'adresser ou non au Conseil de Professeurs sur le sujet ;

c) Pratiques de l'enseignant, considérant :

- Le degré de **couverture du programme officiel de contenus** qu'il déclare avoir atteint pour une classe donnée ;³
- Le **pourcentage des heures hebdomadaires déclaré comme étant consacré à la préparation des cours** sur l'ensemble des heures de travail ;
- Le degré (inféré) de **travail en collaboration avec leurs collègues** appréhendé en termes quantitatifs selon le nombre de réunions formelles au semestre/trimestre qu'ils déclarent avoir avec leurs collègues de la même discipline dans l'établissement. Cette variable considère trois modalités : avoir deux ou plus de deux réunions au semestre/trimestre ; avoir une réunion dans la même période ; ne pas avoir de réunions ou tout au plus, une par an. Elle est censée aussi capturer un certain degré d'organisation globale dans le travail de l'école (favorable aux apprentissages) et rendre compte de ce qui a été fait auparavant dans la littérature (cf. Opdenakker et Van Dame, 2000 ; Dupriez, 2003)

³ Consignons qu'à partir des analyses en parallèle au travail ici présenté nous avons trouvé que cette variable est fonction des caractéristiques de la classe, notamment de sa composition sous différentes perspectives. Elle est aussi fortement déterminée par le niveau de préparation que les enseignants déclarent avoir.

- Leur **profil général de pratiques de travail en cours**. Ces profils ont été construits pour chaque discipline (voir annexes) regroupant les professeurs en quatre ou trois sous-groupes selon les disciplines et dont l'une des distinctions possibles tient au fait de se servir des pratiques plus individualistes (i.e. faisant davantage appel au travail individuel de l'élève en cours) ou bien, plus groupales, tout en admettant un possible manque des pratiques définies dans chaque cas ;

2) Du côté de la composition de la classe, considérant :

- La **taille de la classe** (≥ 20) ;
- La **moyenne d'années de scolarité des mères** en tant que variable censée rendre compte du niveau socioéconomique moyen des classes et donc, d'une probable influence entre camarades, soit-elle, d'un point de vue normatif, soit-elle, encadrant ou autorisant des échanges entre élèves différents de ce point de vue. Le choix de cette variable n'est pas neutre à notre avis, surtout du moment où nous croyons disposer de plusieurs indices pouvant rendre compte du même phénomène.⁴ Toutefois, le choix de cette variable peut rendre plus comparables nos résultats avec ceux d'autres recherches menées au Chili utilisant aussi cet indicateur. Notons de plus que cette variable est aussi, de par leurs qualités discutées plus haut, plus simple et plus proche à la fois du réel, en même temps qu'elle est plus lisible car elle est exprimée à une échelle très concrète. Un point annexe dans ce choix est l'inclusion ou pas de l'individu lui-même dans le calcul des valeurs moyennes. Les portées théoriques et empiriques de ce choix sont discutées en annexe, sachant que l'usage d'une ou d'une autre modalité n'altère guère les résultats trouvés ;
- L'**écart-type de la scolarité des mères par classe**, variable qui, de même que la précédente, est censée mieux objectiver les phénomènes de composition au sein de la classe, notamment le cadre qui entoure les probables interactions entre élèves. En effet, rappelons que comme le notent Schneeweis et Winter-Ebmer (2007), l'hétérogénéité socio-scolaire peut être aussi un facteur qui influence les possibilités d'apprendre et les apprentissages eux-mêmes des élèves. De manière semblable que pour la variable antérieure, celle-ci sera souvent exprimée directement en années mais aussi rapportée à sa moyenne générale selon le cas ;

⁴ En effet, à partir de la construction d'un Indice composite du Niveau Socioéconomique des élèves (voir Annexes) nous avons construits des typologies par décile de NSE, puis par quintile, disposant de ce fait de plusieurs choix pour les exprimer au niveau agrégé, que ce soit la classe ou l'établissement (par exemple, prendre la moyenne de l'Indice de NSE ou le pourcentage appartenant à tel ou tel quintile, etc.). Consignons que nous avons conduit nos analyses en nous servant aussi de ces options, tout en trouvant des résultats très semblables, voire des fois presque identiques. Bref, nous communiquerons ici seulement les résultats utilisant la moyenne de scolarité des mères par classe de fait de la lisibilité de cette variable.

- La **proportion d'élèves classés comme étant en condition de faible bien-être à l'école**.⁵Sachant que cette proportion varie selon les types d'établissement mais qu'elle est finalement assez présente partout, nous la postulons comme une manière de capturer un climat dans la classe plus ou moins paisible (i.e. avec plus ou moins de facilité à l'enseignant pour faire la classe, pour trouver des élèves engagés, etc.). Comme on l'a vu, ceci est l'un des sujets qui occupe les études sur les effets de pairs. Notons que cette variable est issue directement de celle que nous avons construit pour le niveau individuel. Nous pensons qu'elle constitue une proposition originale qui n'a pas été considérée auparavant dans des études à base économétrique faites sur des données chiliennes. Pourtant, elle est soupçonnée de véhiculer le niveau moyen de compétences des élèves de fait que l'on peut penser que les élèves à bas bien-être réussissent, en moyenne, moins bien à l'école que les autres. Toutefois, ayant intégré aux analyses un proxy du niveau académique du groupe (i.e. la moyenne d'années de scolarité des mères), ce biais serait en quelque sorte tenu sous contrôle ;
- La **proportion d'élèves classés comme étant en condition de haute confiance en soi scolaire**. Cette variable a été intégrée par les mêmes raisons que la précédente mais avec moins d'espoir de trouver grâce à elle des significations claires. En effet, nous l'avons intégré à des fins exploratoires et par souci de symétrie. Pourtant, nous avons vu que les élèves du secteur public semblent avoir plus de confiance en soi scolaire que leur collègues du privé payant et du privé subventionné sélectif et avec FP. Cela pourrait montrer une tendance à l'autoprotection du soi des élèves qui savent qu'ils sont placés dans des écoles à basse, voire très basse, réputation académique. Il ne serait pas inattendu de trouver un impact négatif pour sa part.
- **Niveau moyen de compétences estimé dans la classe**. Cette mesure correspond à la moyenne des valeurs individuelles de la somme des scores aux quatre tests utilisés ici et elle est expliquée plus bas. Cette variable n'aura qu'un usage restreint compte tenu de ses qualités questionnables. Elle sera centrée sur sa moyenne générale. Pour certaines analyses, il sera utilisé aussi une mesure de la dispersion de ces compétences, donnée par l'écart-type (au niveau classe) de cette même variable ;
- **Ecart-type des scores par discipline** au sein de la classe, censé capturer l'hétérogénéité académique vue comme une dimension susceptible d'être différente de son homologue socioéconomique.

⁵ Si cette mesure est parfois centrée sur sa moyenne générale, notamment à des fins de l'analyse multiniveau, dans certains cas et par souci de lisibilité et de comparabilité, nous la prendrons en tant que telle (i.e. variation en pourcentage). Ceci d'autant plus que la transformation en mesures standardisées peut se faire à tout moment pour faciliter l'interprétation des coefficients. Ce commentaire est valable pour le % d'élèves en Haute CSS.

PRATIQUES des ÉLÈVES

(cf. Encadré n°3, TOME I)

Tableau B4

Comparaison entre les déclarations des élèves, redressées et non-redressées. Par pratique d'étude selon discipline (2004 - 8^{ème} Grade) (n=278 476)

Discipline	Pratiques	Proportions	
		Redressées	NON-Redressées
MATHS	Aider les autres	29%	31%
	Être Aidé par les autres	41%	43%
	Étudier Seul	34%	38%
	Demander des Cahiers	14%	18%
	Prêter des Cahiers	10%	13%
	Étudier avec mon Livre	12%	14%
	Rien/ Autre	5%	5%
LANGAGE	Aider les autres	28%	30%
	Être Aidé par les autres	33%	36%
	Étudier Seul	39%	43%
	Demander des Cahiers	15%	19%
	Prêter des Cahiers	11%	16%
	Étudier avec mon Livre	16%	19%
	Rien/ Autre	5%	5%
Scs. de la NATURE	Aider les autres	22%	24%
	Être Aidé par les autres	28%	31%
	Étudier Seul	42%	46%
	Demander des Cahiers	14%	17%
	Prêter des Cahiers	10%	14%
	Étudier avec mon Livre	23%	26%
	Rien/ Autre	5%	5%
Scs. SOCIALES	Aider les autres	22%	25%
	Être Aidé par les autres	27%	30%
	Étudier Seul	42%	46%
	Demander des Cahiers	12%	16%
	Prêter des Cahiers	9%	13%
	Étudier avec mon Livre	25%	29%
	Rien/ Autre	5%	5%

Source : élaboration de l'auteur. Lecture : les élèves qui déclarent aider les autres en sciences sociales correspondent, originalement, à 25% du total. Ce chiffre considère à tous les élèves n'ayant pas répondu correctement la question mais aussi à ceux qui ont bien répondu (i.e. 2 options maximum par discipline). Après redressement, cette quantité descend à 22% (i.e. une fois qu'on a choisi de prioriser les consistances entre pratiques).

Tableau B5

Comparaison entre les déclarations des élèves, redressées et non-redressées, selon leur degré de recouvrement. Par discipline et selon coïncidence entre pratiques d'étude.(2004 - 8^{ème} Grade) (n=278 476)

Discipline	PRATIQUES	Proportions		Corrélations de Kendall		Corrélation Tétrachorique (**)
		R	NR	R	NR	R
MATHS	Aidant Aidé	7.6%	9.4%	-0.18	-0.18	-0.30
	Aidé Seul	5.9%	8.8%	-0.35	-0.33	-0.56
	Aidant Seul	6.2%	9.0%	-0.17	-0.14	-0.30
	Aidé Livre	3.7%	5.8%	-0.07	-0.02	-0.15
	Aidant Livre	1.9%	3.6%	-0.10	-0.05	-0.24
	Seul Livre	3.8%	5.9%	-0.01	0.02	-0.03
LANG	Aidant Aidé	6.6%	8.1%	-0.12	-0.13	-0.22
	Aidé Seul	4.7%	7.4%	-0.36	-0.34	-0.59
	Aidant Seul	4.4%	7.0%	-0.30	-0.26	-0.51
	Aidé Livre	3.6%	6.2%	-0.09	-0.04	-0.19
	Aidant Livre	2.2%	4.5%	-0.13	-0.08	-0.29
	Seul Livre	6.7%	9.5%	0.03	0.06	0.06
Scs.NAT	Aidant Aidé	4.9%	6.4%	-0.06	-0.05	-0.12
	Aidé Seul	4.2%	6.8%	-0.34	-0.31	-0.58
	Aidant Seul	3.9%	6.4%	-0.26	-0.22	-0.47
	Aidé Livre	4.2%	7.0%	-0.11	-0.06	-0.22
	Aidant Livre	2.8%	5.1%	-0.12	-0.06	-0.26
	Seul Livre	9.7%	12.7%	0.01	0.03	0.01
Scs.SOC	Aidant Aidé	5.0%	6.5%	-0.06	-0.05	-0.12
	Aidé Seul	3.8%	6.5%	-0.35	-0.31	-0.59
	Aidant Seul	4.0%	6.6%	-0.26	-0.22	-0.48
	Aidé Livre	4.5%	7.3%	-0.12	-0.07	-0.23
	Aidant Livre	3.2%	5.8%	-0.13	-0.08	-0.27
	Seul Livre	11.0%	14.3%	0.02	0.04	0.03

Source : élaboration de l'auteur. Lecture : les élèves qui déclarent aider les autres en sciences de la nature et qui déclarent à la fois être aidés, sont, originalement 6,4%. Ce chiffre considère à tous les élèves n'ayant pas répondu correctement à la question mais aussi à ceux qui ont bien répondu (i.e. 2 options maximum par discipline). Après redressement (R), cette quantité descend à 4,9%. A droite on trouve les respectives corrélations de Kendall selon si la population a été redressée ou pas. (**) De même, on y voit les corrélations tétrachoriques, mais ne portant que sur la population redressée.

Tableau B6

Contributions partielles à l'inertie de chaque dimension ou colonne extraite par ACM (cf. graphique n°3) de la part des variables disjonctives reflétant les pratiques d'étude des élèves. Par discipline. (2004 - 8^{ème} Grade) (n=278 476)

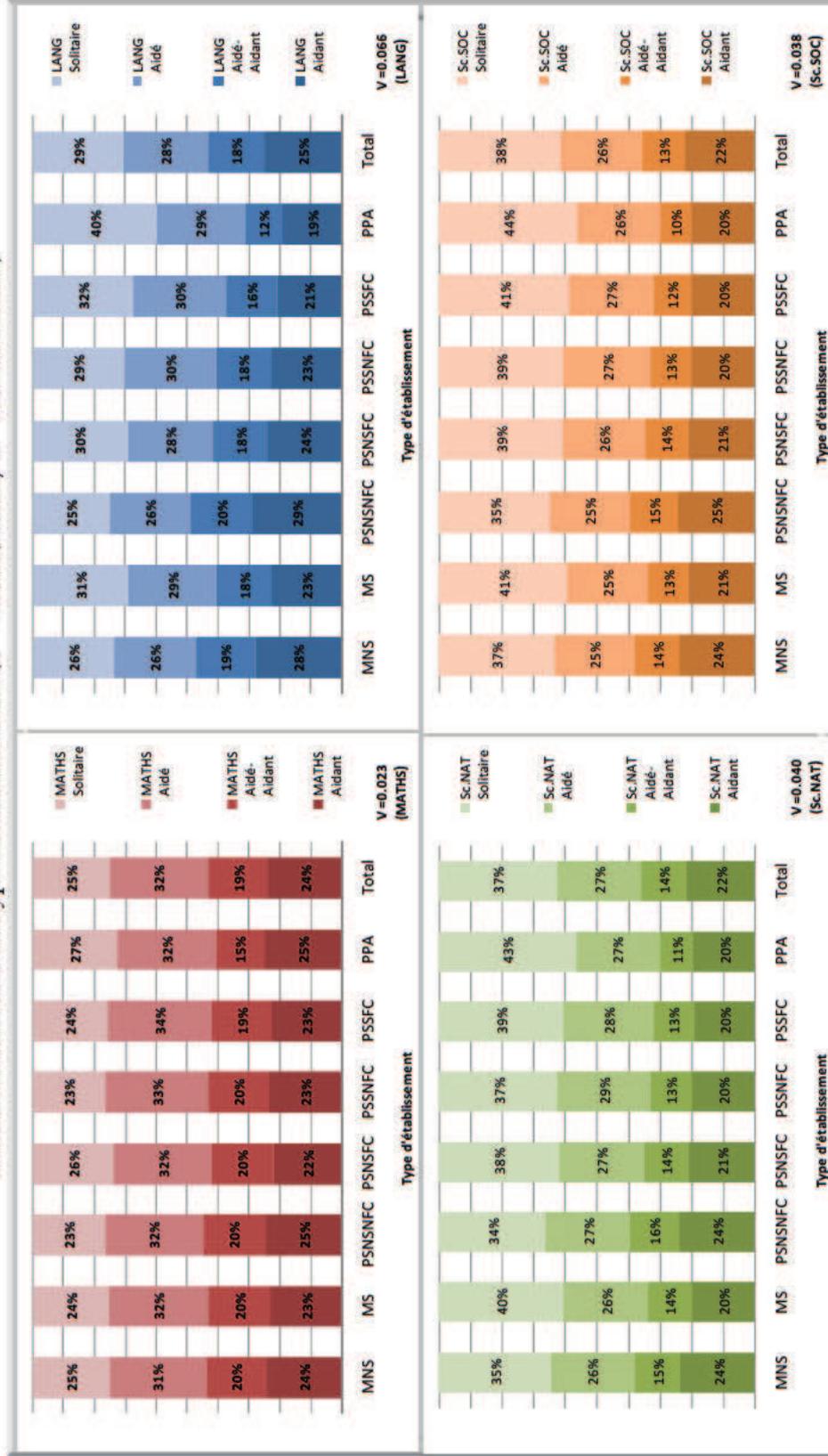
Mathématiques				Langage			
<i>Variable</i>	<i>Dim1</i>	<i>Dim2</i>	<i>Dim3</i>	<i>Variable</i>	<i>Dim1</i>	<i>Dim2</i>	<i>Dim3</i>
Aidant	0.03	0.30	0.02	Aidant	0.10	0.26	0.04
NO_Aidant	0.01	0.12	0.01	NO_Aidant	0.04	0.10	0.01
Aidé	0.27	0.00	0.03	Aidé	0.18	0.14	0.03
NO_Aidé	0.18	0.00	0.02	NO_Aidé	0.09	0.07	0.01
Solitaire	0.20	0.13	0.00	Solitaire	0.29	0.00	0.00
NO_Solit	0.10	0.07	0.00	NO_Solit	0.19	0.00	0.00
Demande_cahiers	0.14	0.03	0.06	Demande_cahiers	0.02	0.16	0.11
NodemCAH	0.02	0.00	0.01	NodemCAH	0.00	0.03	0.02
Prête_cahiers	0.03	0.23	0.01	Prête_cahiers	0.01	0.20	0.03
NOpretCAH	0.00	0.03	0.00	NOpretCAH	0.00	0.03	0.00
Livre	0.01	0.07	0.29	Livre	0.07	0.02	0.33
NO_Livre	0.00	0.01	0.04	NO_Livre	0.01	0.00	0.06
Autres	0.01	0.02	0.48	Autres	0.01	0.00	0.34
NOAutres	0.00	0.00	0.03	NOAutres	0.00	0.00	0.02

Sciences de la Nature				Sciences Sociales			
<i>Variable</i>	<i>Dim1</i>	<i>Dim2</i>	<i>Dim3</i>	<i>Variable</i>	<i>Dim1</i>	<i>Dim2</i>	<i>Dim3</i>
Aidant	0.09	0.27	0.04	Aidant	0.09	0.25	0.05
NO_Aidant	0.02	0.07	0.01	NO_Aidant	0.03	0.07	0.01
Aidé	0.20	0.09	0.04	Aidé	0.19	0.12	0.05
NO_Aidé	0.08	0.04	0.02	NO_Aidé	0.07	0.04	0.02
Solitaire	0.26	0.00	0.00	Solitaire	0.25	0.00	0.00
NO_Solit	0.19	0.00	0.00	NO_Solit	0.18	0.00	0.00
Demande_cahiers	0.04	0.20	0.06	Demande_cahiers	0.04	0.20	0.06
NodemCAH	0.01	0.03	0.01	NodemCAH	0.01	0.03	0.01
Prête_cahiers	0.01	0.24	0.02	Prête_cahiers	0.01	0.24	0.02
NOpretCAH	0.00	0.03	0.00	NOpretCAH	0.00	0.02	0.00
Livre	0.08	0.02	0.24	Livre	0.09	0.01	0.23
NO_Livre	0.02	0.01	0.07	NO_Livre	0.03	0.00	0.08
Autres	0.00	0.00	0.47	Autres	0.00	0.00	0.45
NOAutres	0.00	0.00	0.02	NOAutres	0.00	0.00	0.02

Source : élaboration de l'auteur. Lecture : chacune des sept affirmations proposées, par discipline, est notée ici de manière disjonctive (i.e. en considérant autant ceux qui ont coché chaque option que ceux qui ne l'ont pas coché). Les valeurs >0,1 ont été remarquées.

Graphique B1

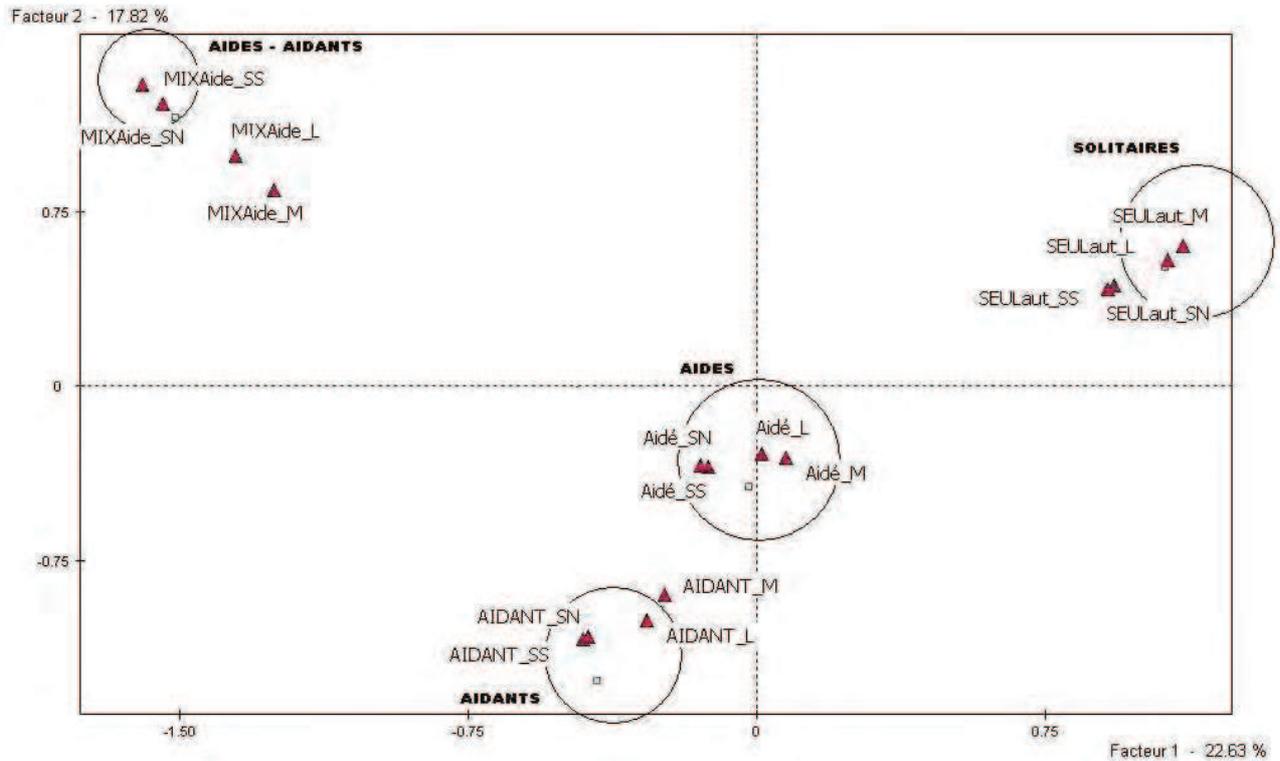
Pourcentages d'élèves selon profils exhaustifs par discipline, construits d'après leurs interactions et leurs manières d'étudier/déclarées. Par type d'établissement. (8^{ème} Grade, 2004 / n = 278 476 élèves)



Source : élaboration de l'auteur. Notons qu'en gros, les déséquilibres ne sont pas trop forts, sauf en langage (privé payant notamment). A l'opposé, en Mathématiques les profils exhaustifs construits semblent se montrer plus uniformes quel que soit le type d'établissement.

Graphique B2

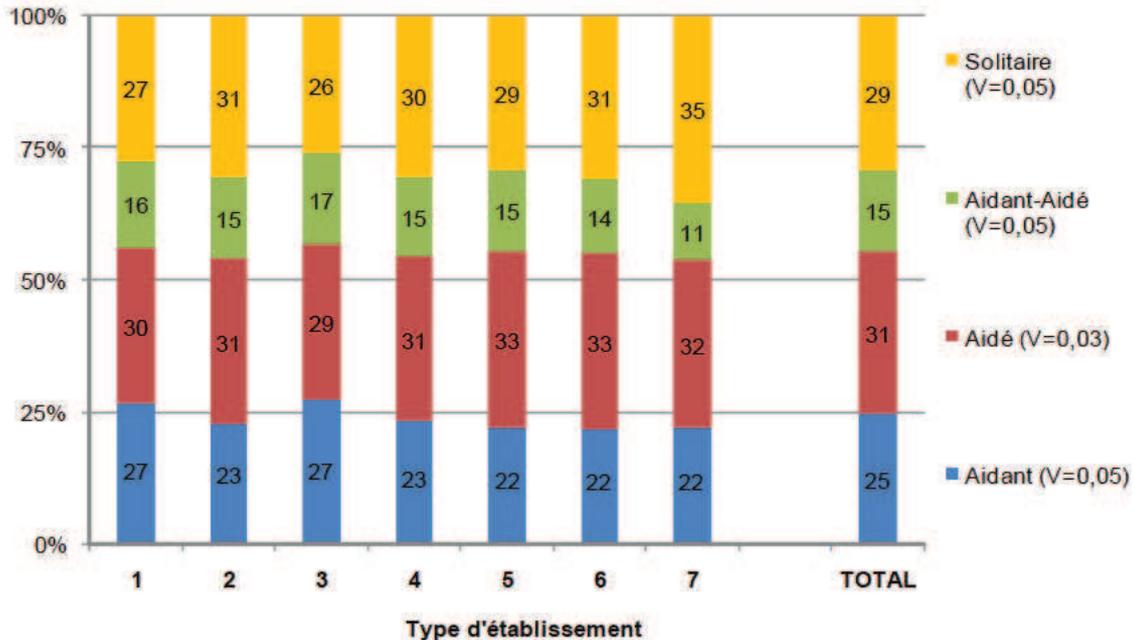
Typologie générale de regroupement des pratiques d'étude et d'interaction entre élèves sur un plan factoriel (ACM).(n=278 476 élèves) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : ce graphique complète le graphique n°4 au Tome I. Il illustre les dimensions n°1 et n°2 issues d'une analyse de correspondances multiples pour tester la séparation entre profils d'élèves (cf. chapitre n°8).

Graphique B3

Pourcentage d'élèves par typologie générale de regroupement des pratiques d'étude et d'interaction entre élèves. Selon type d'établissement.
(n=278 476 élèves) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

Lecture : pour le 8^{ème} grade en 2004, nous trouvons 25% d'élèves classés dans la catégorie générale « Aidant ». Ce pourcentage varie selon le type d'établissement où les élèves sont inscrits. Le coefficient V de Cramer noté à côté de chaque catégorie nous informe sur l'importance (entre 0 et 1) de ces différences. Par exemple, il y a 35% de « Solitaires » chez le privé payant tandis qu'ils ne sont que près de 27% chez les établissements non-sélectifs et gratuits, avec une moyenne nationale de 29%. Cet éloignement à la situation d'indépendance est significative quoi que pas trop forte (V de Cramer = 0,053). Chez les privés payants, on observe aussi une proportion réduite d'élèves en conditions d'interaction mixte (« Aidant-Aidés ») avec 11% contre 15% en moyenne. De même, les établissements gratuits non-sélectifs (privés ou publics) montrent une légère supériorité à l'heure de concentrer des élèves au profil « aidant » (27% contre 25% en moyenne). Une situation semblable s'avère vraie pour le cas des élèves en position d'interaction mixte. Notons enfin que les élèves de profil « Aidé » seraient en proportion similaire partout (près d'un élève sur trois).

Tableau B7

Proportion d'élèves ne déclarant aucune pratique d'étude principale.
Par discipline et selon type d'établissement. En pourcentage. (n=236113; 8^{ème} grade - 2004)

<i>Type d'école</i>	<i>n</i>	<i>MATHS</i>	<i>LANG</i>	<i>Scs. NAT</i>	<i>Scs. SOC</i>	<i>Ensemble (*)</i>
(1) M_NS	104105	15	13	17	16	5.5
(2) MS	17934	12	12	14	13	5.1
(3) PSUBV_NSNFP	13087	15	13	17	16	5.6
(4) PSUBV_NSFP	24875	14	13	15	15	6.0
(5) PSUBV_SNFP	8026	11	11	12	11	4.0
(6) PSUBV_SFP	51083	11	11	12	11	4.2
(7) Privé PAYANT	17003	9	10	10	9	3.7
Total	236113	13	12	15	14	5.1
<i>V de Cramer</i>		0.06	0.04	0.07	0.07	0.03

Source : élaboration de l'auteur. Lecture : dans le secteur public non-sélectif (n°1), 16% d'élèves ne déclarent aucune des pratiques d'étude définies ici comme principales pour sciences sociales. Ils ne sont que 14% à être dans la même situation pour le total d'élèves. (*) La catégorie « ensemble » désigne ici les élèves ne déclarant aucune des quatre pratiques pour aucune des disciplines. Elle n'est mise qu'à titre indicatif. Tous les tests Khi2 sont significatifs à p<0.000.

Tableau B8

Proportion d'élèves ne déclarant aucune pratique d'étude principale.
Selon genre et condition de redoublement. Par discipline et en pourcentage.
(8^{ème} grade - 2004 - Élèves dans des classes urbaines ≥20)

	<i>Filles</i>		<i>Garçons</i>		<i>TOTAL</i>	<i>V de Cramer</i>
	<i>NR</i>	<i>R</i>	<i>NR</i>	<i>R</i>		
MATHS	12.0	19.1	13.1	18.5	13.5	0.07
LANG	11.5	15.8	12.0	15.3	12.3	0.04
Scs. NAT	13.6	20.7	14.2	19.9	14.8	0.07
Scs. SOC	12.6	20.3	13.6	19.7	14.1	0.07
Ensemble (*)	4.1	6.6	5.2	7.6	5.1	0.05
<i>n</i>	103452	15092	96473	21096	236113	-

Source : élaboration de l'auteur. Lecture : NR = non-redoublant, R = redoublant. Tous les tests Khi2 significatifs à p<0.000. Globalement, on peut noter que le genre marque peu des différences bien que les garçons montreraient une toute légère tendance à se trouver davantage dans la situation « aucune pratique » et ceci, pour toutes les disciplines. Les différences fondamentales sont établies par la condition de redoublement (entre 15% et 20% contre 12% et 14%, garçons et filles tous confondus). Les filles (que l'on sait, sont toujours moins nombreuses à se trouver dans cette situation) montrent toutefois des proportions très légèrement plus hautes dans l'intersection des deux variables (autour de +1% en termes relatifs pour toutes les disciplines). Les redoublants semblent ainsi plus démunis scolairement. Il serait donc intéressant de tester si le fait d'interagir avec leurs camarades semble associé à des bénéfices en termes d'acquis scolaires.

Tableau B9

Corrélations de Pearson par classe entre les proportions déclarant chacune des pratiques d'étude et d'interaction, par discipline puis en général, et les proportions par classe qui déclare travailler souvent en groupe ainsi qu'avec la taille des classes.
(7 061 classes urbaines à plus de 19 élèves) (2004 - 8^{ème} grade)

% par Classe déclarant travailler souvent en groupes				
<i>% selon pratique déclarée par classe</i>	<i>Maths</i>	<i>Lang</i>	<i>Scs. NAT</i>	<i>Scs. SOC</i>
	23%	31%	43%	36%
% Aidant	<i>0.01</i>	0.12	0.06	0.05
% Aidé	<i>-0.02</i>	<i>-0.01</i>	<i>0.01</i>	0.05
% Et. Seul	<i>-0.21</i>	<i>-0.20</i>	<i>-0.11</i>	<i>-0.17</i>
% Et. avec son Livre	<i>-0.07</i>	<i>-0.18</i>	<i>-0.12</i>	<i>-0.16</i>
Taille de la Classe				
<i>% selon pratique déclarée par classe</i>	<i>Maths</i>	<i>Lang</i>	<i>Scs. NAT</i>	<i>Scs. SOC</i>
% Aidant	-0.03	-0.12	-0.06	-0.07
% Aidé	0.14	0.13	0.07	0.05
% Et. Seul	0.03	0.05	0.06	0.07
% Et. avec son Livre	-0.04	<i>0.01</i>	0.05	0.07
Profils généraux				
% Aidants (G)			-0.13	
% Aidants-Aidés (G)			<i>-0.01</i>	
% Aidés (G)			0.11	
% Solitaires (G)			<i>0.02</i>	

Source : élaboration de l'auteur. Lecture : en maths, entre la proportion d'élèves déclarant travailler souvent en groupes en cours et le pourcentage d'élèves déclarant être aidés par les autres, il y a une corrélation de Pearson de -0,02 qui n'est pas significative (à $p < 0,05$). En revanche, entre cette dernière proportion, toujours pour les mathématiques et la taille de la classe, il y a une corrélation significative de +0,14. Les chiffres en italique ne sont pas significatifs (à $p < 0,05$), ceux en gras le sont à $p < 0,05$, et tous les autres le sont à $p < 0,001$.

DIMENSIONS SUBJECTIVES

(cf. Chapitren°9, TOME I)

Le bien-être à l'école et la confiance en soi scolaire (remplissage des données manquantes)

Un point fondamental est celui du remplissage des données manquantes. Dans le cas du bien-être à l'école, nous avons vérifié que si 11% d'élèves (n=30 570) n'avaient pas répondu à l'ensemble des huit propositions faites, seulement 0,9% (n=2630) n'avaient répondu à aucune et 2% (n=5455) n'avaient pas répondu à la moitié (i.e. <4 items). Compte tenu de ces faits et du souci de perdre le moins possible d'information dans le total, nous avons décidé de prendre quelques risques pour le remplissage, sachant que les choix faits peuvent mener à des artefacts ou à des visions trop éloignées d'une réalité qui est déjà approximative. Mais sachant aussi que la non-réponse peut aussi être une source de biais. Nous avons donc fixé deux critères consécutifs. 1) Remplissage des données manquantes par corrélation maximale. Autrement dit, si l'élève avait répondu quelques options, on assignait la même valeur aux questions manquantes, mais si, et seulement si, les deux options avaient une corrélation maximale entre elles observée pour les données originales complètes. 2) Le cas échéant, le remplissage est fait en faisant appel à la moyenne générale de chaque proposition. Nous pouvons apprécier que les matrices de corrélations résultantes pour ces deux groupes (i.e. la population avec remplissage d'une part, et le 89% ayant tout répondu, d'autre part) sont fortement semblables, ce que l'on peut consulter dans les tableaux suivants.

Nous avons aussi constaté que les proportions des données manquantes étaient semblables quand on prenait l'échantillon d'élèves placés dans des classes urbaines de taille égale ou supérieure à 20. Puis, nous avons regardé les rapports entre nos indices et les classements construits à partir d'eux. D'un côté, en considérant toute la population (i.e. avec les remplissages faits) et d'un autre côté, en les construisant seulement avec les données remplies en entier (ici, le 89% ayant tout répondu). Les coefficients r de Pearson entre ces mesures étaient de 0,990 et les mauvais classements qui s'en suivaient étaient tous proximaux (i.e. jamais entre catégories extrêmes) atteignant le chiffre de 1,6% d'individus. Toujours est-il que nous allons garder comme nécessaire mesure de contrôle le fait d'éliminer ce 11% comme test de consistance pour nos analyses ultérieures (voir plus loin).

Dans le cas de la dimension de confiance en soi scolaire, 6,7% d'élèves (n=18760) n'a pas répondu en entier aux huit propositions, avec 1,1% (n=3 140) n'ayant répondu à aucune et 2,9% (n=7 991) ayant répondu à moins de la moitié. Pour le remplissage des données manquantes, nous avons suivi la même démarche et les mêmes critères que pour la question liée au bien-être scolaire (i.e. corrélation maximale entre propositions, puis moyenne générale pour chacune). Ceci, tout en gardant la possibilité d'éliminer par la suite les individus n'ayant pas répondu à tout et en ayant regardé aussi la consistance des classements faits avec une ou autre population (i.e. données originales et données remplies ; voir tableaux suivants). A cet égard, nous avons trouvé 0,96% de mal classés (jamais entre classes extrêmes), et un r de Pearson de 0,999 entre les indices construits sur les deux populations (avec des données redressés et avec des données non-redressées).

Tableau B10 : Matrices de Corrélations selon remplissage des données. Par type de corrélation. **Bien-être à l'école.** (2004 - 8^{ème} grade)

	MATRICES AVEC DES DONNÉES ORIGINALES								MATRICES AVEC DES DONNÉES REMPLIES							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Moyenne</i>	3.3	3.4	3.2	3.2	3.0	3.2	2.7	2.8	3.3	3.4	3.2	3.2	3.0	3.2	2.7	2.8
<i>Ecart-Type</i>	0.7	0.7	0.7	0.8	1.0	1.0	1.0	1.1	0.7	0.7	0.7	0.8	1.0	1.0	1.0	1.1
<i>n</i>	247910								278476							
SPEARMAN																
(1) J'y fais des amis	0.31								0.31							
(2) Je m'y sens bien	0.37	0.31							0.38	0.30						
(3) Je suis sympa	0.22	0.34	0.22						0.22	0.36	0.23					
(4) J'aime bien d'y aller	0.08	0.18	0.09	0.07					0.07	0.18	0.09	0.07				
(5) Je n'y me sens pas étrange	0.17	0.22	0.20	0.08	0.41				0.16	0.21	0.19	0.08	0.44			
(6) Je n'y me sens pas seul	0.07	0.19	0.09	0.20	0.24	0.30			0.07	0.18	0.09	0.19	0.25	0.30		
(7) Je ne m'y ennui pas	0.07	0.08	0.10	-0.01	0.20	0.23	0.17		0.07	0.08	0.09	-0.01	0.20	0.23	0.17	
(8) Pas de surnom fâcheux																
PEARSON																
(1) J'y fais des amis	0.32								0.32							
(2) Je m'y sens bien	0.37	0.31							0.38	0.30						
(3) Je suis sympa	0.20	0.33	0.21						0.20	0.35	0.21					
(4) J'aime bien d'y aller	0.06	0.16	0.08	0.05					0.06	0.15	0.08	0.05				
(5) Je n'y me sens pas étrange	0.16	0.20	0.19	0.06	0.40				0.15	0.19	0.19	0.06	0.43			
(6) Je n'y me sens pas seul	0.07	0.19	0.09	0.18	0.24	0.30			0.07	0.18	0.09	0.17	0.25	0.31		
(7) Je ne m'y ennui pas	0.07	0.08	0.09	-0.02	0.19	0.23	0.17		0.06	0.07	0.09	-0.02	0.19	0.22	0.17	
(8) Pas de surnom fâcheux																
Alpha de Cronbach = 0,6347																
Alpha de Cronbach = 0,6319																

Source : élaboration de l'auteur à partir des données SIMCE, 2004. Matrices de corrélation de Spearman et de Pearson estimées sur le logiciel SAS® pour les huit propositions à l'origine de la dimension de Bien-être à l'école (échelle Likert à quatre niveaux). Toutes les corrélations notées sont significatives (à p < 0,0000). Les Alpha de Cronbach correspondent aux estimations sur des valeurs standardisées (sur des valeurs brutes, ces coefficients passent, respectivement, à 0,619 et 0,618).

Tableau B11 : Comparaison des Matrices de Corrélation selon remplissage des données. Par type de corrélation.
Dimension **Confiance en soi scolaire** (*academic self-concept*) (2004 – 8^{ème} grade)

	MATRICES AVEC DES DONNÉES ORIGINALES								MATRICES AVEC DES DONNÉES REMPLIES							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Moyenne</i>	3.0	3.3	3.6	3.4	2.6	3.2	3.4	2.9	3.0	3.3	3.6	3.4	2.6	3.2	3.4	2.9
<i>Ecart-Type</i>	0.8	0.8	0.6	0.8	1.0	0.9	0.8	1.0	0.8	0.8	0.7	0.8	1.0	0.8	0.8	1.0
<i>n</i>	259720								278476							
SPEARMAN																
(1) Je me sens capable...									0.44							
(2) Je peux faire un excellent travail...	0.44								0.28	0.42						
(3) Je peux bien apprendre tout	0.28	0.42							0.25	0.35	0.42					
(4) Je peux ne pas avoir de mauvaises	0.25	0.35	0.42						0.06	0.09	0.06	0.05				
(5) Je ne me déconcentre pas	0.06	0.09	0.06	0.04					0.31	0.39	0.33	0.31	0.11			
(6) Je fais des efforts même si c'est difficile	0.32	0.39	0.33	0.31	0.10				0.28	0.39	0.36	0.32	0.13	0.59		
(7) Je m'efforce pour apprendre	0.28	0.39	0.37	0.32	0.12	0.58			0.25	0.28	0.22	0.20	0.07	0.39	0.38	
(8) Je cherche de l'information si besoin	0.25	0.28	0.23	0.21	0.08	0.39	0.38									
PEARSON																
(1) Je me sens capable...									0.45							
(2) Je peux faire un excellent travail...	0.45								0.29	0.44						
(3) Je peux bien apprendre tout	0.29	0.43							0.23	0.33	0.41					
(4) Je peux ne pas avoir de mauvaises	0.24	0.33	0.41						0.06	0.09	0.05	0.03				
(5) Je ne me déconcentre pas	0.06	0.08	0.06	0.02					0.31	0.38	0.32	0.28	0.10			
(6) Je fais des efforts même si c'est difficile	0.31	0.38	0.33	0.28	0.10				0.27	0.37	0.36	0.28	0.12	0.59		
(7) Je m'efforce pour apprendre	0.28	0.38	0.37	0.28	0.11	0.58			0.25	0.28	0.23	0.18	0.07	0.39	0.39	
(8) Je cherche de l'information si besoin	0.25	0.28	0.23	0.19	0.08	0.39	0.39									

Alpha de Cronbach = 0,747

Alpha de Cronbach = 0,747

Source : élaboration de l'auteur à partir des données SIMCE, 2004. Matrices de corrélation de Spearman et de Pearson estimées sur le logiciel SAS® pour les huit propositions analysées au départ pour établir une dimension d'Estime de soi scolaire (échelle à quatre niveaux). Toutes les corrélations notées sont significatives (à $p < 0,0000$). Les Alpha de Cronbach correspondent aux estimations sur des valeurs standardisées (sur des valeurs brutes, ces coefficients passent, respectivement, à 0,7466 et 0,7465).

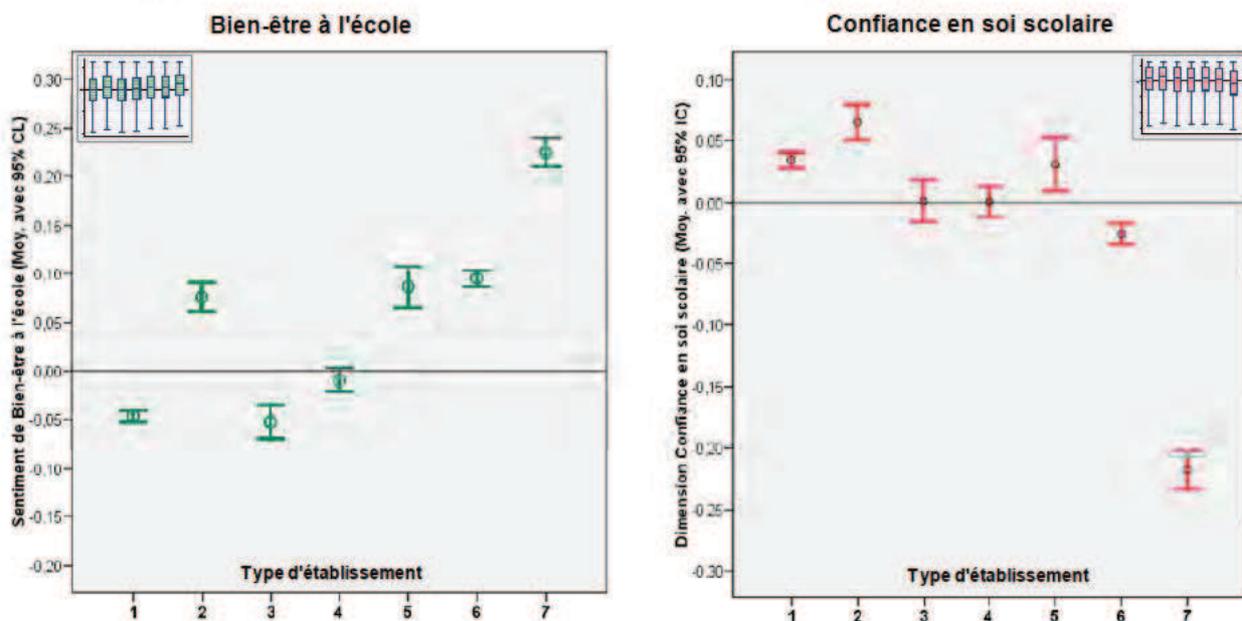
La distribution du BEE et de la CSS selon type d'école

Le graphique B4 illustre les différences par catégorie d'établissement entre les scores moyens sur les échelles de bien-être à l'école et de confiance en soi scolaire. Les rapports illustrés correspondent à un zoom des distributions par type d'école sur les échelles construites. Comme les variations sont plutôt faibles nous ne montrerons pas l'échelle originale (de 1 à 4) ce qui servirait à regarder s'il y a des variations qu'impliquent un changement de niveau. Ici le rang va de 3,05 à 3,25 rappelant que la moyenne de l'ensemble est de 3,09. Nous voyons que les différences sont légères (R^2 de 0,005 et de 0,007 respectivement) quoi qu'indicatives de différences.

Graphique B4

Moyennes avec leurs limites de confiance des deux principales dimensions subjectives, par type d'établissement.

(23 6113 élèves urbains dans des classes à taille ≥ 20) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC ou MUNICIPAL Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). En haut sur chaque graphique on voit une vision de rappel sur l'étendue de chaque distribution, dont les limites de confiance des moyennes représentées sur le graphique ne constituent qu'un zoom. Les scores ont été standardisés ($\mu=0$; $\sigma=1$) et la ligne centrale de chaque graphique représente la moyenne générale de chaque dimension.

Tableau B12

Proportions d'élèves en condition de haute et de basse Confiance en soi scolaire (CSS) et de Bien-être à l'école (BEE), selon type d'établissement. En pourcentage. (236 113 élèves urbains dans des classes à taille ≥ 20) (2004 - 8^{ème} grade)

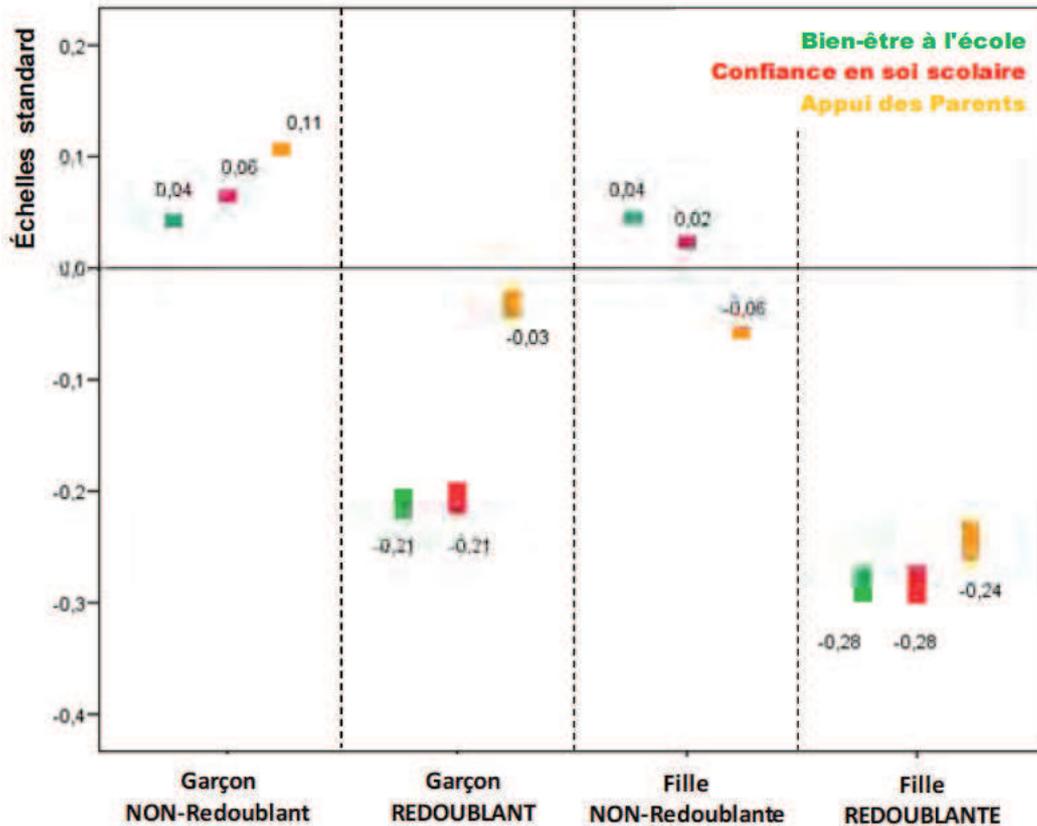
Type d'établissement	Bas BEE	Haut BEE	Bas CSS	Haute CSS
(1) M_NS	35	31	31	36
(2) M_S	30	36	30	36
(3) PS_NSNFP	35	31	33	35
(4) PS_NSFP	33	33	32	34
(5) PS_SNFP	29	37	32	36
(6) PS_SFP	29	37	33	33
(7) PPAY	25	44	42	26
TOTAL	32	34	33	34
<i>V de Cramer</i>	0.07	0.08	0.06	0.06

Source : élaboration de l'auteur. Lecture : dans les établissements publics non-sélectifs (M_NS ou n°1), 35% d'élèves est classé en condition de bas Bien-être à l'école tandis que 31% l'est dans le cas contraire (haut BEE). Dans la moyenne générale, pour cette population, ils ne sont respectivement que 32% et 34%

Le résultat le plus saillant est l'écart opposant les élèves du privé payant et les autres (cf. graphique B4). Les élèves du privé payant montrent un score moyen davantage positif (+0,22) pour l'échelle de bien-être, mais à la fois, un score moyen de confiance en soi scolaire tout à fait opposé. Globalement ce qui ressort des valeurs moyennes illustrées sur le graphique B4 est que les élèves des établissements sélectifs (n°2, 5 et 6) se situent tous au-dessus de la moyenne générale de bien-être. En même temps, la confiance en soi scolaire est moins forte chez les catégories privées à public fort favorisé (n°6 et n°7, avec -0,03 et -0,22 respectivement), ce qui semble conforter l'hypothèse du BFLPE. Notons enfin que les catégories du privé subventionné non-sélectif sont dans la moyenne nationale mais que le public non sélectif se détache légèrement vers le haut (+0,036). Le tableau B12 ne fait qu'exprimer les mêmes rapports mais pour les catégories construites à partir des échelles (cf. chapitre n°9).

Graphique B5

Valeurs moyennes avec limites de confiance pour trois dimensions subjectives du vécu scolaire selon genre et condition de redoublement.
(236 113 élèves urbains dans des classes à taille ≥ 20) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : En moyenne les garçons redoublants obtiennent un score sur l'échelle de Bien-être à l'école de -0,21 points d'écart-type, valeur dont les limites de confiance se situent entre -0,20 et -0,22 (à $p < 0,05$). Les R^2 pour ces trois dimensions rapportées à l'interaction des variables genre et redoublement sont équivalentes, respectivement, à 1,1%, 1,1% et 1,0% de variance expliquée. Les scores ont été standardisés et la moyenne générale pour chacune des mesures est « 0 ».

Autant sur l'échelle de bien-être que sur celle de confiance en soi scolaire, les redoublants (tous genres confondus), obtiennent -0,24 points en moyenne contre +0,04 chez les non-redoublants. En revanche, filles et garçons ne sont pas véritablement éloignés vis-à-vis de ces deux dimensions. Les filles dépassent à peine à leurs camarades en bien-être (écart de 0,01) tout en se montrant moins confiantes sur elles-mêmes, en moyenne (écart de 0,03 points entre les deux groupes, significatif à $p < 0,001$). La condition de redoublante chez les filles semble accentuer les écarts. Si les hommes redoublants totalisent -0,21 points pour ces deux dimensions, les filles dans cette condition montrent un score moyen de -0,28 pour les deux cas. Enfin, la seule variable modulée par les deux facteurs en même temps est

l'appui des parents ressenti (cf. annexes plus loin). Les filles montrent un score moyen de -0,08 pour l'ensemble de la variable genre, mais les filles redoublantes (significativement moins nombreuses) obtiennent une valeur moyenne de -0,24 pour ce sentiment, et les autres -0,06. Les garçons communiquent une plus forte présence des parents dans leurs études (+0,08 comme moyenne de l'ensemble) et leurs redoublants restent proches de la moyenne générale (-0,03). Ces constats pourraient avoir trait à un métier d'élève mieux exercé par les filles et/ou à des parents qui font plus de confiance aux filles.

CONSTRUCTION des VARIABLES et des MESURES COMPLEMENTAIRES

Proposition d'une typologie originale d'établissements

D'après notre revue de littérature (cf. chapitre n°1 et n°6), un premier cadre à définir ce sont les différents niveaux contextuels et les variables qui leur sont attachées. Dans ce sens, aborder en premier lieu la problématique de la typologie d'écoles nous semble obligatoire. Bien que, d'après certains au Chili, ceci reste un sujet pas vraiment important (cf. Gallego et Sapelli, 2007), depuis quelques années le public accueilli par chaque type d'établissement, ainsi que les mécanismes liés à la conformation de ces publics ont été de plus en plus pris en compte. Premièrement, pour contrôler la présence des possibles effets de pairs (cf. chapitres n°2 et n°6). Puis, en considérant le fait que les familles choisissent non seulement l'école de leur goût et/ou possibilités, mais qu'à la fois, beaucoup d'établissements choisissent aussi leur public.

En effet, ce dernier aspect a été longtemps négligé dans les analyses, soit par manque d'information, soit par absence d'intérêt. Nous considérons toutefois qu'il est incontournable dans le cas chilien.⁶ Un autre sujet capital est celui du « financement partagé » (FP) que certains établissements peuvent demander sous la forme des frais de scolarité en plus des subventions octroyées par l'État (cf. chapitre n°6). S'il n'existe pas à notre connaissance assez de recherches portant sur l'efficacité nette de ce type de dispositif, nous savons qu'il a été la plupart du temps pris comme une variable isolée censée influencer directement les résultats des élèves. Néanmoins, il est clair qu'il a augmenté la déjà forte segmentation sociale du système scolaire (cf. chapitre n°6).

Ainsi, d'un point de vue théorique on peut s'atteindre au fait de « payer » soit relié à la motivation, aux pratiques, aux dispositions et aux attentes des familles vis-à-vis de la réussite à l'école et/ou des apprentissages. De ce fait, il se peut que les familles prenant cette modalité soient, en moyenne, plus exigeantes ou attentives, soit, envers leurs enfants, leur travail et leurs comportements scolaires, soit, envers le personnel directif et/ou les enseignants des écoles ; soit enfin, les deux à la fois. En outre, d'un point de vue de la théorie économique, la plus ou moins adéquate combinaison des facteurs productifs peut aussi expliquer les différences observées entre résultats, en l'occurrence, les acquis des élèves. On sait que la plupart du temps cette combinaison est difficilement observable et

⁶ De manière pareille, nous pensons qu'on devrait pouvoir rentrer aussi dans deux sujets presque inexplorés : les lycées non-mixtes et la constitution des classes de niveau au sein des établissements. Ce dernier aspect n'a pratiquement pas été documenté par la recherche locale ni encore moins au niveau officiel. Pourtant, nous ne traiterons pas ici ces deux problèmes.

parfois non-mesurable. En conséquence, la propriété des écoles ou leurs cadres normatifs, parmi d'autres caractéristiques des établissements scolaires, peuvent être utilisées pour s'en approcher. Ceci suppose d'établir des typologies, qui à leur tour peuvent être utilisées comme des variables cible et/ou comme facteurs de contrôle.

La catégorisation proposée pour classer les établissements au Chili

D'après ce qu'on vient de dire, nous avons cru pertinent de proposer une typologie d'établissements ancrée dans l'intersection de la propriété/administration de ceux-ci, de demander ou pas des frais de scolarisation, et enfin, de la sélectivité académique à l'entrée. A partir de là, nous avons décidé de considérer pour nos analyses huit types d'établissements, puis sept. Ceci comme une variable prioritaire de contexte. Pour y parvenir, nous avons pris en considération trois critères. Premièrement, la division la plus évidente et la plus abordée par la littérature, entre le secteur privé payant et le secteur subventionné par l'État, à son tour sous-divisé en propriété privée (privé subventionné) et propriété publique (municipale). Nous avons ainsi trois groupes distincts, dont la division entraîne des différences liées à la propriété, l'histoire, l'orientation, la régulation et à l'administration des établissements.

Deuxièmement, nous avons considéré que le fait que les établissements puissent choisir leur public en termes académiques marque une distinction fondamentale, notamment, depuis la perspective des interactions et des normes construites par les communautés scolaires au sein de ces établissements (cf. chapitre n°1).⁷ En effet, il n'est pas possible de faire fi des implications probables d'un tel mécanisme sur l'esprit de travail, les normes ou les attentes au sein de ces établissements. Il en est de même avec les rapports au comportement des élèves ainsi que pour l'implication des parents dans la scolarisation de leurs enfants et dans les relations avec le corps enseignant et le personnel directif. Plus encore, rappelons que Bidwell (2000) note que la *narration* et la vision que chaque corps enseignant se fait de leur propre mission est fortement déterminée par des moments dits fondateurs pour un établissement. Il est donc fort probable qu'un certain esprit commun sous une configuration du type *functional community* (cf. chapitre n°1),⁸ mais plus

⁷ Une distinction cohérente avec ceci aurait été établie par Merton (1957, cité par Dreeben, 2000) quand il insiste sur le fait que l'agrégation des caractéristiques ne constitue pas forcément un climat normatif : « *Social categories can become transformed into cohesive collectivities, but the distinction between collectivities and categories suggest that aggregating individual characteristics into social categories does not necessarily indicate a normative climate.* » (p. 112). En effet, Dreeben (2000) met en exergue la distinction entre les concepts de catégorie et de collectivité proposée par Merton. Si la première peut contenir des personnes ayant des caractéristiques communes (âge, sexe, race, etc.), la deuxième désigne un groupe capable d'interactions, de la participation et de la tolérance vis-vis de normes plus ou moins partagées.

⁸Toujours est-il que nous avons laissé hors de ce travail le phénomène des établissements religieux. Il faut dire néanmoins que c'est justement le caractère religieux des écoles celui qui se trouve à l'origine de la notion *functional community*. Ceci sachant que la deuxième notion (*value community*) recouvre la première et que l'important est donc le cadre normatif qu'elle installe et les interactions que entérinent et construisent ce cadre (Dreeben, 2000).

précisément du type *value community* (cf. chapitre n°1) soit présent à partir de cette sélectivité. Nous postulons alors que ces conditions émergent justement de par l'ensemble des mécanismes d'encadrement et de travail qui sont en cohérence avec le fait d'accueillir une population académiquement triée. Par conséquent, nous avons décidé de regarder la variable sélection académique à l'entrée des établissements, non-pas de manière isolée comme cela a été fait précédemment (e.g. Contreras et al., 2007), mais plutôt en interaction avec les autres paramètres déjà cités.

En dernier lieu, et pour des raisons semblables au critère précédent, nous avons décidé de considérer d'une façon spécifique le fait qu'à l'intérieur du sous-groupe d'établissements privés subventionnés, il y ait des écoles où les parents payent des frais supplémentaires de scolarisation.⁹ A notre avis, cette caractéristique est susceptible de représenter des volontés et des dispositions particulières, autant de la part des propriétaires de ces écoles que des acteurs à leur charge. En effet, on pourrait s'attendre à des comportements (de la part des enseignants, du personnel directif, des élèves et des familles), modulés par le fait que « l'on paie » quelque chose, d'autant plus qu'on a été admis par un mécanisme quiconque, surtout académique.

Alors, un certain esprit corporatif du type *functional community* pourrait émerger à partir de l'intersection de ces situations. Pourtant, ce qui nous intéresse davantage et qui nous pousse à considérer ces trois critères à la fois et non pas séparés, c'est la constitution des populations scolaires intra-établissements assez homogènes. Ceci représente un défi majeur pour les estimations reliées aux effets de pairs (cf. chapitre n°2). Cela devrait augmenter la diversité des publics entre établissements (en termes de tonalités scolaire et sociale), tout en les diminuant à la fois au sein de ceux-ci. Cette dernière hypothèse opérationnelle sera testée en général à l'heure de considérer la distribution de certaines variables parmi notre typologie d'établissements, dont les principales seront : scores aux tests ; scolarité des parents ; proportion de redoublants inscrits ; et attentes des acteurs. Rappelons que la distinction entre catégories d'école prend sa légitimité, d'une part, depuis une perspective analytique dans le sens où la réalité se présente en quelque sorte sous cette forme et qu'un décorticage donné serait donc pertinent. D'autre part, cette distinction reste importante dans la mesure où nous pouvons soupçonner la présence d'une combinaison particulière des facteurs, attachée à chaque type d'établissement et qu'il n'est pas possible de saisir autrement.

Enfin, rappelons que la recherche a noté qu'il existe de la diversité dans tous les sous-secteurs scolaires au Chili (cf. chapitre n°6) et qu'il y a même de la sélection académique au sein du secteur public. Pourtant, peu de choses ont été dites vis-à-vis de la signification de

⁹ Chez les municipaux nous avons obvié cette considération en attention à la loi. Pourtant, on peut soupçonner de mécanismes semblables ou équivalents chez certaines écoles, par exemple, voie les quotas de collaboration avec l'association de parents de l'établissement.

ceci en termes des publics accueillis, des résultats obtenus, et encore moins, par rapport aux perceptions sur le vécu scolaire des acteurs. De fait, les établissements publics sont, la plupart du temps, pris comme un ensemble homogène (cf. chapitre n°6) alors que l'on sait qu'ils tendent aussi à se différencier. Bref, eu égard ces différentes considérations théoriques nous croyons que notre proposition d'une typologie singulière, non construite auparavant, se justifie de plein droit.

Stratégie d'identification des établissements

Notre stratégie d'identification empirique des établissements qui sélectionnent leurs élèves et/ou qui sont attachés au mécanisme de financement partagé, a suivi quatre pas. D'abord, nous avons pris la typologie de référence fournie par le Ministère d'Éducation (privés payants, privés subventionnés, municipaux). En second terme, et sachant que la sélection académique n'est pas officiellement consignée ni communiquée, nous avons utilisé l'information fournie par les parents en référence aux exigences que leurs enfants ont subi au moment d'entrer à l'établissement. Comme nous pouvons l'apprécier sur le tableau B13, les parents ont été confrontés à neuf questions portant sur les exigences faites pour entrer à chaque école. Ces exigences allant de quelques papiers administratifs demandés – certificat de naissance ou de mariage civil – à la passation des différents types de test de sélection, l'accréditation des évaluations préscolaires ou la fourniture de certificats des notes antérieures ; en passant par l'accréditation des situations sociales données – certificats de revenus, de baptême et/ou de mariage religieux – ou bien, par l'entretien direct avec les parents.¹⁰

Compte tenu du fait que notre intérêt porte sur l'identification la plus précise possible du phénomène de sélection académique en dépit des limitations existantes, nous avons choisi, parmi ces options, celles qui gardent relation directe avec les capacités scolaires des élèves. Ainsi, un élève est considéré comme ayant été académiquement sélectionné si ses parents déclarent qu'il est passé par un type de test ou de séance des jeux, ou bien, s'ils ont dû fournir des évaluations préscolaires pour l'inscrire. Nous avons décidé de ne pas considérer l'exigence des certificats des notes antérieures car il peut s'agir d'une mesure administrative pas forcément vouée à la sélection. Conséquemment, nous avons décidé de qualifier un établissement comme étant académiquement sélectif si la moitié des parents ou plus, déclaraient que leurs enfants avaient subi au moins l'un des trois types de sélection. D'ailleurs, nous pouvons apprécier sur le tableau B13 que, bien qu'il existe une cohérence

¹⁰ Notons que, en ce qui concerne la sélectivité académique des établissements, déclarée par les familles, les proportions des colonnes des sous-totaux du Tableau B12 sont en phase avec la littérature préalable. Toutefois, nos pourcentages sont plus élevés que ceux notés pour le 4^{ème} grade par Contreras et al. (2007), notamment dans le cas du secteur municipal. Ils sont plus bas que ceux notés pour le 10^{ème} grade par Bellei (2009). De fait, Contreras et al. (2007) travaillent avec des élèves de 4^{ème} grade en 2005 et communique un 5% de sélectivité chez le secteur public et un 48% chez le secteur privé subventionné. Ils n'analysent pas le secteur privé payant. Pour sa part, Bellei (2009) note, respectivement, un 56%, un 72% et un 85% pour le secteur public, privé subventionné et privé payant. Ces deux études n'établissent pas de sous-catégories.

avec nos estimations (i.e. que chez les établissements classés comme étant académiquement sélectifs les déclarations d'au moins deux parents sur trois s'ajustent aux critères choisis) il existe aussi une marge d'erreur importante de fait des imprécisions dans ce type de déclaration, pourtant seul disponible.

De manière semblable, et en troisième place, nous avons calculé la moyenne des frais de scolarité pour tous les établissements. Ceci à partir de ce que les parents déclaraient devoir payer par mois sur une échelle ordinaire de dix-neuf niveaux, allant de zéro à plus de deux cents mil pesos chiliens (Ch\$) (i.e. de 0 à ≈315 €).¹¹ Au départ, nous avons classé un établissement comme étant demandeur des frais de scolarisation si cinquante pour cent ou plus des parents déclaraient payer quelque somme. Puis, nous avons comparé nos chiffres avec des données administratives additionnelles du Ministère comptant les établissements officiellement inscrits sous la modalité de financement partagé. Avec ces informations nous avons décidé de classer un établissement comme adscrit à cette modalité si : a) il l'était officiellement ; b) il appartenait au sous-secteur privé subventionné¹²; et c) si dans nos estimations il apparaissait comme demandeur des frais, mais, avec un degré minimal de cohérence (i.e. avec une tendance à l'unanimité dans les déclarations des parents).¹³

Le tableau B14 montre les totaux d'effectifs, notés au niveau établissement, pour nos huit catégories. Aux estimations faites pour la distribution du total des effectifs de l'enseignement primaire chilien (1^{er} à 8^{ème} grade), s'ajoutent celles des élèves et des établissements enquêtés en 2004. En même temps, nous communiquons les principaux chiffres estimés par rapport aux frais de scolarisation (en 2004) pour chacune des catégories concernées par cette modalité. Ces estimations nous permettent d'avoir une vision générale sur la division du système scolaire selon la catégorisation que nous présentons comme base pour notre travail. Pour la suite des analyses, sauf quelques exceptions, ces huit catégories d'établissement céderont la place à sept. En effet, nous allons fusionner les établissements privés payants dans une seule catégorie, compte tenu de la faiblesse des effectifs du privé payant académiquement non-sélectif (0,8% des enquêtés) et des caractéristiques assez semblables entre ces deux sous-secteurs.

¹¹ Notons que l'analyse détaillée que nous avons entreprise nous a permis de trouver quelques mauvais classements de la part des informations officielles fournies par le Ministère. En effet, 77 établissements (1,4% du total) n'avaient pas été bien classés selon leur régime de propriété. Nous avons corrigé et fait noter ce détail.

¹² Le secteur municipal étant exclu par la Loi. Remarquons aussi que le degré de cohérence entre ces deux chiffres (i.e. celui issu du Ministère et celui issu de nos estimations) était assez élevé pour ce sous-groupe d'établissements (r de Spearman = 0,82), considérant, de plus, que 93% des non-coïncidences ($n=276$; 4,9% du total) ont lieu dans le sens « n'apparaît pas dans la base du Ministère mais apparaît dans nos estimations ».

¹³ Cette tendance a été mesurée, 1) à l'aide du coefficient de variation par établissement (i.e. le ratio entre l'écart-type et la moyenne des frais déclarés), notamment pour des $CV < 100$; autrement, on a estimé que l'établissement n'avait pas cette politique ; et 2) sur un montant de frais moyen supérieur à la moyenne des établissements publics (i.e. >5,5 € par mois).

Tableau B13

Proportions (en %) des familles/parents déclarant que leurs enfants ont été objet de différentes exigences au moment de les inscrire dans l'établissement où ils préparent le 8^{ème} grade (2004). Selon la typologie des établissements proposée.

Type d'Etablissement/ Exigence faite par l'établissement	1 MNS	2 MS	Subtotal Municipal	3 PS_NSNFP	4 PS_NSEF	5 PS_SNFP	6 PS_SFP	Subtotal Privé Subventionné	7 PPA_NS	8 PPA_S	Subtotal Privé Payant	TOTAL
A Certificat NAISSANCE	96.5	96.6	96.5	95.8	96.0	96.8	96.5	96.3	92.2	94.1	93.8	96.3
B Certificat BAPTEME (**)	1.6	1.7	1.6	8.2	7.4	40.1	29.2	21.2	11.3	38.3	35.2	11.5
C Certificat MARIAGE	8.6	7.6	8.5	9.0	8.6	18.9	18.6	14.5	14.1	30.9	29.0	12.2
D Certificat REVENU	6.1	7.2	6.3	5.2	9.0	9.3	15.9	11.9	5.8	7.0	6.9	8.5
E INTERVIEW Parents	19.3	26.5	20.1	26.5	30.4	46.0	52.7	42.2	58.9	80.9	78.4	32.6
F EXAMEN d'Admission	5.7	40.3	9.9	7.7	15.1	59.5	67.3	43.7	21.2	76.6	70.3	27.1
G Eval. voie SEANCE DE JEUX	1.9	6.1	2.4	2.3	2.2	9.3	8.4	5.9	3.7	19.2	17.5	4.8
H Evaluations Précolaires	21.5	37.3	23.4	16.7	20.5	35.8	36.5	29.2	21.0	39.7	37.6	26.6
I Certificat des NOTES	39.8	54.6	41.6	50.4	61.9	44.0	56.0	55.6	64.2	49.5	51.1	47.7
Sél. Acad. (F à H) + Notes (I)	54.9	84.0	58.4	61.4	74.4	85.5	90.3	81.2	77.0	92.7	90.9	69.4
Sélection Sociale (B à E)	28.3	34.4	29.1	36.1	40.5	65.7	65.3	54.3	65.1	84.5	82.3	42.4
Sél. Académique (F à H)	25.8	66.8 (*)	30.8	23.2	31.6	73.3	77.5	56.6	34.9	86.2	80.4	44.2
N° Elèves (Parents)	125304	17440	142744	16145	25743	8350	50237	100475	2026	15891	17917	261136
N° Etablissements	2839	194	3033	544	579	130	801	2054	94	396	490	5577

Source : élaboration de l'auteur. **Lecture (*)** : en moyenne, les parents de 66,8% d'élèves inscrits dans les 194 établissements municipaux que nous avons qualifiés comme étant académiquement sélectifs, déclarent que pour y inscrire leur enfant, il a dû passer au moins l'une des suivantes épreuves : rendre un examen d'admission, avoir participé à une séance de jeux ou avoir fourni des évaluations préscolaires. Ces trois critères constituent la base de notre classification, donc, ils sont mis en gras. Notons que sur le total d'établissements (n=5 612) sur 35 ces données n'ont pas pu être estimées. De même, 2 847 parents (1,1% des répondants) n'ont marqué aucune option et 1 400 n'ont pas été communiqués (0,53% du total des questionnaires des parents (cf. tableau N°3)). Les établissements manquants ont été assignés aux catégories respectives de base et les individus ont suivi, soit cette même règle, soit ils ont été assignés aux conditions établies à partir des données existantes. **Les catégories d'établissement seront en général associées aux numéros (1 à 7 et des fois, 1 à 8)**. Pour plus de clarté, les sigles signifient : M=Municipal (ou Public) ; PS = Privé Subventionné ; PPA = Privé Payant. Les sous-fixes NS = Non Sélectif ; S = Sélectif ; NFP= Non Financement Partagé ; FP= Financement Partagé. (**) Le certificat de Baptême pouvait être aussi Certificat de Mariage Religieux.

Tableau B14

Distribution d'effectifs inscrits à l'enseignement primaire et d'effectifs enquêtés en 8^{ème} grade (2004) pour les tests SIMCE. Selon typologie proposée d'établissements. Frais de scolarisation estimés et nombre d'établissements concernés.

Type d'établissement		Effectifs Niveau PRIMAIRE (1er à 8ème Grade) en 2004		Effectifs Enquêtés (8ème Grade) en 2004		Etablissements en 2004 (Enquêtés)		Montant moyen des frais de scolarisation demandés (2004) (en ≈€)			
Code	Type	n	%	n	%	n	%	Moy	E-T	p90	p10
1	MNS (**)	1014762	47,0	133417	47,9	2860	51	-	-	-	-
2	MS	127231	5,9	18361	6,6	194	3,5	-	-	-	-
3	PS_NSFP	144973	6,7	17721	6,4	544	9,7	-	-	-	-
4	PS_NSFP	218966	10,1	27557(*)	9,9	593	10,6	18	14	41	6
5	PS_SFP	73101	3,4	8623	3,1	130	2,3	-	-	-	-
6	PS_SFP	427987	19,8	52777	19,0	801	14,3	27	18	58	10
7	PPA_NS	16555	0,8	2236	0,8	94	1,7	97	54	167	34
8	PPA_S	137270	6,4	17784	6,4	396	7,1	165	60	247	90
TOTAUX		2160845	100	278476	100	5612	100	57	67	247	6

Source : élaboration de l'auteur sur des données du Ministère de l'Éducation et bases SIMCE 2004.

Lecture (*) : Un total de 27 557 d'élèves de 8^{ème} grade, inscrits dans l'un des 593 établissements privés subventionnés non-académiquement sélectifs et adscrits au système financement partagé (n°4), ont passé les tests et questionnaires SIMCE 2004. Ils correspondent à 9,9% du total d'élèves enquêtés ; leurs familles payent, en moyenne, 18 Euros par mois comme frais de scolarité. Note : les chiffres de la colonne « Effectifs Niveau Primaire en 2004 » ne sont qu'une estimation. Les colonnes p90 et p10 correspondent aux percentiles 90 et 10 des distributions par type d'établissements des montants de frais demandés selon les déclarations des parents. (NB : 1€ = 700 Ch.\$). Note (**): dans tous les textes ici présentés Municipal (M) est équivalent à Public.

En général nous voyons que les pourcentages d'effectifs totaux du primaire (1er à 8^{ème} grade) et ceux du 8^{ème} grade ayant passé les tests SIMCE en 2004 se correspondent entre eux. Une légère différence s'observe au niveau des effectifs des établissements publics académiquement sélectifs. Cela tient au fait qu'une partie importante de ces établissements commence leur offre d'enseignement seulement à partir du 7^{ème} grade. Leur caractère académiquement sélectif semble lié aussi à une prestigieuse réputation, ce que paraît aussi se refléter par le relativement faible nombre d'établissements (3,5%) en comparaison aux effectifs accueillis (6%).¹⁴

¹⁴Comme nous l'avons noté au chapitre n°5, ce type d'établissement — à 91% urbain — reste très considéré dans le débat public, d'abord parce qu'il s'agit pour la plupart des lycées fondés il y a de nombreuses décennies, presque deux cent ans pour certains. Ils se réclament laïques et pluriels et sont vus comme la fierté du secteur public et comme l'espoir pour ceux que la pensée commune indique comme les « pauvres intelligents ».

Du côté des frais de scolarisation, nous apprécions une différence presque cinq fois plus importante entre les catégories privé payant et privé subventionné (165 € contre 27 € en moyenne et 97 € contre 18 €), différence modulée aussi par le fait que le caractère académiquement sélectif serait associé à des frais plus importants dans chacune de celles-ci. Notons quand-même que les établissements privés subventionnés ont tendance à couvrir une diversité plus importante que les privés payants vis-à-vis des frais demandés. Ces derniers montrent des frais plus élevés et plus concentrés à la fois. Cette situation peut s'apprécier à l'aide des coefficients de variation (ratio écart-type/moyenne) pour chaque catégorie : 0,78 et 0,66 pour les premières contre 0,56 et 0,36 pour les secondes. Toutefois, notons que certains établissements privés subventionnés sélectifs demandent des frais supérieurs à certains privés payants non-sélectifs. Si cette dernière catégorie qui, à notre connaissance n'a jamais été explorée, n'est pas très importante en nombre, elle semble remplir une certaine fonction sociale au sein du système, comme nous le verrons à partir des proportions de redoublants que chaque type d'école accueillait en 2004.

Notons enfin que les élèves placés dans des établissements gratuits représentent 64% de la population de 8^{ème} grade, dont 85% est inscrit dans un établissement public. Pour leur part, les établissements académiquement sélectifs, toutes catégories confondues, accueillent 35% des élèves répartis comme suit : 18,8% pour le public, 18,2% pour le privé payant et 63% pour le privé subventionné. A l'intérieur de chaque sous-groupe, le privé payant est à 89% sélectif, contre 57,6% pour le privé subventionné et 12,1% pour le public. Ce déséquilibre assez significatif, suivi aussi par la proportion d'établissements pour chacune des sous-catégories (respectivement, 81%, 45% et 6,4%), conforte la littérature existante et montre qu'en général le privé s'oriente à accueillir un public trié d'un point de vue académique. En même temps, ce secteur semble augmenter les frais demandés aux familles de manière concomitante à cette caractéristique (cf. tableau B13) sachant qu'il contemple aussi un tri social (religieux) et économique (cf. tableau B12, lignes B à E). Toujours est-il que, pour nos analyses, la sélection académique restera la caractéristique la plus importante de fait que nous la suspectons comme l'expression synthétique d'un regroupement cumulatif de facteurs socio-scolaires.

Une vision générale des caractéristiques des établissements selon la typologie proposée

Les différences brutes de scores aux tests par type d'établissement

Le graphique B6 montre les scores IRT des quatre tests passés en 2004. Nous pouvons y observer que, comme attendu, notre typologie d'établissements semble bien corrélée avec les acquis des élèves manifestés aux tests. Les différences brutes qui montrent les « boîtes à moustache » respectives sont toutes significatives, sauf entre les catégories gratuites non académiquement sélectives (i.e. le privé

subventionné non-sélectif sans FP et public non-sélectif).¹⁵ Cela est valable pour toutes les matières sauf en mathématiques, où le sous-secteur public ($m=238,4$; $\acute{e}-t=44,4$) dépasse légèrement son homologue du privé ($m=237$; $\acute{e}-t=45$). En même temps, le seul type d'écoles qui se détache des moyennes nationales respectives correspond aux établissements qui appliquent la sélection académique. Cependant et malgré ces différences de moyennes, il faut noter que la plage de scores couverte pour chaque type d'école, tend à être semblable pour la majorité des différentes catégories. Ce constat est consistant avec ceux d'autres auteurs (cf. Ramírez, 2007 ; chapitre n°6).

Notons toutefois que le privé payant se montre comme l'exception à cette tendance majoritaire. En effet, nous y trouvons une homogénéité plus marquée des résultats et à la fois, des scores sensiblement plus élevés pour toutes les matières. De fait, les établissements privés payants dépassent à la catégorie suivante (privé subventionné sélectif et avec FP), en moyenne, entre 20 points (Scs. Sociales) et 33 points (Maths), ce qui équivaut, respectivement, à entre 40% et 66% d'écart-type. Plus significatif encore, ce dépassement des résultats moyens aux tests de la part du privé payant est supérieur ou proche d'un écart-type quand on le compare avec les deux catégories qui ne sont pas académiquement sélectives et qui correspondent à des établissements gratuits. Pour sa part, le public sélectif tend à être plus divers que toutes les autres catégories avec des écarts-type par matière toujours supérieures au respectif écart-type national.

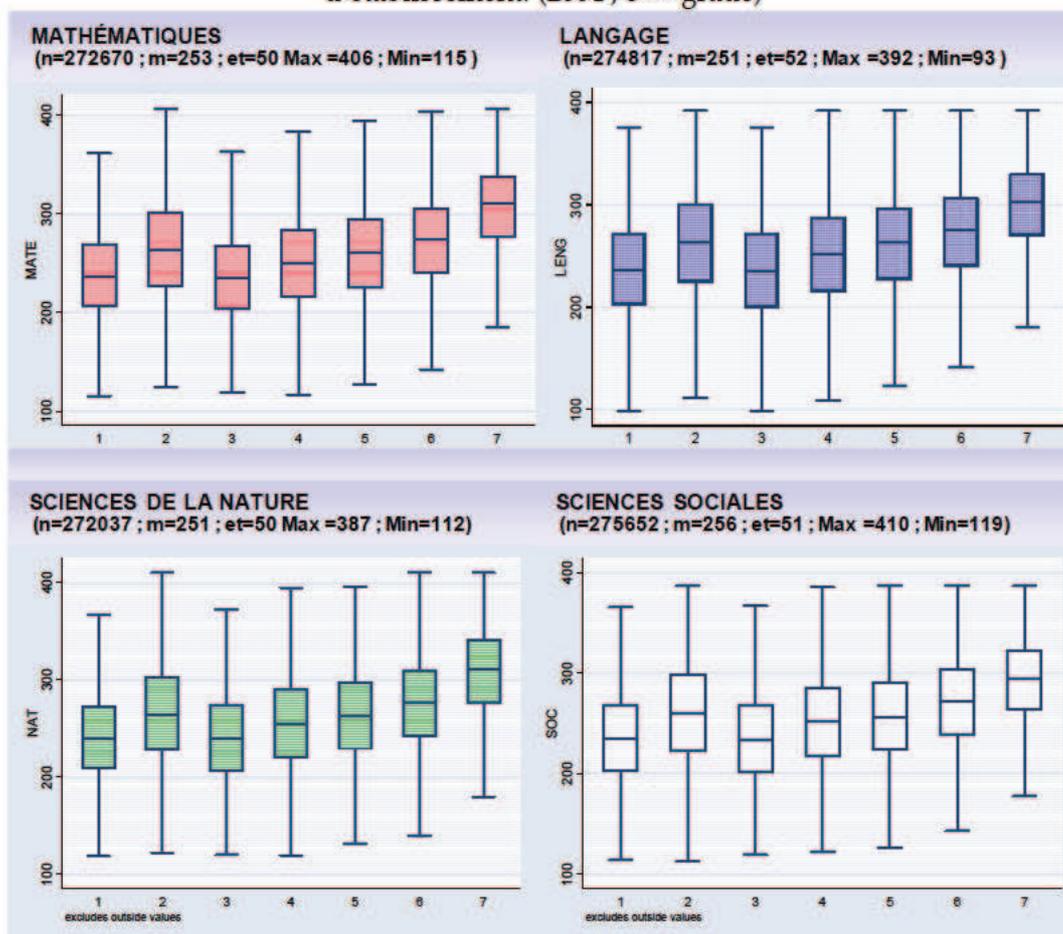
Enfin, notons que les coefficients R^2 résumant les différences de moyennes selon la catégorisation des écoles sont, respectivement, de 0,16, 0,13, 0,14 et 0,12 en maths, langage, sciences de la nature et sciences sociales.¹⁶ Ceux-ci sont très légèrement plus hauts ($\approx +0,002$) quand on considère la division à huit types d'établissement plutôt qu'à sept.

¹⁵ Notons que si l'on fait une analyse avec nos huit catégories de départ, les écoles non sélectives du privé payant (≈ 2200 élèves) sont derrière, en moyenne, de leurs homologues sélectives. Pourtant, la première catégorie reste supérieure à toutes les autres, sauf en langage, où elle obtient, en termes bruts et en moyenne, des scores semblables à ceux du privé subventionné sélectif avec FP.

¹⁶ Ces coefficients R^2 sont issus d'une analyse de la variance via la procédure GLM de SAS®. Ils résument parfois des comparaisons simultanées entre moyennes faites à l'aide du test Ryan-Einot-Gabriel-Welsch (REGW) et/ou par le test Tukey. Le premier test représente un compromis entre l'erreur de première et de seconde espèce. Toutefois, cette procédure a été utilisée pour le cas des tests de type illustratif et exploratoire des différences entre catégories, autant sur des variables continues qu'éventuellement ordinales. Le test Tukey, en revanche, peut être mieux adapté quand le déséquilibre d'effectifs est assez important entre catégories. En général nous nous sommes servis des deux à la fois.

Graphique B6

Distributions des scores IRT individuels au niveau pays, par matière et selon type d'établissement. (2004 ; 8^{ème} grade)

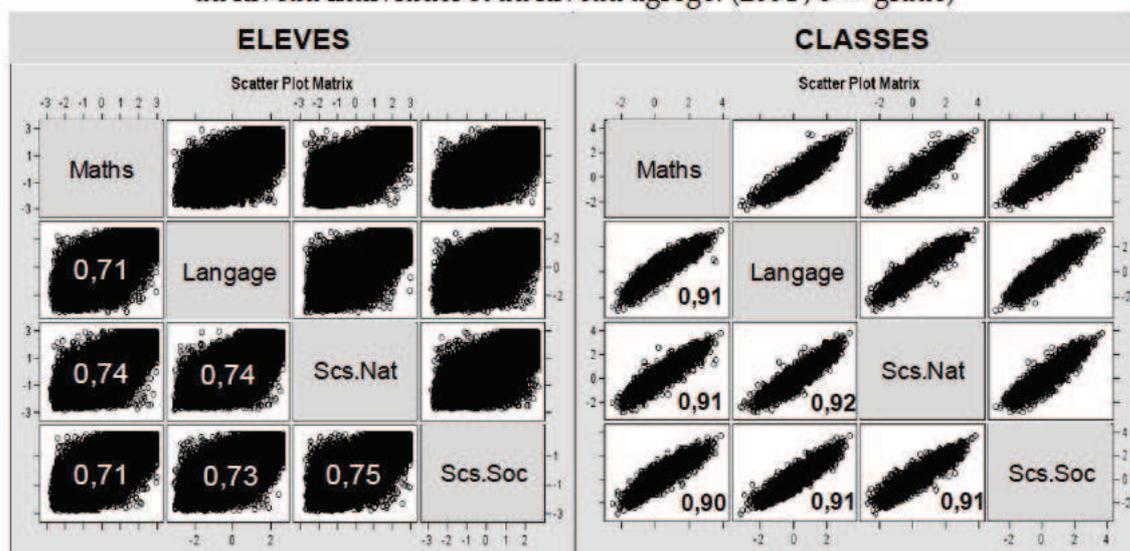


Source : élaboration de l'auteur. **Lecture :** chaque « boîte aux moustaches » représente la distribution des scores de 98% d'élèves sur l'échelle construite pour communiquer ces résultats de manière officielle (i.e. fixant la moyenne nationale de l'année de départ à 250 points). La boîte montre l'écart interquartile ainsi que 50% des individus autour de la médiane (ligne horizontale). Note : les « n » correspondent à 100% des scores considérés (voir tableau A1). Chaque légende montre la moyenne, l'écart-type et les valeurs extrêmes du total de la distribution nationale. **Type d'établissement :** (1) PUBLIC (Municipal) Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

La corrélation entre scores aux tests peut être aussi appréciée de manière indirecte sur le graphique B6. Mais une vision plus précise de ces relations brutes, tant au niveau élèves que entre classes est présentée sur le graphique B7. Consignons que ces corrélations sont un peu moins fortes quand on les regarde au sein de chaque catégorie d'établissement à l'exception du sous-secteur public sélectif.

Graphique B7

Corrélations de Pearson entre les scores aux tests
au niveau individuel et au niveau agrégé. (2004 ; 8^{ème} grade)



Source : élaboration de l'auteur. Lecture : au niveau des élèves, entre les scores en mathématiques et ceux en sciences sociales le coefficient r de Pearson est égal à 0,71. Les élèves correspondent à entre 275652 et 272037 et les classes sont entre 9410 et 9412. Pour chaque niveau et matière, les scores ont été standardisés avec une moyenne égale à 0 et un écart-type égal à 1.

L'emplacement géographique de l'offre scolaire

Tableau B15

Distribution des établissements selon ruralité. (8^{ème} grade - 2004)

Catégorie d'Établissement	TOTAL	Urbains	Ruraux
	n	%	
(1) Public NS	2860	56	44
(2) Public S	194	91	9
(3) Privé Subv. NS NFP	544	62	38
(4) Privé Subv. NS FP	593	92	8
(5) Privé Subv. S NFP	130	92	8
(6) Privé Subv. S FP	801	99	1
(7) Privé Payant	490	97	3
TOTAL	5612	72	28

Source : élaboration de l'auteur. Lecture : 97% des écoles du privé payant se trouvent en milieu urbain. Ce qui équivaut à 475 établissements.

Le nombre d'années d'étude des parents d'élèves

Nous avons considéré nécessaire, avant de commencer nos analyses, d'illustrer nos estimations à cet égard compte tenu de l'importance et du rôle que nous assignerons par la suite à la scolarité des mères. Ceci à partir du constat que cette variable est largement utilisée dans la littérature ayant recours aux modèles économétriques, mais surtout qu'elle nous permet d'avoir une base comparative avec ce qui a été écrit sur le système scolaire chilien (cf. chapitre n°6). Il y a aussi d'autres raisons assez importantes pour que nous nous centrons sur cette variable, sans négliger, comme nous le verrons, d'autres facteurs tenant aux capitaux humain et économique des familles.

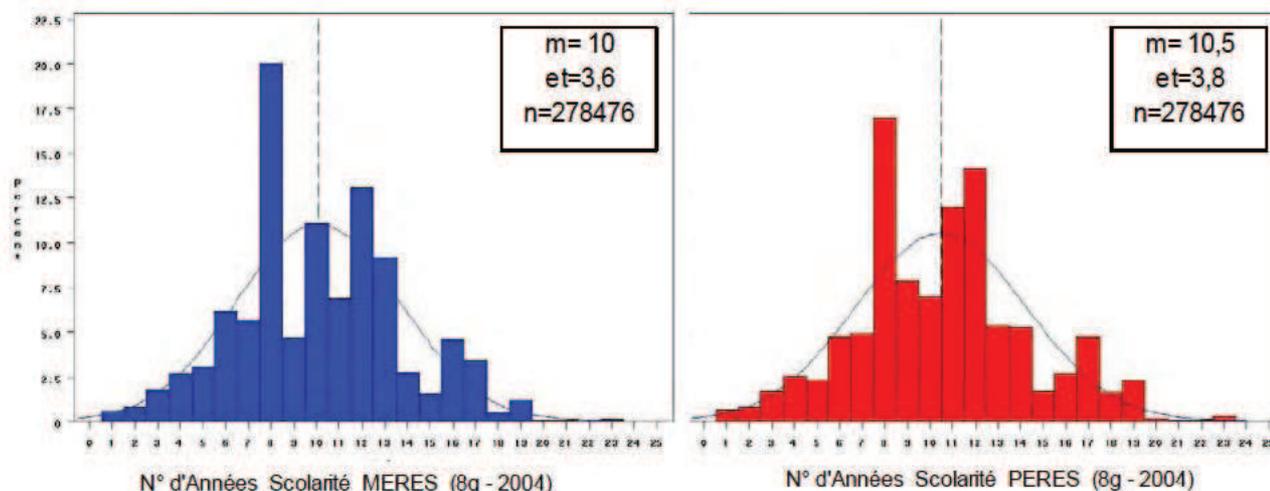
Le graphique B8 illustre les distributions de la scolarité des deux parents mesurées en années d'étude. Les mères ont, en moyenne, presque cinq mois de scolarité de moins que les pères (moins 44% d'année). L'allure de ces variables n'est pas tout à fait normale car une partie importante de cette population a arrêté ses études au 8^{ème} grade (i.e. à la fin du primaire). Cette situation est plus marquée pour les femmes ; elles sont aussi moins présentes dans les niveaux plus hauts d'étude : supérieur (plus de 15 années) et secondaire (entre 9 et 12 années). En effet, les hommes vont plus loin une fois qu'ils sont déjà entrés dans un cycle. Cette donnée démographique ainsi que la différence entre sexes, pourraient, bien entendu, être en train de varier avec l'extension de la scolarisation obligatoire et la demande croissante pour plus de formation (cf. chapitre n°6). Néanmoins, la population de parents d'élèves qui nous intéresse ici ne semble pas encore touchée par ce possible changement.¹⁷

Un autre constat important est la forte corrélation entre la scolarité des mères et celle des pères par élève (r de Pearson=0,70 et Alpha de Cronbach = 0,83). De plus, ces deux variables sont fortement reliées, autant au niveau de revenu déclaré par les familles qu'aux biens présents dans le foyer. Par conséquent, nous avons choisi, d'une part, de construire un Indice composite du statut socioéconomique des familles ou Indice NSE (cf. voir plus loin), puis de l'exprimer au niveau des individus-élèves en termes de catégories du type décile et quintiles (cf. voir plus loin). D'autre part, nous allons considérer souvent seulement l'indicateur fourni par la scolarité des mères. D'abord, parce qu'il est, en partie, un proxy d'autres facteurs. Autrement dit, si l'on perd de l'information par rapport à des indices composites ou à l'utilisation simultanée de variables (e.g. scolarité du père et de la mère), cette perte n'est pas très importante. Puis, parce que la littérature accorde de l'importance à cette variable en elle-même (cf. Duru-Bellat et Van Zanten, 2006).

¹⁷Pour d'autres pays, notamment développés, la majeure scolarisation des filles et l'allongement du parcours scolaire pour tous sont des tendances avérées (cf. Duru-Bellat et Van Zanten, 2006).

Graphique B8

Allure des courbes de distribution du nombre d'années de scolarité des parents (déclarées) des élèves du 8^{ème} grade en 2004.



Source : élaboration de l'auteur. Lecture : En moyenne les mères déclarent avoir 10 années d'étude pour un range qui varie entre 1 et 23 années de scolarisation, et pour une dispersion légèrement inférieure à 4 ans. Comme référence la courbe normale est notée.

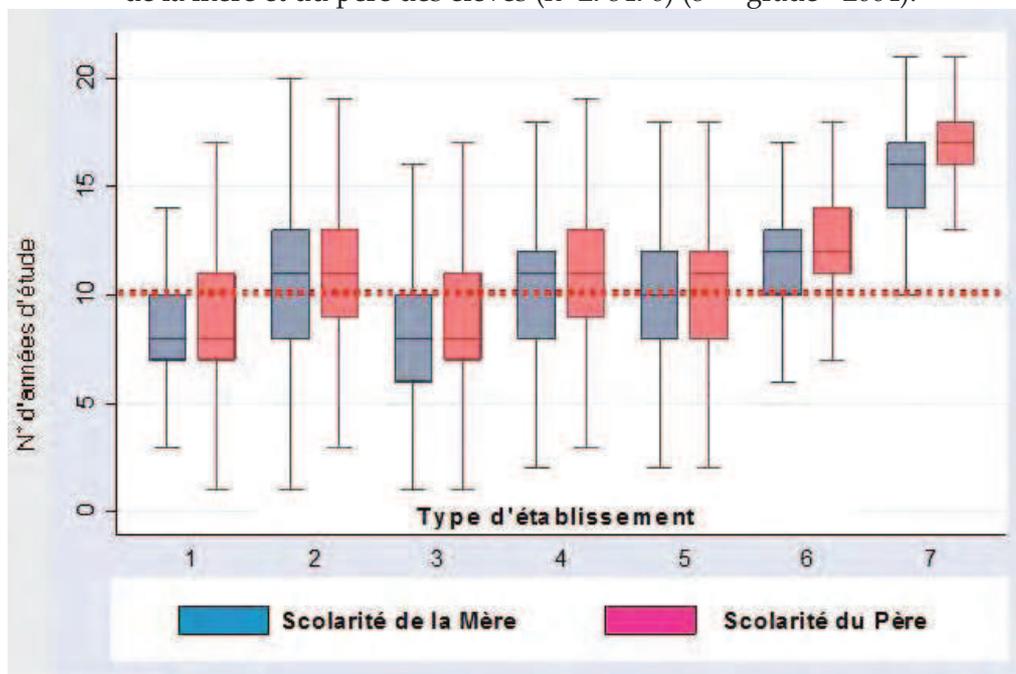
De fait, disons en termes très pratiques que si tout le monde n'a pas forcément un père, tous ont eu une mère. Ainsi, nous voudrions reconnaître l'importance de l'impact potentiel sur la réussite scolaire et les apprentissages associés au niveau culturel et éducatif des mères, par exemple, pendant la grossesse et les premières années de vie des enfants. Par ailleurs, notons que selon les chiffres de la CEPAL,¹⁸ 27,1% des foyers urbains du Chili avaient à leur tête une femme (dont 6,1% parmi les foyers biparentaux). Ce chiffre atteindrait 87,9% pour les foyers monoparentaux et 55% pour les foyers unipersonnels. Autrement dit, il existe une réalité touchant près d'un quart des foyers où les mères prennent en charge toutes seules leurs enfants.¹⁹ Ensuite, le graphique B9 illustre la distribution des scolarités des deux parents selon notre typologie d'établissements. D'une part, on conforte la forte corrélation déjà notée entre ces variables. Ceci illustre une tendance à l'homophilie sociale par niveau d'étude, de même qu'une distribution de la scolarisation des enfants des familles très en ligne avec celle-ci. Certes, il existe une certaine dispersion commune d'un point de vue des plages recouvertes par chaque catégorie d'établissement, notamment parmi les cinq premières. Toutefois, les catégories académiquement sélectives et/ou ayant recours au FP (n°2, n°4 et n°5 sur le graphique B9) tendent nettement à se détacher vers le haut graduellement, et cela pour le cas des deux parents.

¹⁸ Voir : <http://websie.eclac.cl/infest/ajax/cepalstat.asp?carpeta=estadisticas>

¹⁹ D'autres estimations indiquent, qu'en 2002, un total de 33% des foyers avait en tête une femme contre 27% en 1992. Cette réalité serait davantage prégnante chez les secteurs défavorisés urbains. Toujours est-il que le foyer dit « nucléaire » — mère, père et enfants — reste fortement majoritaire au niveau pays (57% en 2002) de même que pour notre enquête des parents des élèves de 8^{ème} grade en 2004 (71%).

Graphique B9

Distribution par type d'établissement du nombre d'années de scolarité de la mère et du père des élèves (n=278476) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : La scolarité moyenne des Mères (10,06 années) est illustrée par la ligne rouge à traits. La ligne horizontale de chaque boîte montre la médiane et chacune des boîtes aux moustaches illustre 98% des distributions en excluant les données extrêmes des deux côtés (1% chacun).

Autrement dit, elles concentrent des parents sensiblement plus éduqués en moyenne que les catégories non-sélectives. Ce détachement est encore plus fort quand il s'agit du privé subventionné avec FP et sélectif (12,3 et 11,9 années pour les pères et les mères en moyenne, plus une dispersion autour de 3 dans les deux cas), de même que pour le privé payant. En effet, ce dernier type d'établissement montre des populations où 98%, voire presque 100% des pères et des mères se trouvent dans la moyenne du pays (16,6 et 15,6 années d'étude en moyenne et une dispersion inférieure à 2,5 années). Clairement, les parents qui ont fait des études supérieures se concentrent dans ce type d'école. Notons aussi que pour ces deux catégories privées (n°6 et n°7) les pères exhibent sensiblement plus d'années d'étude que les mères. La même situation s'avère vraie pour le privé subventionné non-sélectif avec FP. Il semble donc évident que le niveau de revenu, ostensiblement lié au niveau d'étude (voir annexe sur l'Indice NSE), serait davantage lié aux pères de famille.

En ce qui concerne les deux catégories gratuites et non sélectives, que ce soit dans le privé subventionné ou dans le public, il n'y a pas des différences dans le niveau de scolarité des deux parents (8,5 années en moyenne dans les deux cas pour les mères, et 8,8 pour les pères). Celles-ci se placent alors en dessous des moyennes nationales pour ces deux variables.

Le capital culturel des familles

Le capital culturel, pris à l'état *matériel* (Bourdieu, 1979) est souvent mesuré d'après le nombre de livres disponibles à la maison. Ici, nous avons considéré les six catégories dont les parents ont été invités à se classer. Nous les avons ensuite regroupé en trois grandes catégories, chacune proche du tiers de la population. Les six niveaux originaux correspondent, sur l'ensemble de répondants, à : « 0 » livre à la maison (3,0%) ; « 1 à 10 » (28,2%) ; « 11 à 50 » (39,6%) ; « 51 à 100 » (17,1%) ; « 101 à 200 » (6,1%) ; « plus de 200 » (5,9%).²⁰ Les trois catégories résultantes considèrent alors 31,3% (n=87053) en situation de bas capital culturel, 39,6% (n=110396) avec un capital culturel moyen et 29,1% (n=81027) jouissant d'un capital élevé. Notons que si nous allons souvent considérer cette manière d'opérationnaliser la variable, pour certaines analyses, nous reviendrons toujours au découpage en six catégories.

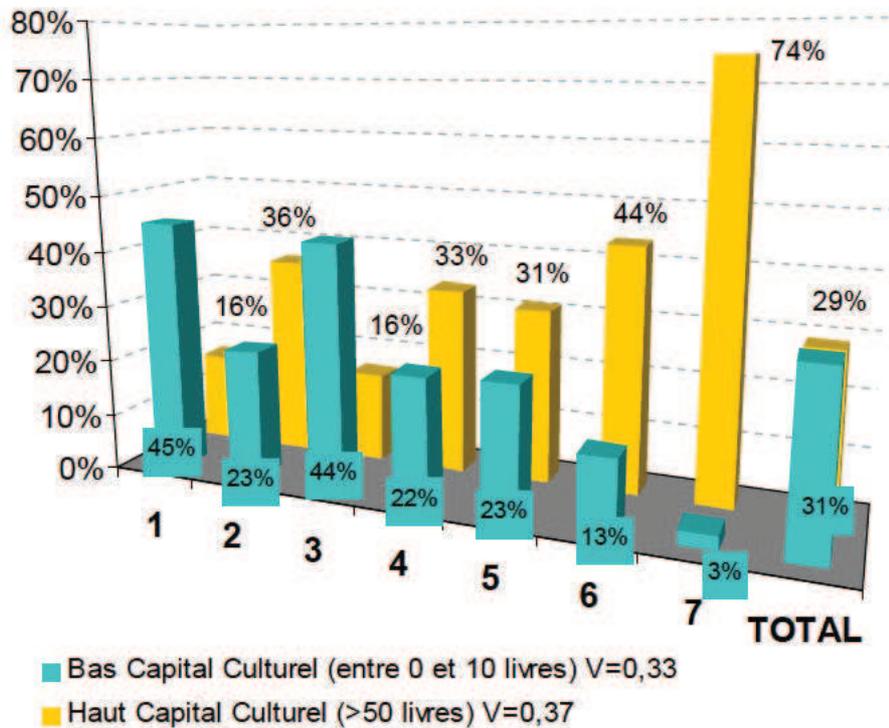
Notons d'ailleurs que le capital culturel est assez lié au nombre d'années de scolarisation des deux parents et au niveau de revenu des familles (r de Spearman entre 0,45 et 0,52). Pourtant, le nombre de livres dont les élèves disposent chez eux reste une variable en elle-même et qui, de plus, n'est pas totalement déterminée par ces deux autres facteurs, ce qui est en phase avec la littérature (cf. Duru-Bellat et Van Zanten, 2006). La distribution des élèves selon niveau de capital culturel d'après notre catégorisation d'établissements est présentée dans le graphique B10.

Comme il était possible de s'y attendre, les différences entre types d'établissement se font sentir ; néanmoins elles sont plus fortes chez les élèves ayant un haut capital culturel (i.e. >50 livres à la maison). Autrement dit, ces élèves se ségréguent un peu plus (V de Cramer = 0,37) par rapport aux élèves en situation de bas capital culturel (V=0,33). Notons aussi, encore une fois, la similitude entre les établissements gratuits, soient-ils publics ou privés, comptant avec près de 45% en niveau bas, 16% en niveau haut et donc, quatre élèves sur dix en niveau moyen.

²⁰ Cette variable comptait un 7,2% de données manquantes. Les proportions originales étaient alors de 3,2%, 30,1%, 37,9%, 15,9%, 6,5% et 6,3% respectivement. Nous avons décidé de remplir ces données manquantes à partir d'un modèle d'analyse de la variance (Proc GLM de SAS®) faisant appel à quatre variables : régions, condition urbaine, sexe du chef de ménage et revenu familial (R² du modèle = 27,8%). Ainsi, on a imputé des catégories aux données manquantes.

Graphique B10

Proportions d'élèves en situation de Bas et de Haut Capital Culturel à la maison.
Selon type d'établissement. (n=278476 ; 8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture : 45% des élèves chez le public non-sélectif (noté n°1) appartient au niveau Bas Capital Culturel tandis que 16% appartient au niveau le plus élevé. Les coefficients V de Cramer (V) illustrent entre 0 et 1 la force des déséquilibres pour chaque total.

Le redoublement

Quand on fait des analyses sur des coupures transversales, il est fort difficile d'objectiver l'impact net du redoublement sur les acquis et les progressions scolaires. Toutefois, cette variable (i.e. le fait qu'un élève ait redoublé antérieurement) peut être aussi considérée comme un indicateur des compétences scolaires individuelles (e.g. Belleï, 2009). L'impact de cette mesure scolaire (i.e. le fait d'avoir fait redoubler un élève) a été largement étudié en milieu français (e.g. Crahay, 2003²¹; Paul et

²¹Crahay, M.; 2003 ; Peut-on lutter contre l'échec scolaire ? ; De Boeck, Bruxelles ; 378 pp.

Troncin, 2004²²; Bernard et al., 2005²³). Les conclusions des études sont plutôt univoque : on n'améliore pas les apprentissages sur le long terme des élèves ayant été objet de cette décision à différents moments de leur parcours. Ainsi, l'usage généralisé de cette mesure, comme une manière de résoudre les difficultés scolaires de certains élèves, reste très discutable et plutôt déconseillé,²⁴ à l'exception de cas assez situés. Cela serait d'autant plus important que d'autres effets sont à atteindre sur les rapports à l'école et les perceptions de soi entretenus par les élèves redoublants. Il faut, toutefois, considérer l'âge, le niveau scolaire atteint et d'autres situations susceptibles de moduler ces potentiels impacts. Au Chili, Belleï (2009, cf. chapitre n°6) a montré une relation significative entre bas résultats scolaires et le fait d'être déjà redoublant au 4^{ème} grade. La littérature montre que cette variable peut être socialement très typée (e.g. Duru-Bellat, 2003a).

En définitive, la variable « redoublement » est riche de différentes significations. Elle montre un certain actif scolaire cumulé dans le passé ; elle peut contenir des approches plus ou moins favorables au travail scolaire, aux disciplines et à l'école en général ; elle peut fonctionner comme un proxy des performances académiques individuelles, mais aussi, de l'âge scolaire des élèves. Hélas, nous ne pouvons pas ici appréhender cette variable dans sa version longitudinale, mais plutôt comme une réalité déjà présente au moment des tests. Malheureusement, nous n'avons pas rencontré des études faites au Chili décrivant plus soigneusement le phénomène. Cela nous a obligé, en partie, à destiner quelques efforts à le comprendre davantage. Ainsi, et ne pouvant pas jouir de plus de précisions nous allons utiliser la condition de redoublement des élèves comme un contrôle indispensable des résultats en termes d'acquis scolaires, comme un facteur éventuellement modulateur d'autres variables, et enfin, comme une typologie spécifique d'élève. En termes concrets, disons qu'autant les parents que les propres élèves ont été invités à indiquer combien des fois l'élève avait redoublé une année avant le 8^{ème} grade en 2004. Cela nous a permis d'identifier quatre sous-catégories (i.e. jamais, une fois, deux fois, trois fois ou plus). Néanmoins, nous avons décidé de saisir cette variable d'une manière binaire (a redoublé ou pas).

Compte tenu des limitations des données, nous ne pouvons pas qualifier avec exactitude l'ampleur de cette pratique en tant que telle de la part des établissements. En revanche, nous pouvons objectiver les proportions de redoublants accueillis par chaque établissement. A la lumière de notre catégorisation, cela devrait nous indiquer en quelque sorte la prévalence de celle-ci et /ou le destin scolaire, en termes

²²Paul J.J. ; Troncin T. ; 2004 ; *Les apports de la recherche sur l'impact du redoublement comme moyen de traiter les difficultés scolaires au cours de la scolarité obligatoire* ; Rapport du Haut conseil de l'évaluation de l'école (HCEE), n°14 ; 44 pp.

²³Bernard, J-M. ; Simon, O. ; Vianou, K. ; 2005 ; *Le redoublement : mirage de l'école africaine?* ; Programme d'Analyse des Systèmes Educatifs de la CONFEMEN ; 96 pp.

²⁴ Comme le notent Bernard et al. (2005) : « Le véritable enjeu du redoublement réside plutôt dans les apprentissages, c'est-à-dire que le coup de pouce souhaité vise à permettre aux élèves en situation d'échec scolaire de suivre les cours de façon normale. La question de fond qui se pose alors est la suivante : le fait qu'un élève améliore sa position relative dans la classe jusqu'à réussir comme la moyenne des élèves veut-il nécessairement dire que ses difficultés d'apprentissage sont résolues ? » (p. 56).

d'établissement d'arrivée, des élèves ayant fait l'objet de cette mesure de la part d'autres établissements qui expulseraient leurs élèves les moins performants (cf. chapitre n°6).

Le graphique B11 nous montre, d'un côté, les proportions de redoublants accueillis par les différents établissements selon leur catégorie, et d'un autre côté, le nombre d'élèves redoublants et les proportions qu'ils représentent pour chaque catégorie et au niveau national. Si les établissements publics non-sélectifs accueillent 48% des élèves du 8^{ème} grade, ils concentrent en même temps, 62% des redoublants à ce même niveau. Il est assez probable que cela soit l'effet cumulé et conjugué des pratiques de sélection académique et socioéconomique présentes dans le système d'éducation au Chili, mais aussi de la pratique du redoublement par elle-même. En effet, certains établissements, notamment privés se permettent de ne pas accueillir et/ou de se défaire des élèves avec plus de difficultés et moins des capitaux familiaux et par ce fait, plus chers à éduquer (cf. chapitre n°6). Une explication plausible serait, donc, que la pratique du redoublement est nettement moins fréquente parmi les sous-catégories sélectives. Hélas, nous n'avons pas ici les moyens pour départager cette question. Toutefois nous nous sommes inclinés à penser, conformément à la littérature, qu'une large partie de ces 80% de redoublants placés chez les écoles non-sélectives est l'expression du tri que les établissements sélectifs mettent en place dans tout le système, soit, en éliminant de leurs inscrits les élèves les plus faibles ; soit, en ne les acceptant pas dès l'entrée.

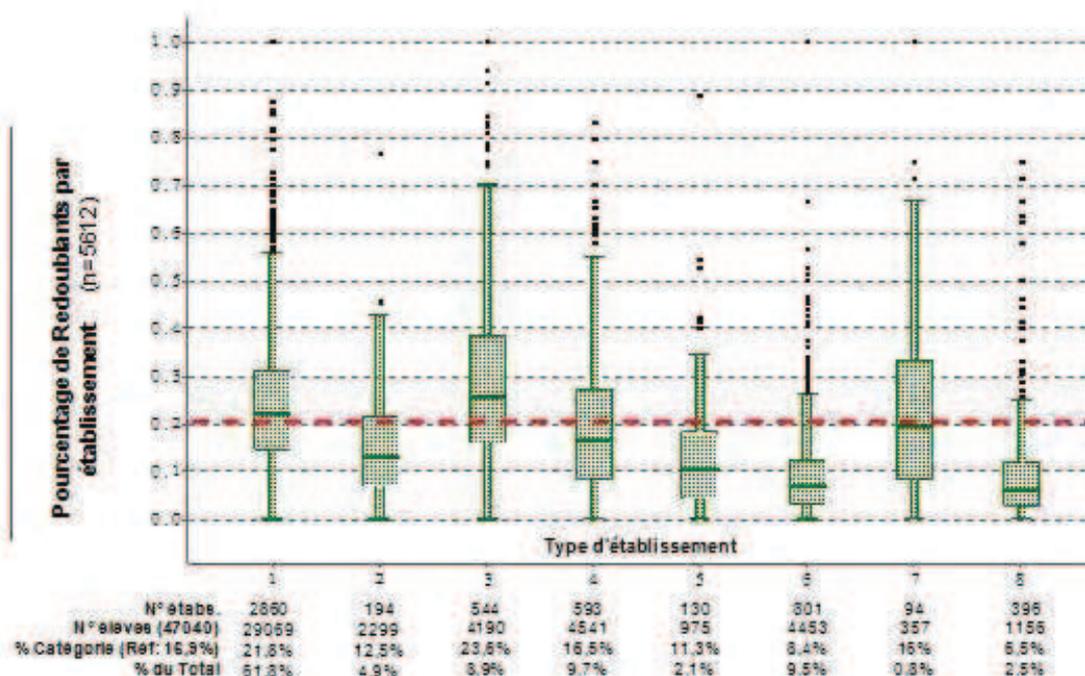
Nonobstant, la pratique du redoublement pourrait faire aussi partie des dispositifs de gestion académique dans des établissements que l'on sait, accueillent un public majoritairement défavorisé (catégories n°1 ou n°3 sur le graphique B11). Ceci serait une manière de faire face aux problèmes rencontrés vis-à-vis du travail, de l'ordre et de la convivialité scolaire. Le mécanisme en œuvre serait donc, « la menace du redoublement », permettant, entre autres, l'expression des arrangements des acteurs, notamment de la part des enseignants, vis-à-vis de ce public.

Nous n'allons pas commenter ni reproduire en détail les résultats illustrés par le graphique B11. Retenons toutefois que cette variable, *a priori*, purement scolaire, ne se repartit pas de manière équilibrée parmi les types d'école, même s'il y a une certaine variabilité dans toutes les catégories.²⁵ Rappelons aussi que le pourcentage général de redoublants correspond à 16,9% (8^{ème} grade, 2004) et que la proportion moyenne par établissement est de 20,5%. Ceci permet d'apprécier, à simple vue, les écarts des catégories à la situation d'indépendance.

²⁵ Notons au passage, le cas des 94 établissements du privé payant non-sélectif. En effet, les politiques de ces écoles vis-à-vis du redoublement et de l'accueil des redoublants ne semblent pas symétriques à celles de leurs homologues du privé payant sélectif. Au contraire, le privé payant non-sélectif semble voué à accueillir des élèves moins performants académiquement et/ou avec un passé scolaire difficile, mais, issus des familles capables de payer d'importantes sommes.

Graphique B11

Proportions de redoublants inscrits en 2004 au 8^{ème} grade par établissement.
Selon catégorie d'établissements. (n = 5612)



Source : élaboration de l'auteur. **Lecture :** Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT Non-Sélectif ; (8) PRIVÉ PAYANT Sélectif. Note : la ligne « N°d'élèves » correspond au nombre total de redoublants en 2004 (n=47.040), puis, par catégorie. En haut de celle-ci, on trouve le nombre d'établissement par catégorie.

Lecture en haut : La moyenne nationale du pourcentage de redoublants par établissement (20,5%) est indiquée par la ligne rouge à traits. Cela signifie que, en moyenne, les établissements du pays ont 20,5% d'élèves dans cette condition. Chaque boîte aux moustaches montre la médiane (ligne verte) et le 50% autour de celle-ci (i.e. la boîte). Par exemple, la proportion médiane de redoublants chez les établissements du privé subventionné sélectif sans FP (n°5) est de 10% avec 50% d'établissements (boîte) concentrant entre 4% et 18% de redoublants parmi leurs effectifs. En revanche, les mêmes chiffres dans le public non-sélectif équivalent à 23% comme médiane, avec 50% de ces établissements accueillant entre 15% et 31% d'élèves redoublants, pour cette année et pour ce grade.

Lecture en bas : nous trouvons le nombre d'élèves redoublants accueillis par chaque catégorie en total, ainsi que le pourcentage de redoublants présents dans la catégorie et la proportion du total de redoublants y placés. La première ligne des pourcentages montre les proportions relatives (i.e. quel pourcentage du total d'élèves de chaque catégorie a redoublé au moins une fois avant 2004 ?). La seconde ligne, illustre les proportions par rapport au pays (i.e. du total national d'élèves dans cette situation, quelle proportion est accueillie par chaque catégorie ?). Par exemple : 62% des redoublants du pays se concentre dans les établissements publics non-sélectifs, cela correspond à 22% du total d'élèves dans cette catégorie d'établissements (29.069 sur 133.417). Enfin, rappelons que la proportion nationale de redoublants, estimée parmi les inscrits en 8^{ème} grade en 2004, correspond à 16,89% (47.040 sur 278.476).

Notons, de plus, que le phénomène de concentration des redoublants est plus fort pour les établissements ruraux, toutes catégories confondues : 26% comme proportion moyenne contre 18,4% en zone urbaine. Enfin et globalement, remarquons, une fois de plus, que sur les 47040 redoublants faisant leur 8^{ème} grade en 2004, il existe une grosse concentration des redoublants dans le public non-sélectif, et en moindre mesure dans le privé subventionné non-sélectif (avec et sans FP). Ces trois catégories cumulent 80,4% du total de redoublants du pays à ce moment, ce qui tend à confirmer la pertinence de notre catégorisation.

Des perspectives complémentaires : le genre et le redoublement

Dans le tableau B16, nous pouvons apprécier que les redoublants ne se répartissent pas également selon le genre des élèves. En effet, si le total d'élèves ici considérés (i.e. tous ceux pour lesquels on a pu identifier le genre) montre une moyenne de 16,5% de redoublants, cette proportion atteint 19,5% dans le cas des garçons, et seulement 13,5% pour les filles. Ces dernières semblent donc, être moins l'objet de cette décision scolaire et/ou moins déterminées par ses causes. Notons aussi qu'une augmentation de la présence des garçons se dessine dans les modalités accrues du redoublement, constituant 64% de la population ayant redoublé deux fois avant 2004 et 68% de ceux qui ont redoublé trois fois ou plus.

Tableau B16

Table de contingence entre redoublement et genre (8^{ème} grade - 2004)

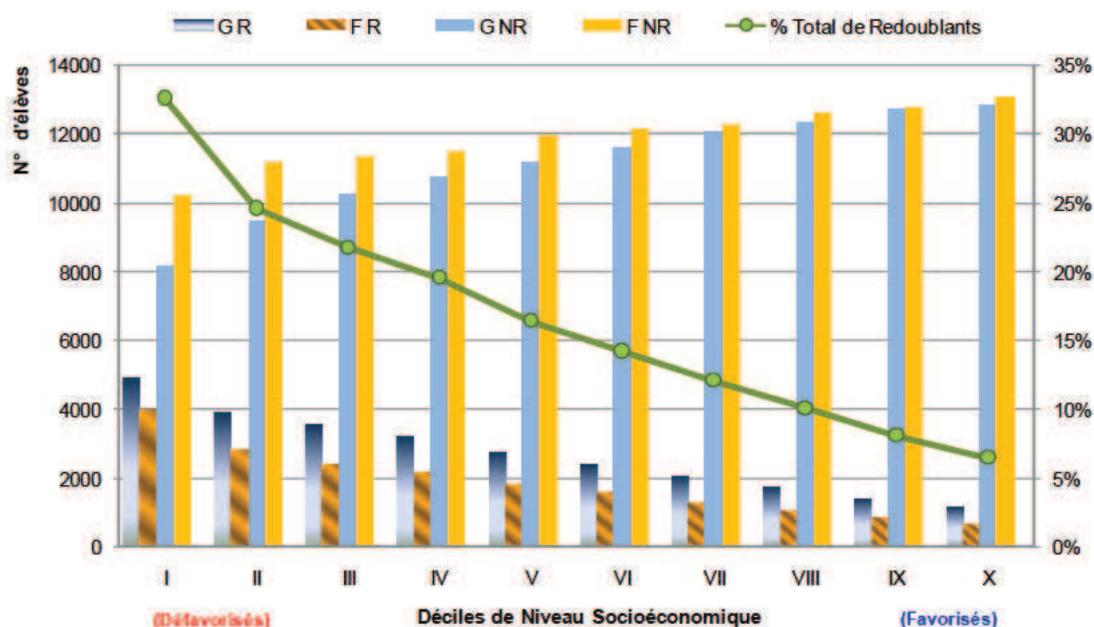
Situation de Redoublement		GARÇONS	FILLES	TOTAL
Non Redoublants	<i>n</i>	111397	118952	230349
	% modalité	48%	52%	100%
	% par Genre	80%	87%	83%
Redoublants UNE fois	<i>n</i>	19855	14533	34388
	% modalité	58% (*)	42%	100%
	% par Genre	14%	11%	12%
Redoublants DEUX fois	<i>n</i>	6059	3481	9540
	% modalité	64%	36%	100%
	% par Genre	4%	3%	3%
Redoublants TROIS fois ou plus	<i>n</i>	1121	538	1659
	% modalité	68%	32%	100%
	% par Genre	0 80%	0 40%	0 60%
TOTAL		138432	137504	275936
<i>V de Cramer = 0.0845 (à p<0,000)</i>		50.2%	49.8%	100%
		100%	100%	

Source : élaboration de l'auteur. Lecture (*) : 58% des 34 388 élèves ayant redoublé une fois avant leur 8^{ème} grade (2004) sont des garçons (n=19.855). Ils correspondent en même temps à 14% du total des garçons considérés.

En outre, nous nous sommes demandés si le redoublement était relié à la condition socioéconomique et si, cette dernière variable modulait ou pas les différences par genre. De fait, on sait que certaines différences scolaires liées au genre des élèves disparaissent ou s'atténuent dans les niveaux socioéconomiques élevés (e.g. Caille et O'Prey, 2005 ; Grisay, 1997). Le graphique B12 illustre cette situation faisant appel à notre indice de niveau socioéconomique ou NSE (cf. voir plus loin). Cette fois ci, on a repris la variable redoublement de manière regroupée ou binaire. On constate à l'œil nu que le redoublement constitue une condition très typée socialement. En effet, parmi le 10% des élèves les plus défavorisés en 2004, on double la moyenne nationale de redoublants (33% contre 16,5%). Et ceci avec une tendance semblable autant pour les filles (28% contre 13,5%) que pour les garçons (38% contre 20%).²⁶

Graphique B12

Nombre de redoublants et de non-redoublants inscrits en 2004 au 8^{ème} grade par genre.
Selon décile de niveau socioéconomique (NSE). (n = 275936)



Source : élaboration de l'auteur. Lecture : Pour le premier décile d'élèves (i.e. les 10% les plus défavorisés en termes de niveau Socioéconomique) on voit à gauche que 4.881 sont des garçons redoublants (GR) et 3.977 sont des filles redoublantes (FR), tandis que 8.124 correspondent à des garçons non-redoublants (GNR) et 10.203 sont des filles non-redoublantes (FNR). A droite on voit que, parmi les 27.185 élèves classés dans ce décile, 32,6% seraient des redoublants.

²⁶Notons au passage que ce premier décile s'écarte, mais très légèrement de la norme générale de parité de genre, totalisant 48% de garçons. Pour leur part, les quatre déciles de plus hauts revenus comptent 51% de garçons.

Le deuxième constat qu'on peut faire en regardant le graphique B12 est que les différences par genre vis-à-vis du redoublement illustrant une moindre présence systématique des filles dans cette condition, reste forte est plutôt constante quel que soit le niveau socioéconomique considéré. Toutefois, de pair que la proportion de redoublants descend à un taux constant au fur et à mesure que le niveau socioéconomique augmente, il faut noter qu'une diminution de la force dans la relation entre le genre et le redoublement se fait aussi évidente, sans pour autant disparaître. Bref, les filles et les garçons redoublent un peu moins inégalement dans les milieux aisés que chez les plus défavorisés. La décroissance dans la force de cette association est capturée par les coefficients V de Cramer qui passent graduellement de 0,11 à 0,07 entre les déciles extrêmes, tout en restant fort significatifs.

L'appui reçu de la part des parents : construction d'une typologie générale

Conforme à l'importance présumable de l'appui des parents perçu par les élèves vis-à-vis du vécu scolaire de ceux-ci, nous nous sommes attardés sur la construction de deux variables susceptibles d'exprimer cette dimension. D'abord, l'extraction d'échelles factorielles permettant de qualifier la quantité et/ou l'intensité de l'appui ressenti. Puis, une classification selon les caractéristiques dominantes dans les déclarations des élèves sur les rapports des parents vis-à-vis de leurs études. Ces échelles et ces caractéristiques sont rapportées globalement à l'aide et aux exigences que les deuxièmes prêtent ou font sur les premiers.

A partir d'une question composée de neuf items, nous avons conduit, d'une part, des analyses factorielles (cf. annexes méthodologiques) pour extraire des dimensions cohérentes sous la forme d'échelles continues. D'autre part, nous avons entrepris une analyse vouée à la construction de clusters ou catégorisation qualitative. La question posée aux élèves était : « Avec quelle fréquence tes parents/tuteurs réalisent les activités suivantes pour t'aider dans ton travail scolaire ? ». Celle-ci se déclinait en neuf affirmations (e.g. « Mes parents m'apprennent comment étudier » ; « Ils exigent de moi des bonnes notes » ; etc.). Les options de réponse étaient sur une échelle Likert à quatre niveaux allant de « toujours ou presque », à « jamais ou presque ». A ces neuf items répondus par les élèves nous avons décidé d'ajouter une réponse donnée par les parents. Celle-ci exprimait la fréquence de révision du carnet des notes de leurs élèves (à 4 niveaux). Cet item a été choisi parmi d'autres possibles (e.g. assiduité aux réunions des parents, entretiens avec les professeurs, etc.) parce qu'il représente, avant tout, une relation parent-élève plus au sein de la famille.

L'Alpha de Cronbach pour l'ensemble d'items pris est de 0,78 et le MSA de 0,86. Notons que l'Alpha serait monté à 0,80 si nous avions éliminé le dixième item issu des déclarations des parents. Néanmoins, nous n'avons pas voulu le faire, compte tenu de l'information extra-élèves (à la manière d'un contrôle sur leurs perceptions déclarées) qu'il apporte, de même que par les propriétés qu'il prêtait aux analyses factorielles (e.g. MSA de l'item=0,89). L'extraction de deux dimensions a été faite sur la méthode ULS (sachant que les résultats des quatre méthodes testées étaient presque identiques). Ensuite, une rotation oblique a été conduite (option Harris-Kaiser). Les résultats donnent deux sous-dimensions corrélées entre elles ($Rho=0,316$) et susceptibles d'être nommées « appui » et « exigence », car la deuxième montre un penchant vers les variables exprimant cette condition perçue. Toutefois, bien que l'AF illustre une structure simple des sous-dimensions, elle est fortement centrée sur le premier axe (« appui »). Donc, c'est celui que nous avons retenu, puisqu'il semble bien résumer l'ensemble. Notons que nous n'avons pas établi des sous-divisions de la population d'élèves à l'aide de cette échelle. Ceci en partie parce qu'à l'opposé des rapports avec, et/ou, au sein de l'école — où nous avons la certitude d'un environnement commun aux élèves qui y appartiennent — les situations

et rapports familiaux semblent plus difficiles à être regardés en termes relatifs et comparatifs sur une échelle générale, du moins, théoriquement.

Conséquemment une construction de clusters a été entreprise en parallèle (avec le logiciel SPAD®, méthode PARTI-DECLA, par l'algorithme de *k-moyennes* ; cf. Annexes méthodologiques). Cette partition a donné quatre profils d'après les dix mêmes items. Disons toutefois que la partition en clusters est associée à un R^2 moyen de 0,37 pour les dix variables-items d'origine. En même temps, ces quatre catégories expliquent 81% de variance ($R^2=0,806$) de la sous-dimension que nous avons choisi d'après l'AF (i.e. « appui »), puis 28% de l'autre axe. Tous les résultats sont illustrés dans les tableaux B17 et B18 et leurs graphiques annexes. Sont aussi indiquées quelques considérations techniques sur la façon de construire les clusters.

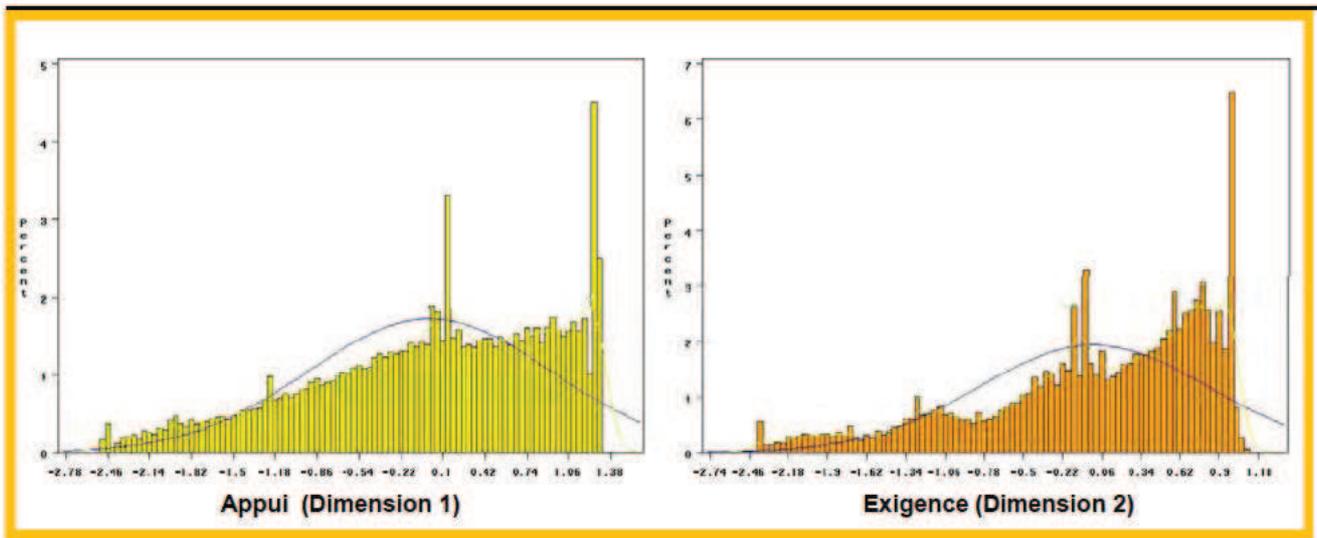
Ainsi, nous allons considérer une classification en quatre types de parents : « souples » (11%), « exigeants » (20%), « aidants » (35%) et « préoccupés » (34%). Bien sûr, cette catégorisation est fondée sur un matériel qui est en fait fortement pollué par les appréciations subjectives des élèves sur eux-mêmes et sur l'attitude de leurs familles envers eux. A notre avis, cette variable est intéressante et indispensable, mais elle doit, quand même, être considérée avec prudence. En effet, il s'agit surtout d'une démarche qui vise à rendre compte des comportements des uns (parents) à partir des appréciations des autres (élèves), sachant que ces derniers font partie directe du phénomène. Cela peut entraîner des fortes distinctions liées à l'interprétation des affirmations proposées, dont les différences d'un élève à l'autre affecteraient, peut-être, le résultat total. Par exemple, il est possible de penser que des élèves faibles à l'école vont sentir que leurs parents les surveillent tout le temps dessus, même si en termes comparatif cela n'est pas le cas. Au contraire, des élèves forts pourraient se sentir plutôt libres de la pression parentale, voire trop libres. Certains contextes pourraient même moduler ces perceptions, par exemple, à partir du niveau d'exigence des écoles. De fait, des effets apparents et contaminés vis-à-vis de la relation entre les familles et la scolarité de leurs enfants sont bien documentés par la littérature (cf. Duru-Bellat et Van Zanten, 2006). En tout état de cause, il s'agit d'un facteur (i.e. l'engagement concret des familles dans la scolarité de leurs enfants et la manière d'y parvenir) qui est, autant difficile à obvier que difficile à être observé et objectivé. Toujours est-il que, n'ayant pas d'autre source disponible dans notre cas, nous allons considérer cette catégorisation, d'une part, comme une source de variation dans les acquis vérifiés aux tests, et d'autre part, davantage comme une représentation du ressenti des élèves, plus que comme une véritable vision des comportements des parents.

Ainsi, et malgré toutes ces limitations, nous pensons que la caractérisation des parents en quatre groupes et/ou via un score d'« appui » scolaire à leurs enfants, nous permettront, du moins, de contrôler d'autres analyses.

Tableau B17

Matrice de Corrélations de Pearson et informations sur l'extraction et rotations factorielles pour la Dimension **Appui des parents**. Histogrammes de deux sous-dimensions extraites.(2004 – 8^{ème} grade)

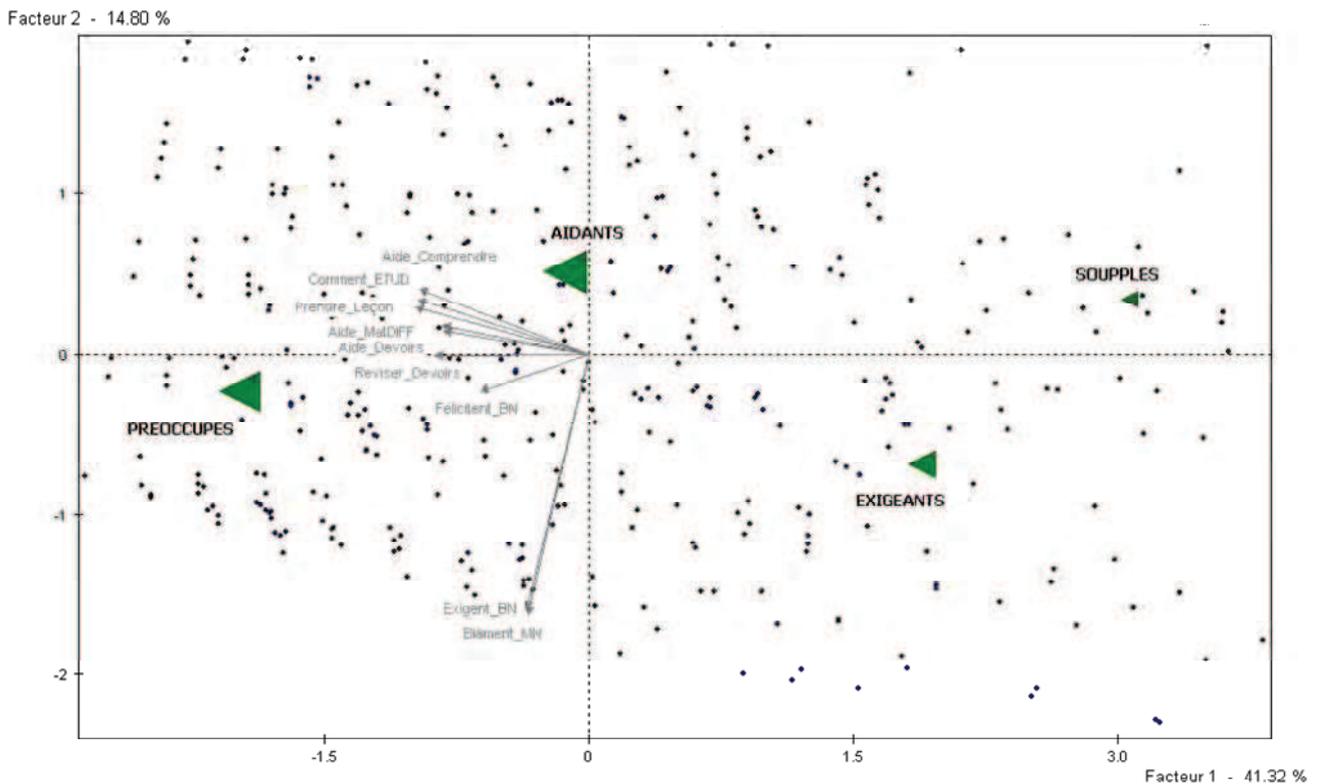
PROPOSITION		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Moyenne	2,9	2,7	2,9	2,8	2,6	3,2	2,8	3,6	3,3	3,3
	Écart-type	1,1	1,1	1,1	1,1	1,1	1,0	1,1	0,8	0,9	0,8
	n	278476									
(Mes parents...)		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) m'apprenent comment étudier											
(2) repasent avec moi les leçons		0,64									
(3) ...m'aident à comprendre ce que je ne comprends pas		0,62	0,62								
(4) ...m'aident à faire les devoirs		0,46	0,45	0,46							
(5) ...revisent si j'ai fait les devoirs		0,47	0,53	0,44	0,43						
(6) ...m'aident à sentir confiance quand je ne comprends pas		0,44	0,43	0,46	0,37	0,41					
(7) ...me réprimentent si je n'ai pas des bonnes notes		0,14	0,15	0,10	0,15	0,20	0,10				
(8) ...me félicitent si j'ai des bonnes notes		0,27	0,28	0,29	0,25	0,26	0,39	0,12			
(9) ...m'exigent des bonnes notes		0,12	0,13	0,10	0,12	0,16	0,13	0,41	0,17		
(10) Je(nous/parents) revise(ons) le cahier des notes		0,07	0,07	0,10	0,06	0,05	0,09	0,02	0,06	0,04	
Alpha de Cronbach = 0,78		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MSA Général = 0,86	MSA	0,87	0,86	0,87	0,93	0,90	0,89	0,65	0,88	0,65	0,89
Structure Factorielle après extraction (ULS)	Dim1	0,78	0,77	0,77	0,64	0,61	0,60	0,41	0,11	0,18	0,20
et rotation (HK)	Dim2	0,22	0,17	0,20	0,29	0,22	0,22	0,25	0,05	0,69	0,59



Source : élaboration de l'auteur à partir des données SIMCE, 2004. Toutes les corrélations notées sont significatives (à $p < 0,0000$) et elles portent sur des échelles allant de 1 à 4. Les graphiques en bas illustrent ce qui est décrit sur le tableau.

Tableau B18
 Catégorisation par CLUSTERS proposée pour la Dimension **Appui des parents**.
 Plan factoriel positionnant les élèves, les variables et les catégories construites.
 (2004 - 8^{ème} grade)

N°	Variables / caractéristiques	Moyenne générale (1 à 4)	CLUSTERS				R ²
			Préoccupés	Aidants	Exigents	Souples	
(Mes parents...)							
(1)	m'apprennent comment étudier	2.91	3.70	3.07	1.76	1.78	0.54
(2)	repassent avec moi les leçons	3.26	3.63	2.81	1.58	1.61	0.56
(3)	...m'aident à comprendre ce que je ne comprends pas	2.70	3.72	3.12	1.80	1.79	0.54
(4)	...m'aident à faire les devoirs	2.87	3.62	2.86	2.04	1.91	0.37
(5)	...révisent si j'ai fait les devoirs	3.55	3.58	2.56	1.79	1.62	0.44
(6)	...m'aident à sentir confiance quand je rencontre des difficultés	2.64	3.81	3.31	2.68	1.92	0.36
(7)	...me réprimandent si je n'ai pas des bonnes notes	2.79	3.29	2.46	2.83	2.29	0.13
(8)	...me félicitent si j'ai des bonnes notes	3.20	3.88	3.64	3.80	1.81	0.57
(9)	...m'exigent des bonnes notes	2.84	3.68	2.93	3.39	2.69	0.16
(10)	Je/nous (parents) révise/ons le cahier des notes	3.25	3.33	3.28	3.21	3.11	0.01
ELEVES		278476	33.5%	35.3%	20.4%	10.8%	0.37



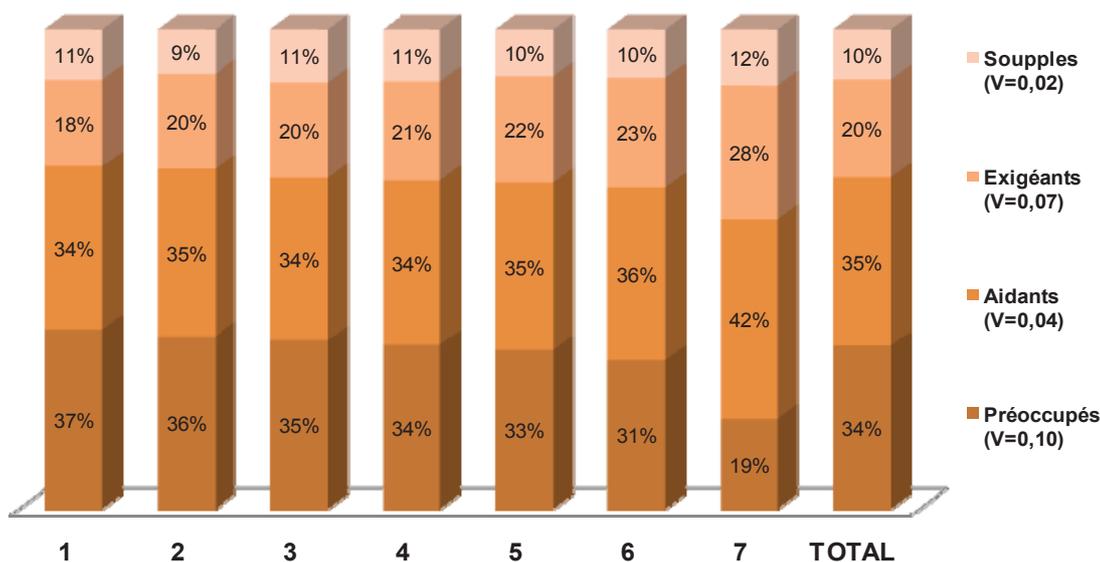
Source : élaboration de l'auteur à partir des données SIMCE, 2004. Sur le tableau, on peut apprécier les valeurs pour chaque catégorie ou cluster des items-variables originaux, ainsi que la proportion d'élèves pour chacune de sous-divisions créées et les R² que la partition faite entretienne avec chacune (R² moyen noté en bas à droite).

L'appui des parents : des profils différents

Sur le graphique B13, nous pouvons apprécier la distribution, générale et par type d'école, des quatre catégories de parents extraites par la méthode clusters. De même que pour les dimensions de bien-être et de confiance en soi scolaire, nous rencontrons une situation polarisant le public qui attend les établissements privés payants et le reste des catégories. En effet, les parents des élèves qui y sont inscrits sont davantage placés dans les catégories « aidants » (42% contre un 35% en moyenne) et désertent considérablement l'autre catégorie majoritaire. En effet, seulement un élève sur cinq aurait des parents « préoccupés » dans le privé payant contre un élève sur trois au niveau national. Toutefois, notons que les différences ne sont pas très fortes selon le type d'école. Cela dit, elles sont bien présentes et arrivent à des niveaux non méprisables dans le cas des catégories « préoccupés » ($V=0,1$) et « exigeants » ($V=0,07$), concentrant leurs principaux déséquilibres dans la catégorie privé payant.

Graphique B13

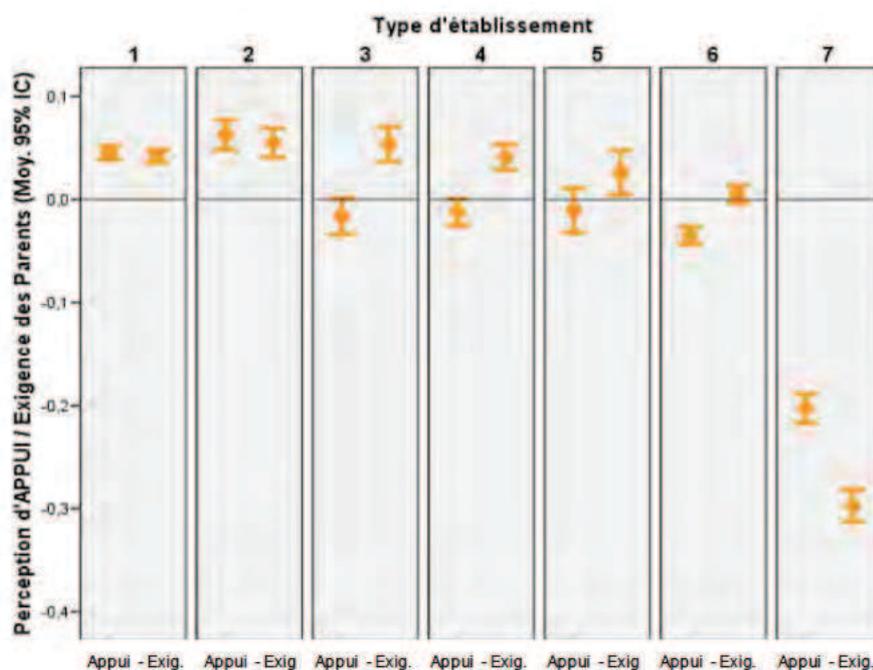
Proportions des catégories de parents construites à partir des déclarations des élèves sur la manière dont ceux-ci les aident/exigent vis-à-vis de leur travail scolaire. Par type d'établissement. (n=23 6113 élèves urbains dans des classes à taille ≥ 20) (8ème grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Les élèves placés dans des établissements privés payants totalisent 19% classifiés comme ayant des parents à profil « Préoccupé », tandis qu'à niveau général ce chiffre atteint 34%.

Graphique B14

Valeurs moyennes des facteurs extraits pour les sous-dimensions **appui et exigences des parents** (ressentis). Avec limites de confiance (à $p < 0,05$) et selon type d'établissement. (278476 élèves - 2004 - 8^{ème} grade)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

Pour sa part, le graphique B14 illustre les moyennes de deux échelles factorielles construites selon notre catégorisation des écoles. Les différences appréciées sont concordantes avec ce que montre le graphique B13. Toutefois, elles font davantage ressortir la différenciation négative, en moyenne et par rapport à toutes les autres catégories, de la part des élèves du privé payant sur les deux sous-dimensions construites.

Les rapports avec les enseignants

De même que pour la dimension reliée aux rapports scolaires avec les parents, les visions entretenues par les élèves sur les pratiques concrètes des enseignants envers eux, restent un sujet incontournable, autant d'un point de vue pratique que théorique (cf. Bressoux, 1994, 2001 ; Bressoux et Pansu, 2003 ; Pansu et al., 2007). Toutefois, nous allons considérer ici ce facteur davantage comme une variable contrôle, faite à partir des déclarations des élèves. Nous n'allons donc pas approfondir sur le sujet, nous contentant d'objectiver des mesures qui soient cohérentes et indicatives à cet égard. Hélas, l'information de base sur les pratiques enseignantes et les rapports entre ces deux acteurs, n'est disponible que pour les mathématiques et le langage. Toutefois, et même si cela crée une légère asymétrie avec les deux autres disciplines, nous nous servons de ces indices.

Nous avons donc analysé une question portant sur huit propositions où les élèves étaient invités à se prononcer sur une échelle d'attitudes à quatre niveaux (i.e. fréquences) pour chaque proposition. Celles-ci étant reliées, d'une façon plutôt spécifique, aux manières d'enseigner de l'enseignant (e.g. « *il explique les erreurs* », « *il révisé les devoirs* », « *il m'encourage* », etc.). Bien entendu, les réponses à ces questions sont, encore une fois, probablement chargées des différents biais (e.g. acquiescence) mais, dans le total, celles-ci reflètent des situations qu'il faut considérer, notamment quand on va s'intéresser aux interactions des élèves entre eux et plus spécifiquement, aux sentiments rapportés au vécu scolaire, qui, on le sait, sont fortement façonnés par ces rapports (cf. chapitres n°5 et n°4).

Sur le tableau B19 et le graphique B15, il est possible de voir les principaux résultats des analyses de base faites pour objectiver cette dimension pour chacune des deux disciplines concernées. Malgré leur liaison prévisible entre les deux ressentis par discipline, nous n'avons pas voulu extraire une autre mesure composite. Ceci compte tenu de la spécificité que les perceptions et les rapports véhiculés par ces mesures peuvent avoir vis-à-vis de l'enseignant lui-même, ou encore, vis-à-vis de la discipline. Approximativement, les indices uniques extraits par une analyse factorielle (cf. annexes méthodologiques) expriment un continu des relations perçues allant des plus mauvaises aux meilleures. Le score de chaque élève est, dans ce cas, rapporté à la moyenne générale, sachant toutefois qu'une approche le reliant plutôt à la classe pourrait être plus heuristique, si nous présentons les deux possibilités. Cependant, certains défis méthodologiques à considérer, notamment lors de l'interprétation des coefficients issus des régressions multiniveaux nécessitant de la centration précise des variables (cf. annexes méthodologiques), nous font préférer pour l'instant l'échelle d'ordre général. De plus, les deux mesures ici illustrées sont plutôt bien corrélées entre elles (r de Pearson ≈ 0.86).

Tableau B19

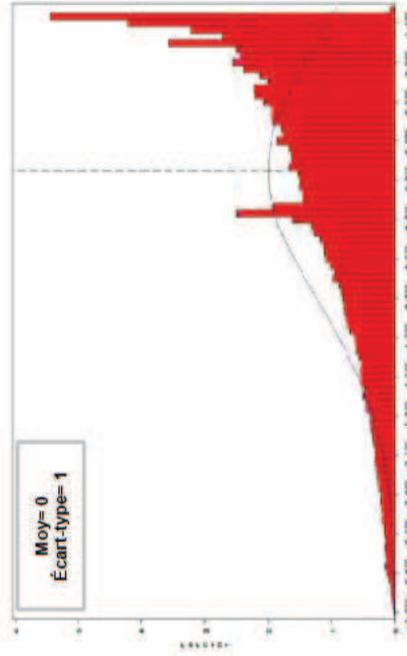
Matrices de Corrélations de Pearson pour les items-variables représentant la dimension **rappports et perceptions des élèves envers leurs enseignants** en Maths et Langage. Informations additionnelles sur l'extraction de scores factoriels à partir de celles-ci (2004 - 8^{ème} grade)

PROPOSITION	MATHÉMATIQUES								LANGAGE							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Moyenne</i>	3.1	3.2	3.3	3.2	3.4	3.3	3.2	3.4	2.9	2.9	3.1	3.2	3.3	3.1	3.2	3.5
<i>Écart-type</i>	1.1	0.9	0.9	1.0	0.9	1.0	1.0	1.0	1.1	1.0	1.0	0.9	0.9	1.0	1.0	0.9
<i>n</i>	278476								278476							
(Mon professeur...)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) ...m'apprenne comment étudier	0.35								0.43							
(2) ...répète plusieurs fois les sujets	0.36	0.39							0.42	0.50						
(3) ...explique des différentes formes les sujets	0.24	0.17	0.23						0.30	0.26	0.30					
(4) ...révise les devoirs	0.32	0.25	0.31	0.38					0.40	0.38	0.41	0.43				
(5) ...explique les erreurs dans les devoirs et contrôles	0.39	0.32	0.35	0.31	0.47				0.46	0.44	0.44	0.34	0.52			
(6) ...m'aide quand on voit un sujet difficile	0.43	0.30	0.34	0.27	0.41	0.59			0.50	0.40	0.41	0.33	0.47	0.64		
(7) ...m'encourage pour que j'étudie et obtienne des bonnes notes	0.10	0.07	0.09	0.09	0.16	0.23	0.28		0.08	0.09	0.09	0.06	0.13	0.18	0.21	
(8) ...n'est pas indifférent si j'ai des mauvaises notes																
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Alpha de Cronbach = 0,807 MSA Général = 0,869								Alpha de Cronbach = 0,772 MSA Général = 0,840							
<i>MSA</i>	0.90	0.88	0.88	0.89	0.88	0.85	0.84	0.83	0.87	0.84	0.86	0.86	0.86	0.82	0.81	0.81
<i>Dim1</i>	0.62	0.69	0.71	0.44	0.60	0.65	0.61	0.12	0.58	0.61	0.64	0.35	0.48	0.55	0.53	0.12
<i>Dim2</i>	0.60	0.54	0.55	0.46	0.65	0.79	0.79	0.23	0.53	0.40	0.47	0.43	0.61	0.76	0.75	0.31

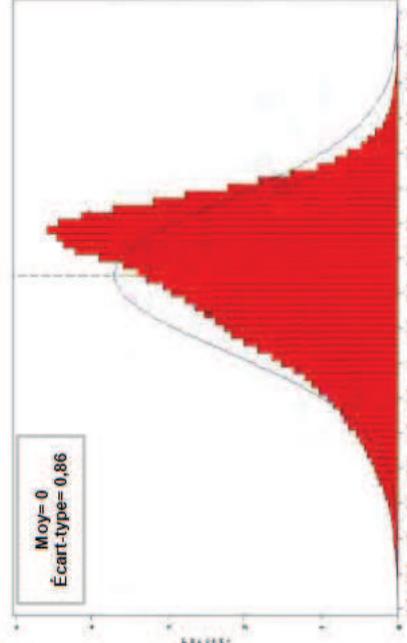
Source : élaboration de l'auteur à partir des données SIMCE, 2004. Matrices de corrélation de Pearson pour les huit propositions analysées pour établir une dimension sur les rapports et appréciations des élèves vis-à-vis des enseignants. Ceci est noté pour les deux disciplines dont cette information a été demandée. L'échelle ordinale considère quatre niveaux pris ici de manière ascendante (i.e. de - à +). Toutes les corrélations notées sont significatives (à p < 0,0000). Les Alpha de Cronbach correspondent aux estimations sur des valeurs standardisées. Il faut consigner que la structure factorielle notée en bas à deux facteurs n'est présentée qu'à titre illustratif puisque nous avons décidé de n'extraire qu'une seule dimension eu égard de ces premiers résultats. En effet, cette structure est « complexe » dans les deux cas.

Graphique B15

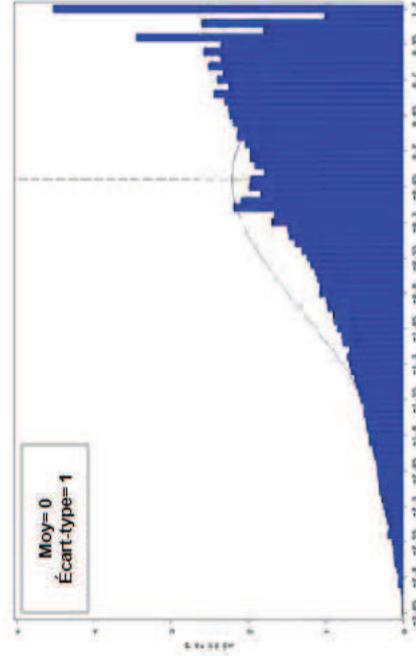
Histogrammes des facteurs uniques extraits pour objectiver une dimension de rapports et perceptions des élèves vis-à-vis des enseignants en Maths et Langage. Selon type de centration. (278476 élèves - 2004 - 8^{ème} grade)



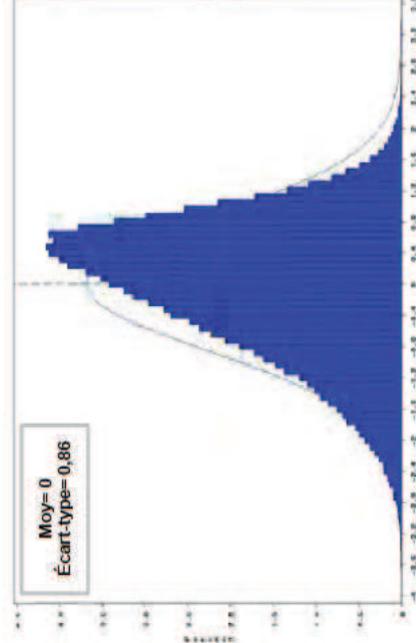
Rapports avec les enseignants (MATHS)
(Facteur unique)



Rapports avec les enseignants (MATHS)
(Facteur unique - Centré au niveau CLASSE)



Rapports avec les enseignants (LANG)
(Facteur unique)

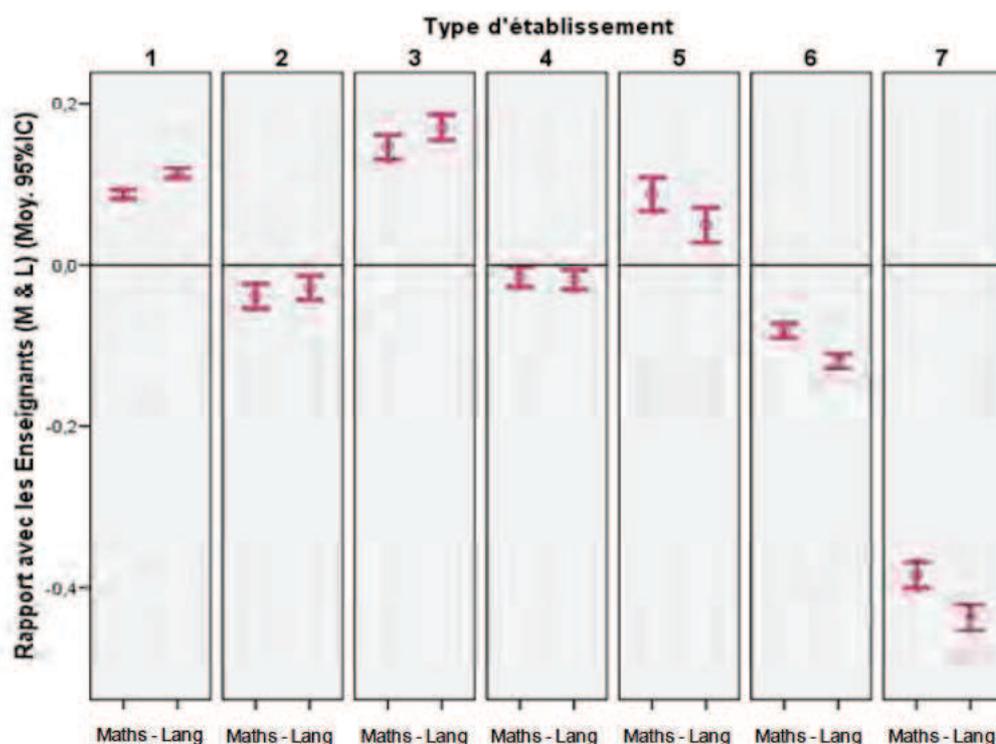


Rapports avec les enseignants (LANG)
(Facteur unique - Centré au niveau CLASSE)

SOURCE : élaboration de l'auteur à partir des données SIMCE, 2004. Ces scores factoriels ont été extraits à l'aide de la méthode ULS. La corrélation entre ces deux mesures est de $r=0,493$ ($R^2=0,24$) pour l'extraction générale, puis de $r=0,536$ ($R^2=0,28$) quand on les regarde sous l'approche de niveau classe (i.e. en centrant le score de chaque élève sur la moyenne de sa classe).

Graphique B16

Valeurs moyennes avec leurs limites de confiance (à $p < 0,05$) des facteurs extraits pour objectiver les rapports et perceptions des élèves vis-à-vis des enseignants en Maths et Langage. Selon type d'établissement. (278 476 élèves - 2004 - 8^{ème} grade)



Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

Enfin, le graphique B16 illustre les différences moyennes entre catégories d'établissement pour les échelles factorielles (à facteurs unique) exprimant les rapports aux enseignants, et selon discipline. Cette fois-ci la différenciation des moyennes est davantage marquée entre le secteur public et privé subventionné non sélectif et gratuit, et le privé payant (écarts autour de $|0,5\sigma|$).

Affinités, importance accordée et fréquence d'étude déclarée par discipline

Parmi les autres dimensions subjectives considérées ici, se trouvent celles reliées aux rapports aux disciplines, notamment du point de vue de la valorisation de celles-ci. Cette valorisation est prise, d'une part, en termes d'affinité/goût, et d'autre part, en termes d'importance accordée à chacune des matières scolaires traitées. Pour les deux cas, il s'agit d'une échelle d'attitudes toujours à quatre niveaux et portant sur le degré d'accord avec des affirmations du type : « *J'aime bien apprendre les mathématiques* » pour le premier groupe, et du type : « *Mathématiques c'est une matière importante pour la vie* » dans le deuxième. De toute évidence la question réunissant ces huit propositions (deux par discipline) semble pensée sous l'optique des théories sur la motivation scolaire (cf. Ryan et Deci, 2000) distinguant entre motivation intrinsèque (« *j'aime bien ...* ») et extrinsèque (« *il est important de...* »). En dépit du fait que ces items puissent être analysés en soi, nous avons décidé de ne les considérer que comme des facteurs de contrôle. En effet, il se peut que certaines interactions et manières d'étudier des élèves soient bien corrélées avec des niveaux plus ou moins importants de motivation sous chacune des approches notées. De ce fait, l'inclusion des variables censées refléter ces appréciations dans des modèles testant l'impact des interactions et d'autres ressentis sur les acquis scolaires, nous semble une précaution tout à fait nécessaire.

La question qui se pose dès lors est « à quel degré de profondeur analyser les variables reliées à la motivation pour les objectiver ? ». En toute logique, signalons que ces mesures sont bien corrélées entre elles, au total et par discipline. Néanmoins, nous avons laissé de côté la construction d'indices composites les reliant pour préférer les penser en termes plus grossiers mais plus proches des déclarations directement faites. De ce fait, nous n'allons prendre pour le moment que les réponses sur les affinités ou le goût déclaré de manière ordinale à trois niveaux.²⁷ Ce choix est fondé sur les corrélations entretenues entre ces déclarations et le fait qu'elles semblent converger dans une même dimension (i.e. Alphas de Cronbach entre 0,58 et 0,70). Tout ceci est montré sur le tableau B20.

Pour les analyses, nous allons retenir principalement l'affinité déclarée envers la discipline car celle-ci peut exprimer, en théorie, un lien au préalable plus fort et explicite avec les apprentissages pour une matière donnée, et qu'elle peut aussi, refléter davantage d'autres attitudes telles qu'une plus grande concentration et/ou plus de collaboration en cours, par exemple. Toujours est-il que nous nous servons des échelles Likert respectives à des fins illustratives sur les rapports entre cette dimension et d'autres variables dans l'analyse.

²⁷ Ces trois niveaux correspondent à : 1) « très en désaccord » + « en désaccord » ; 2) « d'accord » et 3) « très d'accord ». Notons que les proportions les plus éloignées des disciplines (i.e. « très en désaccord ») étaient plutôt faibles en effectifs (entre 5% et 1,8%). Leur fusionnement ne répond qu'à des raisons pratiques vouées à ne pas augmenter davantage les sous-divisions chez les analyses par régression. En outre, notons que la factorisation n'était pas du tout aisée pour l'ensemble de ces mesures. Cela était du notamment à des « communautés » plus grandes que le nombre 1 dans certains cas, ce qui montrerait que la matrice des corrélations de ces items cache de hautes similitudes entre mesures. Ainsi, seule l'extraction par composantes principales était admise (cf. annexes méthodologiques). Nous ne prendrons pas cette voie.

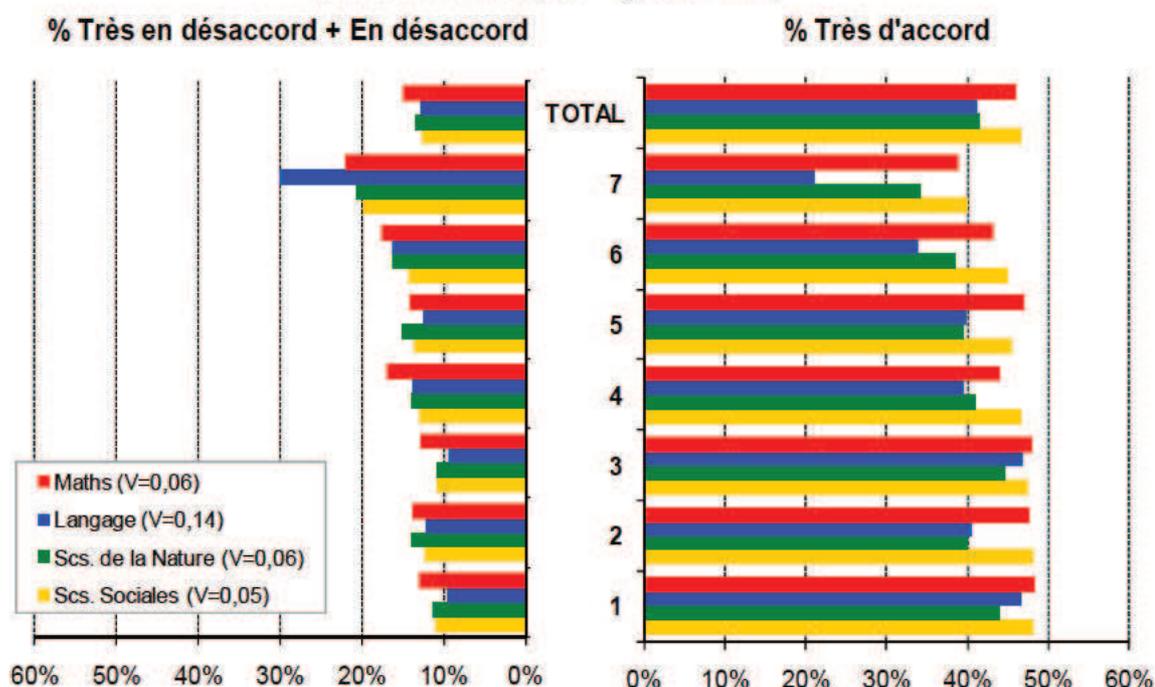
Tableau B20
Matrices de Corrélations de Pearson pour les items-variables représentant l'affinité et l'importance accordées à chaque discipline. (2004 – 8ème grade)

	Maths		Langage		Scs. NAT		Scs. SOC		Général
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Moyenne</i>	3.3	3.6	3.2	3.4	3.2	3.3	3.3	3.4	3.3
<i>Écart-type</i>	0.8	0.6	0.8	0.7	0.8	0.7	0.8	0.7	0.8
<i>n</i>	278476								
PROPOSITION									
(1) J'aime bien apprendre Maths	0.42								
(2) Maths c'est une matière importante pour la vie	0.18	0.14							
(3) J'aime bien apprendre Langage	0.20	0.37	0.46						
(4) Langage c'est une matière importante pour la vie	0.23	0.21	0.31	0.25					
(5) J'aime bien apprendre Sciences de la Nature	0.21	0.33	0.24	0.39	0.54				
(6) Sciences de la Nature c'est une matière importante pour la vie	0.23	0.20	0.32	0.26	0.36	0.27			
(7) J'aime bien apprendre Sciences Sociales	0.19	0.33	0.24	0.41	0.27	0.49	0.54		
(8) Sciences Sociales c'est une matière importante pour la vie	0.30	0.18	0.42	0.29	0.30	0.26	0.28	0.24	
(9) (Je vais à l'école parce que) J'aime bien étudier (*)									
<i>Alphas de Cronbach</i>	Maths		Langage		Scs. NAT		Scs. SOC		-
<i>PAR DISCIPLINE</i>	0,58		0,63		0,70		0,70		-
<i>GLOBAL (a)**</i>									0,78
<i>GLOBAL (b)**</i>									0,79
<i>AFFINITÉS</i>									0,60
<i>IMPORTANCE</i>									0,72

Source : élaboration de l'auteur à partir des données SIMCE, 2004. Matrice de corrélation de Pearson pour diverses propositions liées à la valorisation des disciplines en termes d'affinités et d'importance accordée (sur une échelle à quatre degrés d'accord). Toutes les corrélations notées sont significatives (à p < 0,0000). (*) L'item n°9 correspond à une question directe sur les motivations pour aller à l'école. Nous l'avons intégré dans ces analyses en trouvant un Alpha de Cronbach Global (**) sans cet item (a), puis, en l'intégrant (b). Les valeurs de cet indice de consistance interne sont aussi communiquées par discipline (reliant affinités et importance), puis par type de proposition (affinité ou importance).

Graphique B17

Proportions d'élèves par discipline exprimant leur degré d'accord ou de désaccord face à la proposition « J'aime bien apprendre... ». Selon type d'établissement.
(n=27 8476 élèves) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture : 14,9% du total d'élèves déclare qu'il est en désaccord ou très en désaccord avec l'affirmation « j'aime bien apprendre mathématiques ». Ils sont 46% à être « très d'accord » avec cette proposition. Les coefficients V de Cramer ont été estimés à partir du croisement entre les variables catégorielles à trois niveaux et la typologie d'écoles.

Sur le graphique B17, nous apprécions qu'il existe des différences dans la valorisation des disciplines qui semblent reliées au type d'établissement fréquenté. Ce sont surtout les établissements privés payants qui s'écartent très fortement des autres, avec des pourcentages de refus par discipline supérieurs ou égaux à 20%. En revanche, ces mêmes pourcentages restent proches de 10% pour les deux types d'établissement classés comme gratuits en non-sélectifs. A l'opposé, du côté de l'affinité affirmée, la situation est inverse : les privés payants cumulant de faibles proportions se déclarant « très d'accord » pour les différentes matières ($\leq 40\%$). Ces rapports sont particulièrement forts en langage, discipline qui cumule le plus d'écarts entre les catégories et la moyenne générale.

La fréquence d'étude en dehors des cours déclarée pour chaque matière

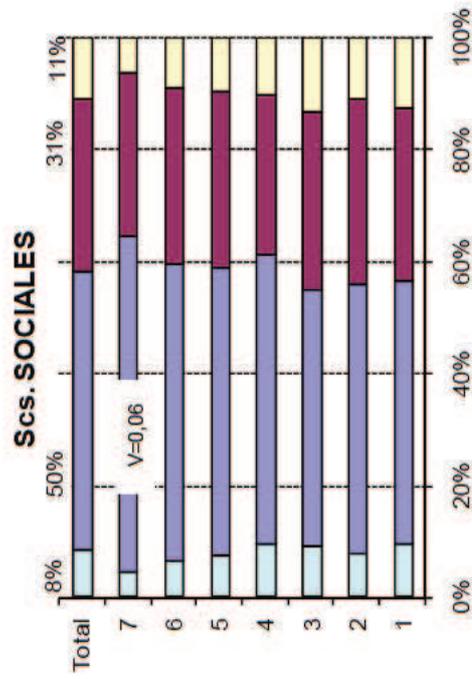
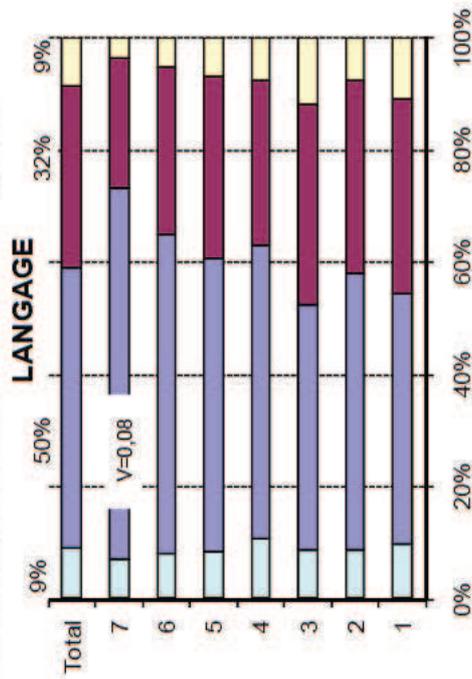
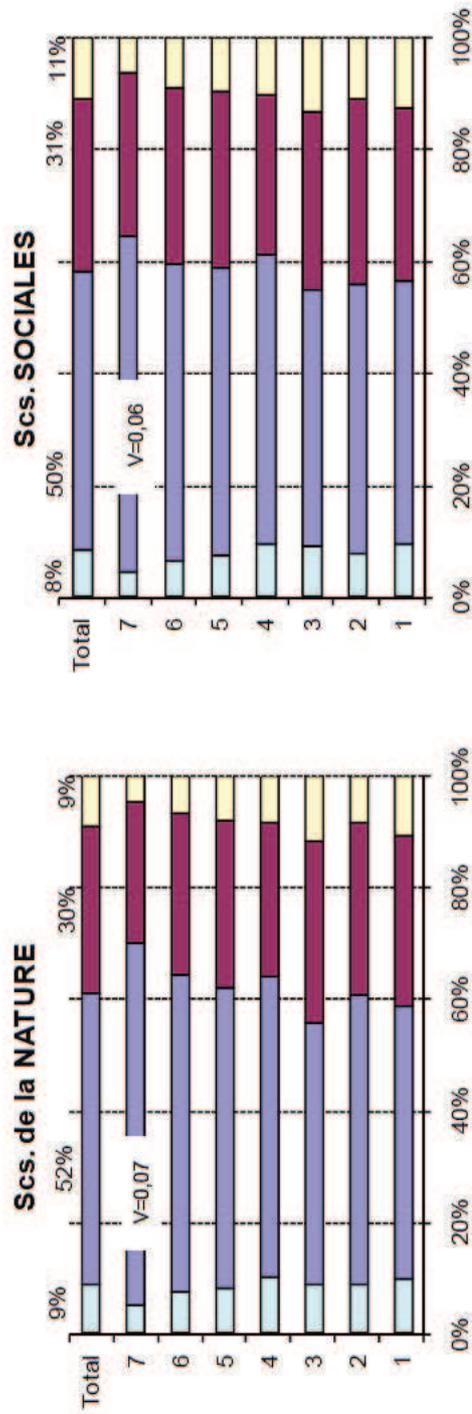
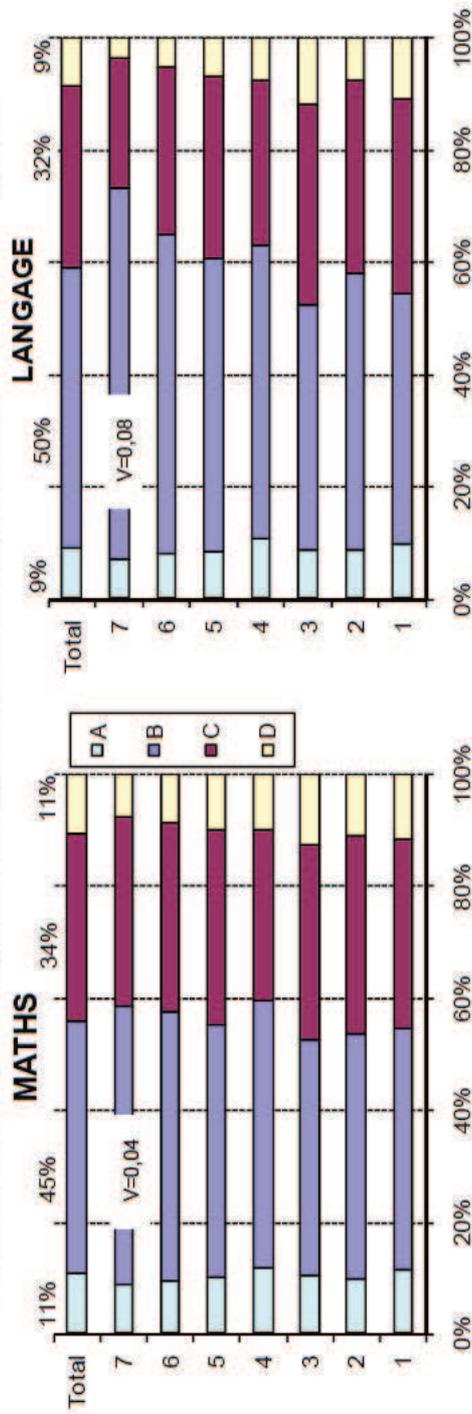
Un autre aspect qui n'est pas trop éloigné de l'importance et l'affinité accordées aux disciplines correspond au temps consacré à l'étude de chacune des disciplines. Ceci à partir des items cherchant une approximation, toujours déclarative, sur la fréquence et/ou les moments et opportunités dans lesquels les élèves se consacraient à étudier chaque matière en dehors des cours. Ainsi, les élèves étaient amenés à répondre sur une échelle à quatre niveaux composée de : « *Je n'étudie jamais en dehors des cours ou presque* » ; « *Je n'étudie que quand il y a des contrôles* » ; « *J'étudie quelques jours par semaine* » ; « *J'étudie tous les jours ou presque* ».

Encore une fois, avec ces estimations sur la fréquence/assiduité d'étude déclarée nous sommes face à une variable entachée d'incertitude. Toutefois, nous ne pouvons pas négliger a priori le probable lien entre la relative intensité du travail scolaire des élèves en dehors des cours et d'autres facteurs. Notamment, avec leurs performances (même si dans des sens parfois difficiles à décrypter), mais aussi, avec leurs rapports aux disciplines et leurs pratiques d'étude. Bien entendu, ces déclarations peuvent être polluées par plusieurs biais, dont celui de sentir que l'on travaille beaucoup pour une discipline dans laquelle on n'excelle pas et qui n'est pas non plus une des préférées. Ou bien, à l'envers, sentir que l'on n'a pas besoin de beaucoup de travail pour une discipline qu'on aime et qui est considérée « facile ».

Quoi qu'il en soit, ce contrôle s'avère important et nous le garderons ici pris d'une manière catégorielle en forme binaire. De ce fait nous allons regrouper les modalités « *Je n'étudie jamais en dehors de cours ou presque* » et « *Je n'étudie que quand il y a des contrôles* » (entre 56% et 61% d'effectifs selon discipline), puis les modalités restantes, représentant 44%, 41%, 39% et 42% respectivement en maths, langage, sciences de la nature et sciences sociales. Ci bas, sur les graphiques B18 et B19, nous communiquons les différences, pour les quatre niveaux de cette variable, associées aux types d'établissement. Notons enfin que les corrélations entre ces fréquences d'étude déclarées par discipline, sont plus élevées que pour d'autres variables ici analysées (r de Spearman allant entre +0,55 et +0,65), totalisant un Alpha de Cronbach de 0,86. Nous avons donc entrepris la construction d'une variable générale correspondant à la moyenne arrondie des quatre disciplines, à utiliser dans certains cas.

Graphique B18

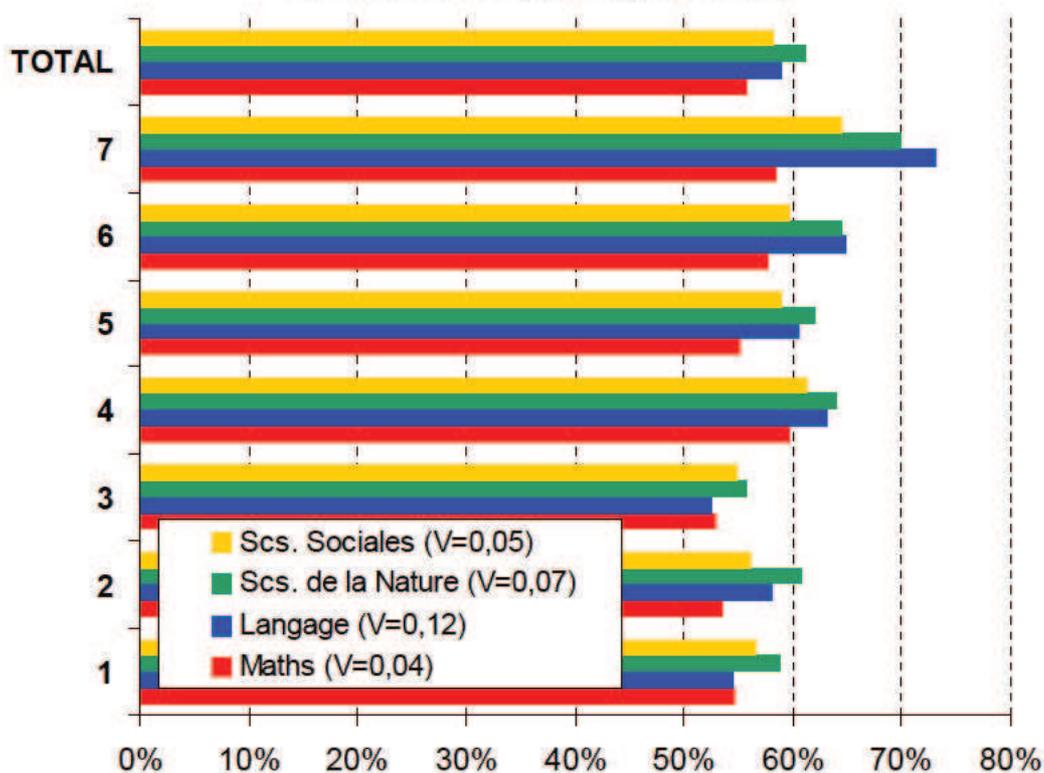
Graphiques exprimant les proportions d'élèves selon les déclarations par rapport à la fréquence d'étude en dehors des cours par discipline. Selon type d'établissement. (278 476 élèves ; 2004 - 8^{ème} grade)



Source : élaboration de l'auteur à partir des données SIMCE, 2004. Les options dans chaque discipline correspondent aux pourcentages déclarant à : (A) « Je n'étudie jamais en dehors de cours ou presque » ; (B) « Je n'étudie qu'en face des contrôles » ; (C) « J'étudie quelques jours la semaine » ; et (D) « J'étudie tous les jours ou presque ». Les V de Cramer correspondent à cette même échelle prise mais comme variable catégorielle. Pour chaque total, les pourcentages sont notés en haut.

Graphique B19

Pourcentages d'élèves par discipline et selon type d'établissement, déclarant ne pas étudier en dehors des cours ou bien de ne le faire qu'en face des contrôles.
(n=278 476 élèves) (8ème grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Notons tout de même que, bien qu'existant des différences entre les élèves d'après le type d'établissement qu'ils fréquentent, celles-ci ne sont pas d'une grande taille. Sauf dans le cas de Langage (V=0,12) les divergences entre proportions d'élèves déclarant une ou une autre fréquence de travail en dehors des cours ne semblent pas très importantes.

Des bonnes raisons pour aller à l'école

Enfin, une dernière question rapportée directement à la dimension de motivation en termes globaux est celle des raisons déclarées pour aller à l'école. Cette motivation générale aurait été saisie à partir de six items pour lesquels on demandait aux élèves de se prononcer, en quatre niveaux d'accord, par rapport aux raisons pour aller dans leurs écoles. Il s'agit d'items du type : « *J'y vais parce que j'aime bien étudier* » ; « *J'y vais pour décrocher un bon travail* » ; ou encore, « *J'y vais parce que j'en suis obligé* » ou « *... parce que c'est ce que je dois faire à mon âge* ». ²⁸ A la lumière de tels propos on voit que cette question semble être davantage explicite vis-à-vis de ce que l'on peut penser comme motivation, soit elle intrinsèque (« *J'aime bien* ») ou extrinsèque (« *J'y vais pour avoir après...* »). Toutefois, celle-ci n'exprime que des situations générales.

Pour ce groupe d'items, nous avons constaté qu'il était difficile de le penser en termes d'une seule dimension (e.g. exprimant différents degrés d'utilitarisme face aux études) montrant un Alpha de Cronbach < 0,50. Ceci est rapporté sur le tableau B21 ci bas. Toutefois, en regardant de manière plus spécifique par sous-regroupements, ces valeurs pourraient s'améliorer (entre 0,63 et 0,55) pouvant autoriser éventuellement la construction d'indices composites ; le cas échéant, justifiant le choix d'un des items comme étant censé refléter, en quelque sorte, les autres.

Toujours est-il que nous avons décidé de prendre chacune de ces propositions comme une variable en soi, tout en regardant leurs proximités pour ne pas dupliquer l'information au cas où celles-ci seraient mises dans des analyses à plusieurs variables. Sur le tableau B21 on peut voir la matrice des corrélations de celle-ci. Tandis que sur le tableau B22, on apprécie une illustration de la valeur de ces variables en relation à notre typologie d'établissements (sur l'échelle Likert). On y voit aussi la signification de ces différences, lesquelles restent plutôt discrètes malgré leur signification statistique. Enfin, disons qu'une analyse des valeurs associées à ces affirmations, en rapport aux profils généraux d'élève selon les pratiques d'étude — i.e. des interactions déclarées avec les autres élèves —, montre, encore une fois, des différences très réduites.

²⁸ Notons que des items semblables ont été utilisés par Van Petegem et al. (2008) comme variables explicatives de leur échelle de bien-être scolaire au 10^{ème} grade.

Tableau B21

Matrices de Corrélations de Pearson pour les items-variables proposant différentes raisons pour aller à l'école. (2004 – 8^{ème} grade)

	(1)	(2)	(3)	(4)	(5)	(6)
Moyenne	3.2	2.2	3.7	2.7	3.6	3.3
Écart-type	0.8	1.1	0.6	1.1	0.7	0.9
<i>n</i>	278476					

PROPOSITION

(Je vais à mon école...)

(1) ... parce que j'aime bien étudier	-0.29					
(2) ... <i>parce que je suis obligé</i>	0.25	-0.09				
(3) ... pour avoir après un bon travail	0.01	0.30	0.06			
(4) ... parce que tous les jeunes comme moi le font	0.16	-0.003	0.46	0.16		
(5) ... pour avoir après un bon salaire	0.05	0.16	0.16	0.44	0.20	
(6) ... parce qu'à mon âge je dois de le faire						

Général

0,48

(3)+(5)

0,63

ALPHA DE CRONBACH

0,61

(1)+(3)+(5)

0,55

(2)+(4)+(6)

0,56

(3)+(4)+(5)+(6)

0,53

(2)+(3)+(4)+(5)+(6)

0,53

(1)+(3)+(4)+(5)+(6)

0,55

Source : élaboration de l'auteur à partir des données SIMCE, 2004. Matrice de corrélation de Pearson pour les six items reliés aux motivations déclarées pour aller à l'école (allant de 1 à 4). Nous présentons ici toutes les variables prises dans l'ordre original, tout en communiquant la consistance interne des différentes possibilités de regroupement de celles-ci. L'inversement de certains items (p. ex. n°2, n°4, n°6), pris sous différentes combinaisons, n'aboutit pas à des Alpha >0,50. Toutes les corrélations notées sont significatives (à p <0,0000) sauf celle marquée en italique (p<0,18).

Tableau B22

Tableau illustrant les valeurs moyennes des items-variables exprimant différentes raisons pour aller à l'école, selon type d'établissement. (278 476 élèves ; 2004 – 8^{ème} grade)

Type d'établissement	J'aime étudier	Je suis obligé	Bon travail après	Tous le font	Bon salaire après	Devoir de l'âge
1	3.30	2.13	3.68	2.73	3.60	3.33
2	3.16	2.20	3.70	2.62	3.59	3.29
3	3.30	2.14	3.68	2.74	3.60	3.32
4	3.09	2.27	3.65	2.65	3.57	3.27
5	3.15	2.12	3.68	2.56	3.54	3.24
6	2.99	2.28	3.65	2.55	3.54	3.24
7	2.60	2.69	3.56	2.56	3.40	3.23
<i>Total</i>	3.16	2.22	3.67	2.66	3.57	3.30
<i>R²</i>	0.065	0.018	0.003	0.005	0.006	0.002

Source : élaboration de l'auteur à partir des données SIMCE, 2004. Lecture Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). L'échelle d'expression de ces variables (allant de 1 à 4) a été prise ici comme continue. Les différences de couleur reflètent des valeurs significativement différentes entre types d'école pour chacune des propositions (à $p < 0,05$; Proc GLM de SAS®, option Tukey). En bas, sont notés tant la moyenne générale respective de chaque item que le coefficient R^2 de l'analyse de variance qui compare, dans chaque cas, les sept moyennes par catégorie d'établissement (toutes les analyses sont significatives à $p < 0,000$). Par exemple, 6,5% de variance des valeurs pour la proposition « (Je vais à l'école parce que) J'aime étudier » est relié à la typologie d'écoles, tandis que seulement 0,6% l'est pour l'item « (Je vais à l'école) pour avoir après un bon salaire ». Les items sont dans l'ordre dont ils ont été présentés aux élèves.

Les attentes des élèves, mais aussi celles des parents et des enseignants

Un autre aspect inéluctable parmi les croyances entretenues par les acteurs en situation scolaire, est constitué par les attentes de scolarisation postérieure qu'ils abritent. La littérature s'accorde à montrer ces croyances comme étant reliées aux acquis des élèves de manière synchronique mais aussi longitudinale (cf. Duru-Bellat et Van Zanten, 2006). En effet, nous avons vu que certaines connaissances de soi sont susceptibles d'influencer fortement le travail et les résultats des individus (cf. chapitre n°4).

Ainsi, nous avons objectivé ces attentes d'une manière presque directe, c'est-à-dire, en prenant l'information telle qu'elle a été fournie par les répondants aux questionnaires, puis, en construisant des variables catégorielles.²⁹ Malgré leur concordance manifeste, nous n'avons pas voulu faire des indices reliant l'avis des trois acteurs interrogés (élèves, parents et enseignants). Notamment, parce que ces opinions transmettent, à la base, une vision qui est singulière et spécifique de l'acteur concerné : de l'élève sur lui-même, des parents vis-à-vis de chaque enfant, puis enfin, des enseignants sur l'ensemble d'une classe. Dans certains cas, et compte tenu de la haute corrélation dans le total entre les attentes des parents et celles des élèves (r de Spearman=0,66), nous avons opté pour omettre l'une des deux mesures dans les analyses, tout en laissant l'opinion des enseignants qui est, d'ailleurs, exprimée depuis une autre perspective (opinion sur l'ensemble).³⁰

Comme cela peut être apprécié sur le graphique B20 (« total »), la plupart des élèves et de leurs parents visent le niveau supérieur. D'abord l'université, avec respectivement, 60% et 49% dans les établissements urbains, et 57% et 45% sur l'ensemble. Ensuite, le supérieur technique avec un total de 18% et de 30% dans les deux cas. De cette manière, on voit que la catégorie « arriver à l'université » devient la principale référence. Puis, nous allons donner priorité à l'avis des élèves car, on peut le voir, ils sont moins déterminés par notre première variable contextuelle (i.e. le type d'école). En outre, remarquons que, cherchant à être davantage près des analyses faites, sur le graphique B20 nous ne voyons que les classes urbaines et ayant plus de 19 élèves. Ceci sachant qu'une partie des différences entre typologies vis-à-vis des attentes des acteurs passe par la condition de ruralité, mais que ces rapports et ces distinctions demeurent semblables quand on ne compare que les établissements urbains.³¹ Tout cela confirme, encore une fois, la pertinence de la catégorisation proposée.

²⁹ Elèves et enseignants ont été interrogés sur cinq niveaux de scolarisation ultérieure, mais les parents n'ont été invités à se prononcer que sur quatre niveaux (omettant le secondaire technique comme option).

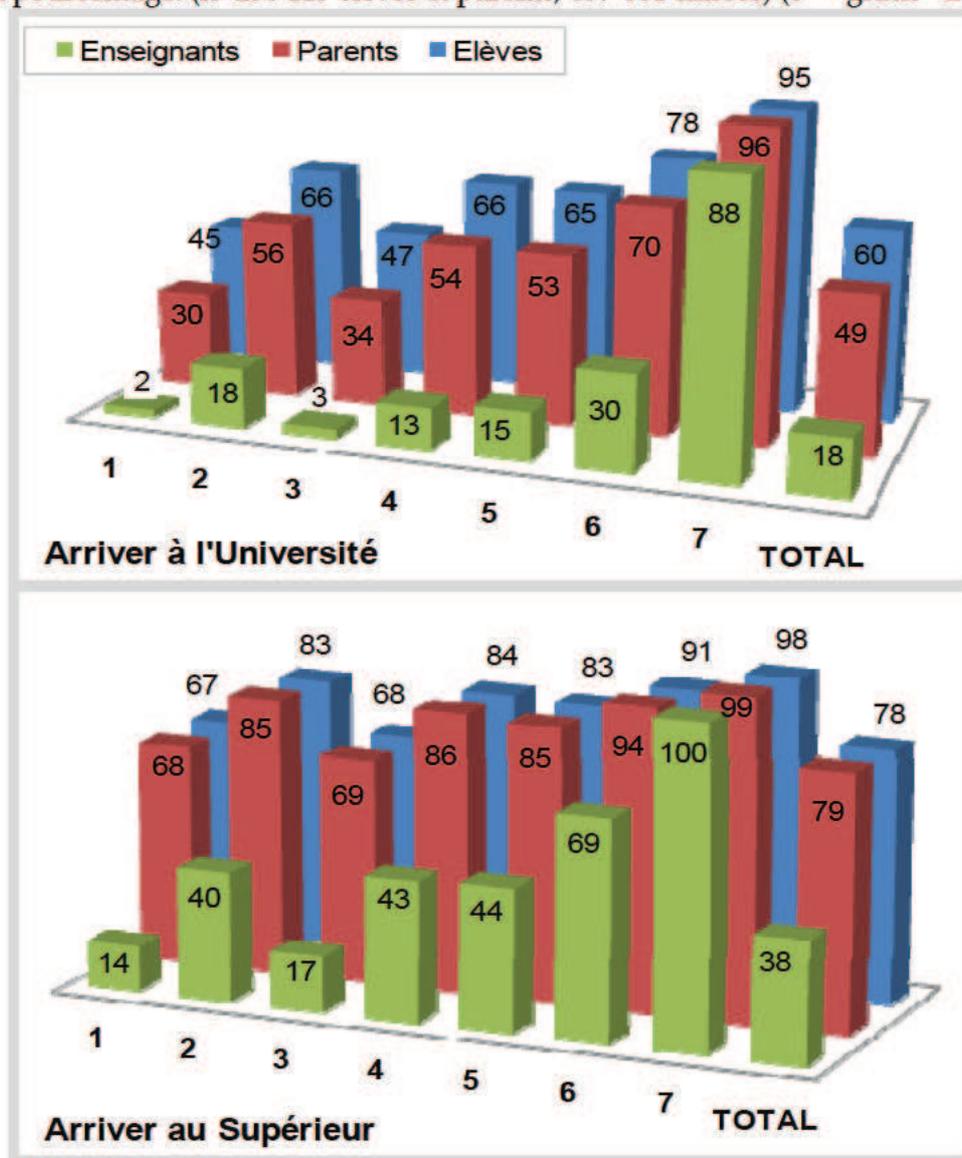
³⁰ En effet, si l'on prend les corrélations par classe ($n=9412$), par la construction de mesures agrégées des attentes des parents et des élèves, les corrélations entre l'avis des enseignants et celui des autres acteurs fluctuent, selon la discipline, entre 0,60 et 0,59 avec les élèves, et entre 0,60 et 0,62 avec les parents. Toutefois ces rapports montent à 0,66 et 0,68 si l'on considère, à son tour, une mesure agrégée des quatre enseignants par classe. Enfin, entre parents et élèves, le rapport entre attentes moyennes par classe serait de $r=0,83$.

³¹ En effet, ces coefficients descendent un peu quand on élimine de l'analyse les écoles rurales, à l'occurrence ils passent de $V=0,34$ et 0,41, respectivement pour élèves et parents dans la déclaration de croire qu'ils iront à

Graphique B20

Attentes de scolarisation après la fin du primaire pour les parents, les élèves et les enseignants,³²selon type d'établissement en milieu urbain.

En pourcentage. (n=236 113 élèves et parents, et 7 061 classes) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Le niveau « Supérieur » (graphique en bas) considère ensemble les niveaux « Supérieur Technique » et « Université ». Dans le cas des enseignants il s'agit de l'opinion moyenne par classe de ceux-ci.

l'université. Au niveau de la moyenne des enseignants par classe, pour cette même déclaration, les rapports passent de $V=0,64$ à $0,63$.

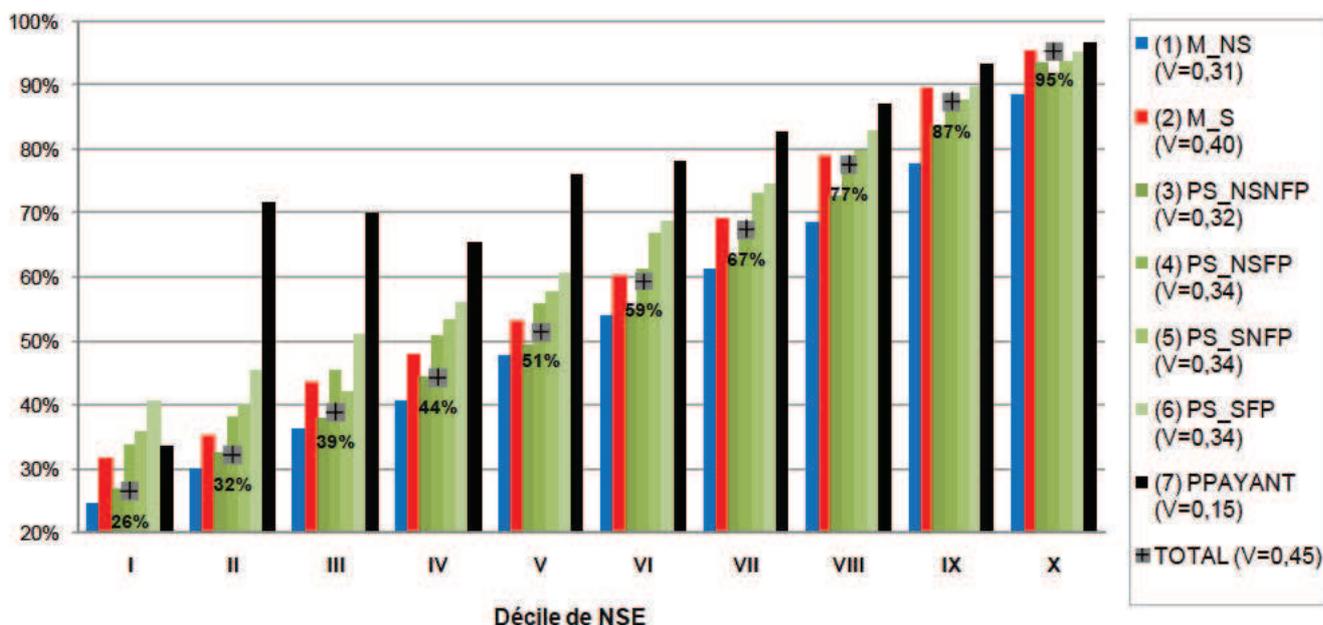
³²Notons que dans le cas des enseignants, ces variables expriment une moyenne arrondie par classe, toujours en forme de catégorie binaire.

Pour leur part, les enseignants se montrent beaucoup plus sévères par rapport aux attentes. En effet, seulement 15,7% parmi eux déclare croire que leurs élèves arriveront à l'université (18% dans les classes urbaines sur le graphique B21). Puis, un enseignant sur trois croit qu'ils arriveront, du moins, au supérieur (38% en situation urbaine). Cependant, comme attendu, ces attentes déclarées sont très fortement différenciés selon le type d'établissement où les enseignants travaillent (e.g. V de Cramer=0,64 pour le niveau universitaire et V=0,61 pour l'ensemble du supérieur).³³ Dans le cas des élèves, ce même coefficient V (brut) reliant les différents types d'école et la croyance qu'on arrivera à l'université n'est que de 0,37, contre 0,44 chez les parents. Dans le total, il est intéressant de noter aussi, que les attentes moyennes des professeurs par classe sont plus en phase avec celles des parents qu'avec celles des élèves, mais ces différences diminuent un peu au niveau agrégé par classe. Quoi qu'il en soit, les élèves semblent plus optimistes vis-à-vis de leur futur scolaire que leurs parents et surtout que leurs professeurs, bien que les rapports entre ces mesures restent tous positifs et hautement significatifs.

Des attentes modulées par le type d'établissement.

Graphique B21

Attentes de scolarisation des élèves après la fin du primaire, selon type d'établissement en milieu urbain et par décile de niveau socioéconomique.
En pourcentage. (n=236113 élèves dans des classes à ≥20) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture type d'établissement : Idem tableau antérieur. Lecture : les élèves du décile de niveau socioéconomique « I » (ligne noire à traits) déclarent, en moyenne, à 26,3% croire qu'ils arriveront jusqu'à l'université. Pourtant, les élèves de ce même décile, mais placés dans des établissements du Privé subventionné sélectif et avec FP (n°6) déclarent la même chose à 40,5%.

³³ Ces mesures reviennent à dire que près de 60% de variance des déclarations des enseignants sur ces niveaux d'attente (contre tous les autres) serait associée, en termes bruts, à la catégorisation des établissements utilisée.

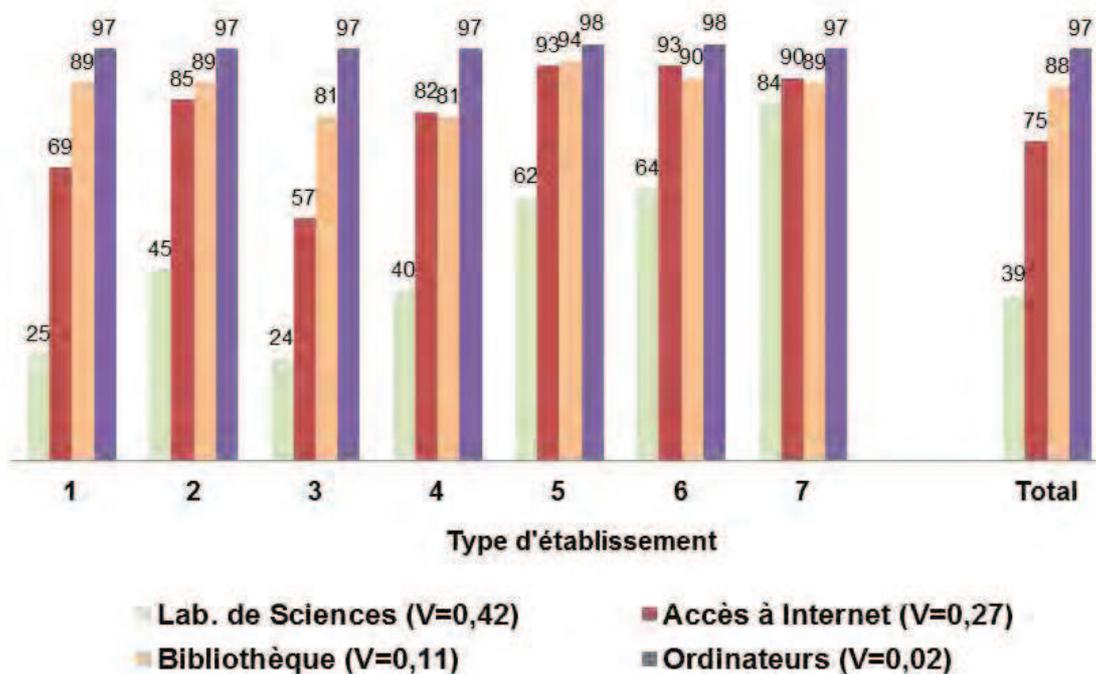
Le Graphique B21 illustre une situation très intéressante pour l'analyse des attentes des acteurs. En effet, nous apprécions que les élèves voient modulées leurs déclarations, en premier lieu, par leur niveau socioéconomique estimé (Indice de NSE, voir plus loin). Ce facteur demeure sans aucun doute, le plus influant. Pourtant, l'appartenance à une catégorie d'établissement ou à une autre semble aussi assez modulatrice des attentes. De plus, cela semble agir, autant pour rehausser les attentes des plus démunis, que pour réduire un peu celles des plus favorisés. Notons que si les coefficients V de Cramer expriment la force des déséquilibres au sein de chaque catégorie d'école selon les déciles de NSE d'appartenance, cela revient à dire que, plus leurs valeurs sont réduites, plus les élèves d'un type donné d'établissement, montrent des déclarations semblables entre eux.

VARIABLES CONTEXTUELLES - NIVEAU ETABLISSEMENT

L'équipement matériel des établissements

Graphique B22

Proportion d'établissements équipés par type d'établissement (en %)
(5 612 établissements - 8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Les coefficients V de Cramer illustrent (entre 0 et 1) la force des déséquilibres et/ou de la relation entre la typologie de sept établissements et chaque variable d'équipement.

Le critère d'information pour construire les variables liées à l'équipement a été la réponse des élèves à la fréquence d'utilisation de cinq outils différents, dont une des alternatives considérait la réponse « il n'y en a pas ». Parmi les questions posées aux élèves, nous avons pris quatre outils qui nous semblaient potentiellement représentatifs et cohérents pour cette problématique.³⁴ Ainsi, si 50% ou plus d'élèves d'un établissement s'accordaient à déclarer qu'ils utilisaient les outils fournis par celui-ci – bibliothèque, ordinateurs, accès à Internet et laboratoire de sciences –, nous avons considéré l'établissement comme en étant équipé.

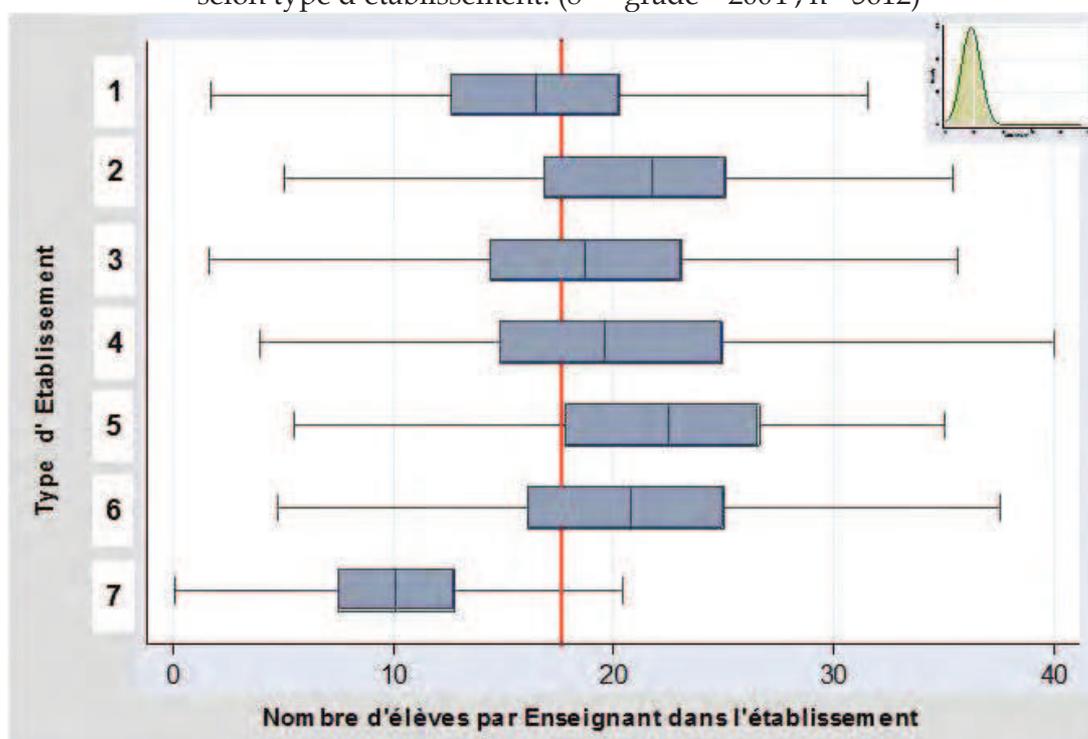
³⁴ Cohérents car la cinquième question sur si l'établissement possède ou non des calculatrices n'est guère informative à notre avis. Il se peut bien que les élèves eux-mêmes possèdent cet outil.

En revanche, dans le cas contraire nous l'avons classé comme non-équipé. Compte tenu des réalités illustrés sur le graphique B22, et des dispersions avérées entre types d'établissements, nous pensons que les outils bibliothèque et laboratoire de sciences nous parlent, d'une part, du degré d'équipement et des moyens disponibles pour un établissement donné, et d'autre part, de facteurs susceptibles d'influencer les acquis des élèves. Cependant, au lieu de construire un indice composite d'équipement, nous avons décidé de prendre un seul facteur à part entière : le laboratoire de sciences, car il pourrait refléter, par soi-même, un degré d'équipement global assez différenciateur entre écoles.

Le taux d'encadrement

Graphique B23

Distribution du taux d'encadrement (nombre d'élèves par enseignant) selon type d'établissement. (8^{ème} grade - 2004 ; n= 5612)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : La moyenne générale est indiquée par la ligne rouge centrale. L'allure de la courbe de distribution générale de cette variable est illustrée en haut à droite.

Une autre variable de contrôle, différente de la taille des classes, est le taux d'encadrement par établissement. Le graphique B23 illustre cette variable par catégorie d'établissement. La moyenne nationale est de 18 élèves par enseignement, avec un écart-type de 7. Ce chiffre arrive à 20 quand on ne considère que les établissements urbains. La différenciation par type d'établissement explique 18% de la variance de cette variable (R^2). Seulement les catégories public non-sélectif (taux=19) et privé payant (13) se détachent significativement du reste (≈ 22) d'après une comparaison simultanée de moyennes.

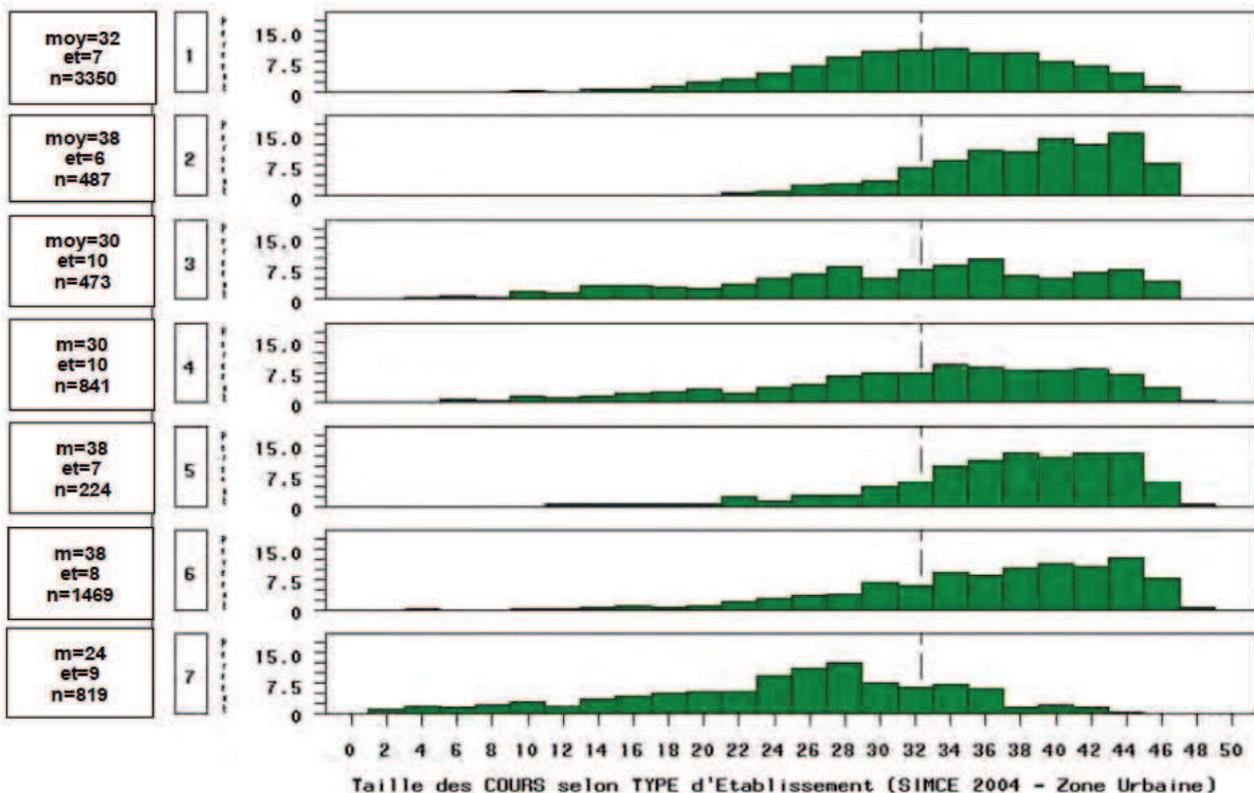
VARIABLES CONTEXTUELLES - NIVEAU CLASSE

Pour l'ensemble des 9 412 classes de 8^{ème} grade en 2004 (7 061 sur l'échantillon analysé), nous avons estimé différents facteurs permettant de les caractériser, notamment à partir des pratiques déclarées et des particularités des enseignants, mais aussi, des caractéristiques du public des classes. Notons, au passage, qu'à la fin du primaire chilien (8^{ème} grade) 11,62% (n=32354) d'élèves se trouvent scolarisés en zone rurale. Si l'on ne considère que les cours ayant 20 élèves ou plus, cette proportion passe à 8%.

La taille des cours

Graphique B24

Distribution de la taille des classes selon type d'établissement.
 Nombre de cours concernés par catégorie, taille moyenne et écart-type.
 (2004 ; 8^{ème} grade ; ZONE URBAINE / n=7663)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : La moyenne nationale pour la zone urbaine (32,4 élèves par cours) est indiquée par la ligne à traits. L'écart-type général est de 8,5 élèves. Le total de classes concernées correspond à 81,4%

Par rapport à la taille moyenne des classes, en 2004, nos estimations pour le 8^{ème} grade donnent 29,7 élèves par cours avec un écart-type de 10,4. Néanmoins, ces chiffres varient significativement selon l'emplacement géographique des écoles. Les établissements urbains, toutes catégories confondues, ont 32,4 élèves par classe (écart-type de 8,5) contre 18,3 (écart-type de 10,1) pour les établissements ruraux. Le graphique B24 montre les différences entre les courbes de distribution des sept sous-catégories d'établissements pour la zone urbaine. Le coefficient R^2 pour cette division est de 0,16.

Vision générale des caractéristiques des enseignants

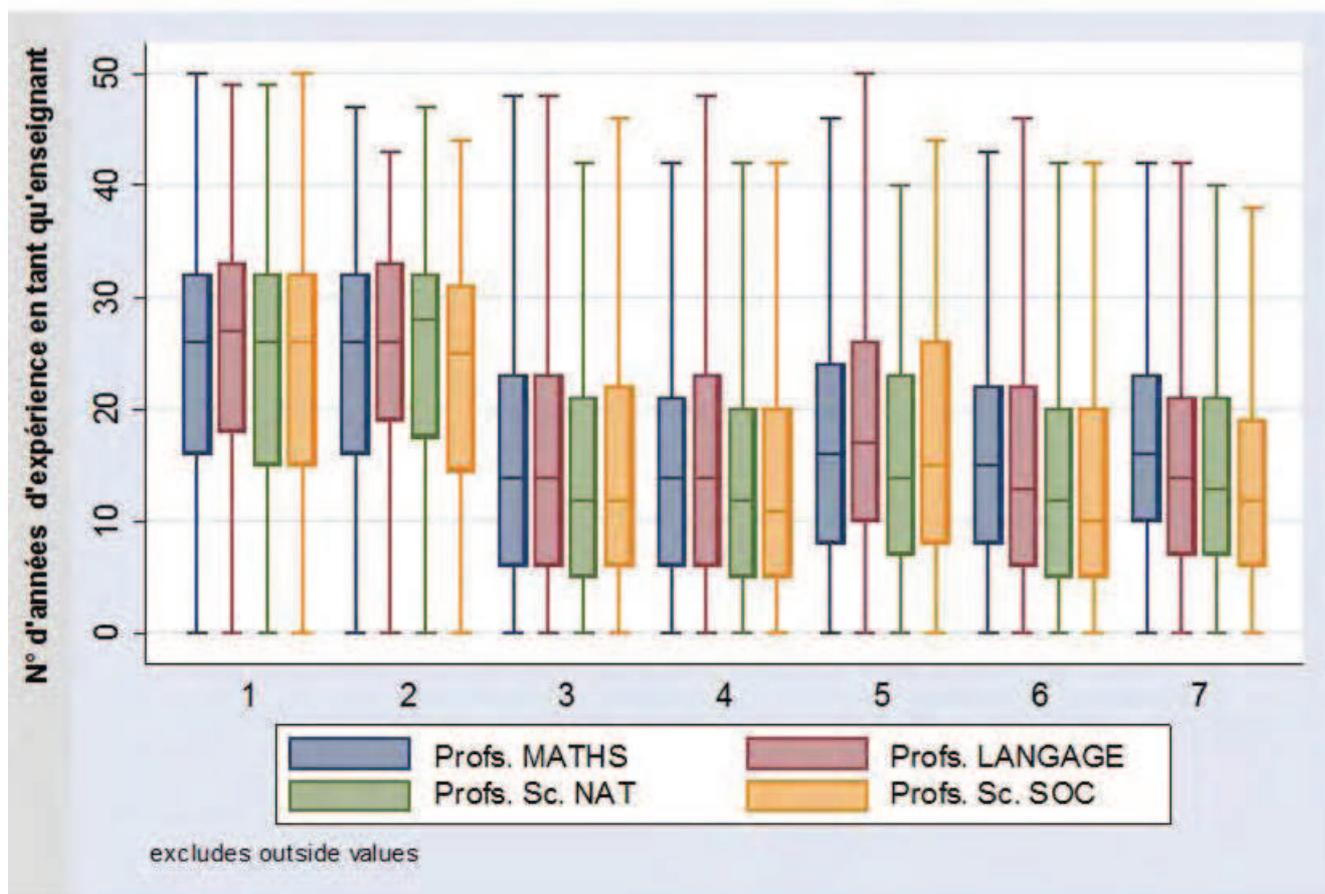
Les enseignants se caractérisent et diffèrent entre eux sur un ensemble de variables que nous avons essayé de saisir le plus soigneusement possible. Nous le présenterons ici souvent liées à notre catégorisation des établissements. En effet, nous ne pouvons qu'accepter l'idée de départ que ces caractéristiques ne sont pas aléatoirement distribuées entre les établissements, même au sein d'une catégorie donnée.

L'expérience des enseignants

Parmi les caractéristiques les plus significatives, nous avons considéré tout d'abord l'expérience, appréhendée à partir de la question directe « *combien d'années avez-vous travaillé comme enseignant ?* ». Ces réponses sont illustrées par le graphique B25. Il nous montre, sur une plage allant de zéro (première année de travail) jusqu'à cinquante années révolues, les distributions de l'expérience déclarée par les enseignants selon le type d'établissement et selon la discipline. Tout d'abord nous apprécions que la catégorie d'établissement l'emporte largement sur la matière concernée. Si bien il existe, chez tous les types d'établissement, une diversité d'enseignants vis-à-vis de leur expérience, les établissements publics (sélectifs ou pas) se distinguent pour avoir un corps enseignant significativement plus ancien en moyenne (entre 22 et 24 ans) et ceci pour toutes les disciplines. En général se distinguent, partout, trois groupes plus ou moins homogènes. D'abord, les établissements publics, puis le privé subventionné académiquement sélectif mais sans FP (entre 16 et 18 années en moyenne) et le privé payant (entre 13 et 17 années en moyenne) mais seulement en mathématiques et en sciences de la nature (14,9 et 16,7 années respectivement). Finalement, les types d'établissements restants ne se différencient pas significativement entre eux et ceci pour toutes les disciplines. Bien que cela dépasse nos objectifs, il faut noter que le régime contractuel des professeurs semble bien avoir un effet sur cette caractéristique du corps enseignant (cf. chapitre n°6). Conséquemment, nous trouvons aussi significativement plus d'enseignants issus des écoles normales dans ces établissements (entre 25% et 16%) quand on les compare avec tout le secteur privé (entre 11% et 1%), et notamment avec le privé payant (pas plus de 3%).

Graphique B25

Distributions autour de la médiane du nombre d'années d'expérience déclarées par les enseignants des quatre matières, selon type d'établissement
(n=9 412 classes de 8^{ème} grade - 2004)



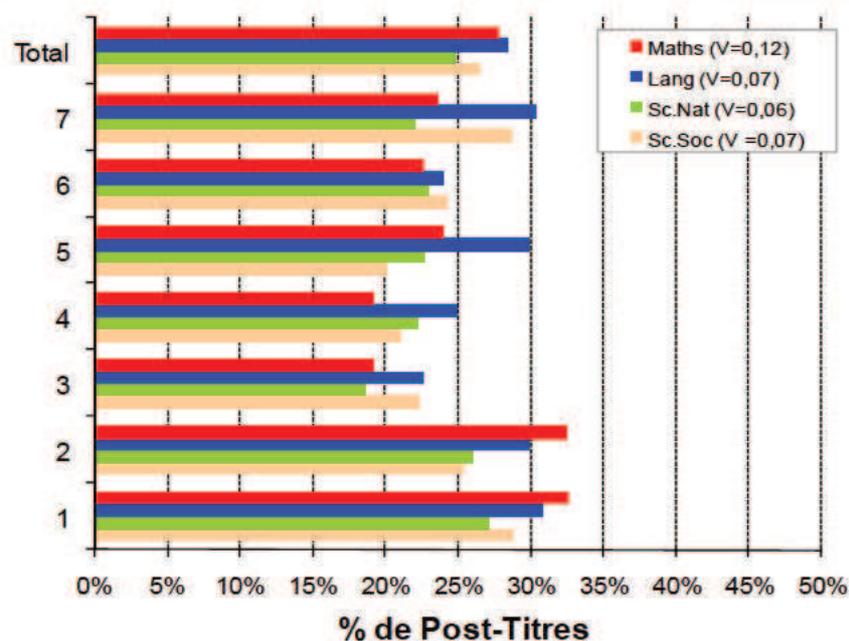
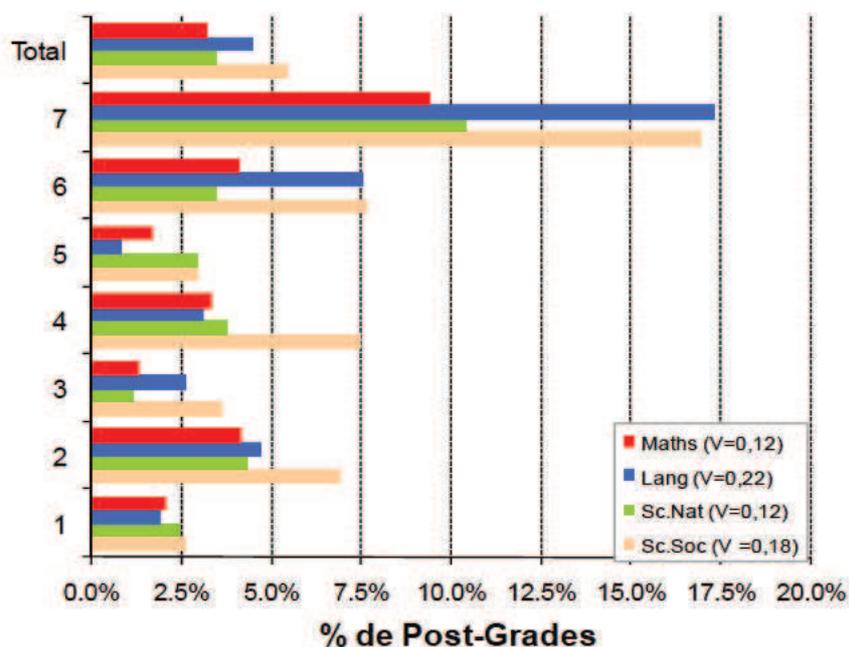
Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

La formation des enseignants

Sur nos données, nous avons rencontré tout d'abord que, pour chaque matière, près de 97% des enseignants déclarent avoir un diplôme pour exercer le métier, ce qui est concordant avec la loi nationale. Ce chiffre arrive à 99% si l'on considère ceux qui déclarent n'avoir pas encore fini leurs études de pédagogie (i.e. qu'ils obtiendront bientôt le diplôme). Par conséquent, nous avons décidé de nous concentrer sur les qualifications après l'obtention du diplôme, soit des post-grades, soit des post-titres ou qualifications dites professionnelles. De cette manière nous avons construit une catégorisation ordinaire à trois niveaux : sans formation spécialisée ; post-titre en éducation et/ou dans la discipline ; et enfin, post-grade en éducation et/ou dans la discipline.

Graphique B26

Proportion d'enseignants avec un Post-grade et un Post-titre par type d'établissement et selon discipline (n=9412 classes ; 8^{ème} grade ; 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

Le graphique B26 illustre les pourcentages d'enseignants possédant un post-grade et un post-titre. De toute évidence, les différences les plus frappantes se trouvent dans la

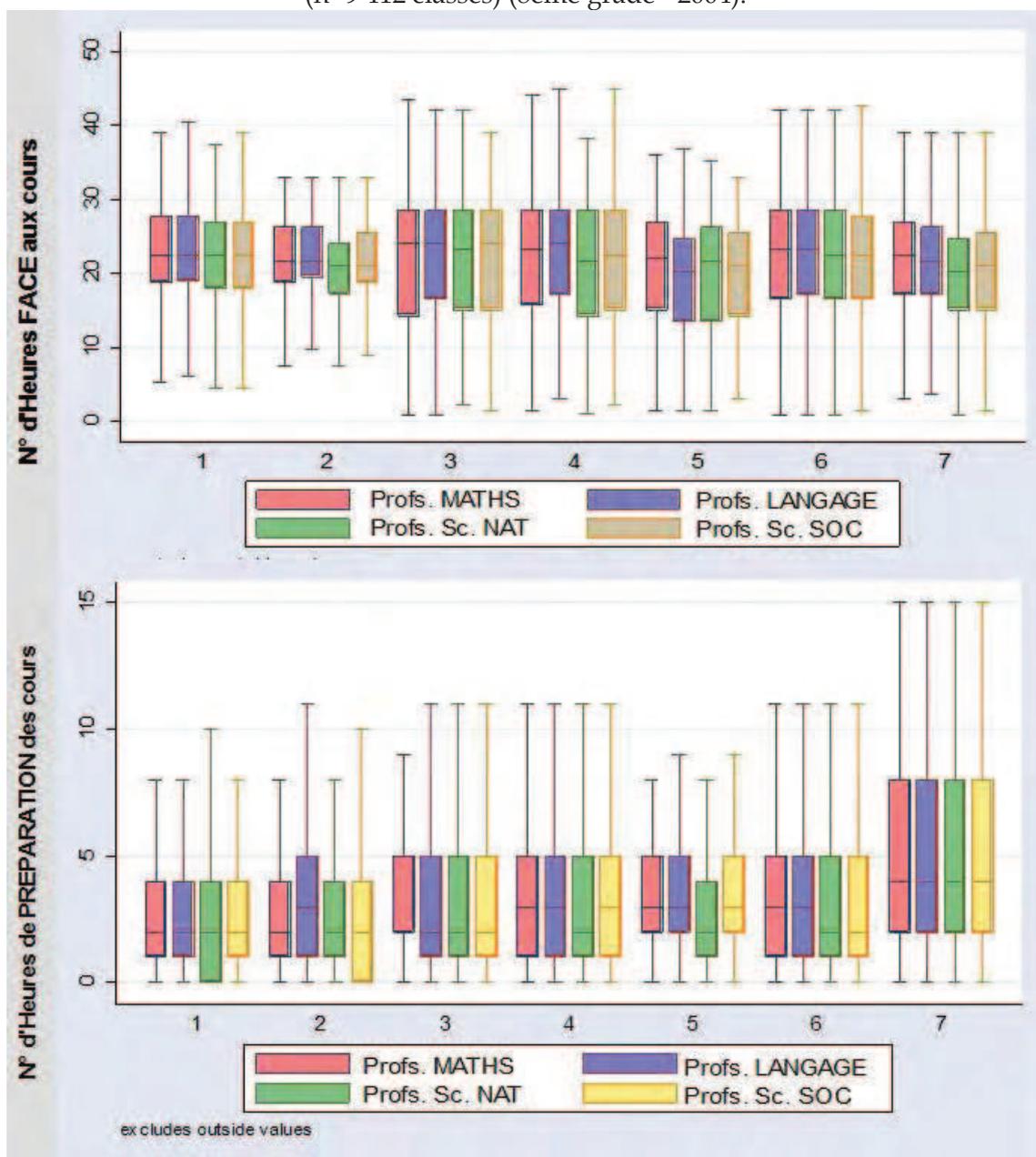
répartition des enseignants post-gradués et ceci pour toutes les disciplines mais spécialement en mathématiques et en langage. Tandis que le total des enseignants avec un post-grade (i.e. un Master et/ou un Doctorat) ne dépasse pas 5% des enseignants en maths (3,2%), sciences de la nature (3,5%) et langage (4,5%), et arrive à peine à 5,5% en sciences sociales, ces chiffres s'élèvent à 10%, pour toutes les disciplines – exception faite des maths avec 9,4% – chez le privé payant. A l'évidence, cette catégorie dépasse largement et significativement toutes les autres dans les quatre matières.

Le nombre d'heures déclarées pour la préparation des cours

Deux autres variables importantes seraient le nombre d'heures hebdomadaires déclarées dont disposent les enseignants pour préparer leurs cours, ainsi que le nombre d'heures par semaine en face d'une classe. Ces distributions sont illustrées par le graphique B27. On peut y apprécier, plutôt tout d'un coup, la dispersion et le niveau moyen de ces heures par type d'établissement. Nous pouvons ainsi voir que les heures en face des cours ne montrent pas des patrons tout à fait évidents. Toutefois, nous n'allons pas prendre cette dernière variable pour la suite de nos analyses, et nous allons nous concentrer sur les heures de préparation, susceptibles d'affecter davantage les acquis des élèves et pointées par certains travaux au Chili comme un phénomène problématique (cf. Llaña et Escudero, 2000 ; Corvalán, 2005 ; chapitre n°6).

Graphique B27

Distribution d'heures hebdomadaires déclarées comme étant « *En face des cours* » et de « *Préparation des cours* », par type d'établissement et selon discipline (n=9 412 classes) (8ème grade - 2004).

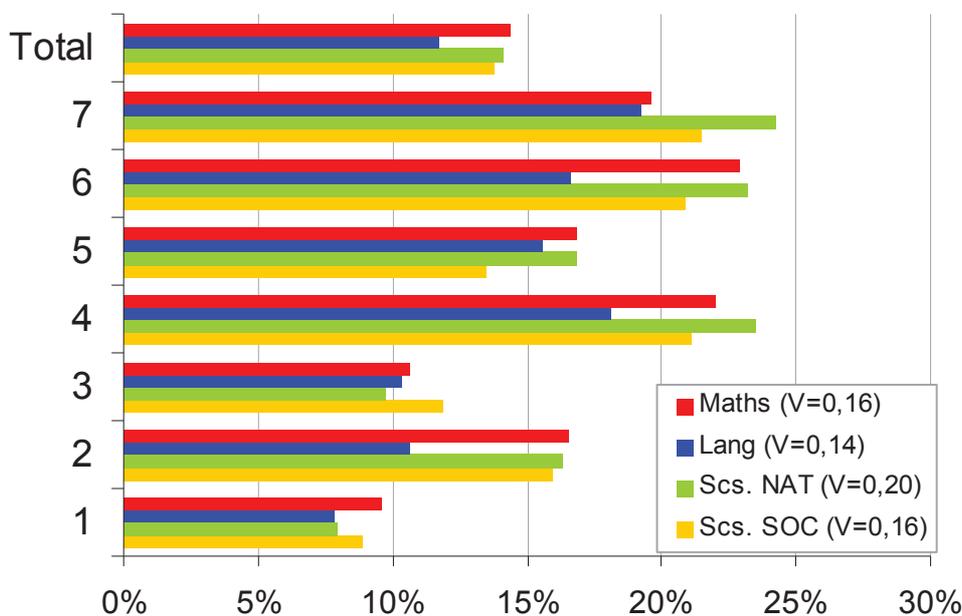


Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

Les enseignants ayant plus d'un travail et/ou travaillant dans plus d'un établissement

Graphique B28

Pourcentage d'enseignants ayant un autre travail différent à celui de l'établissement concerné. Par type d'établissement et selon discipline.
(n=9 412 classes) (8ème grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

Les croyances et perceptions des enseignants

Les préoccupations générales des enseignants (discipline, méthodes, contenus et résultats) par une approche quantitatif du phénomène

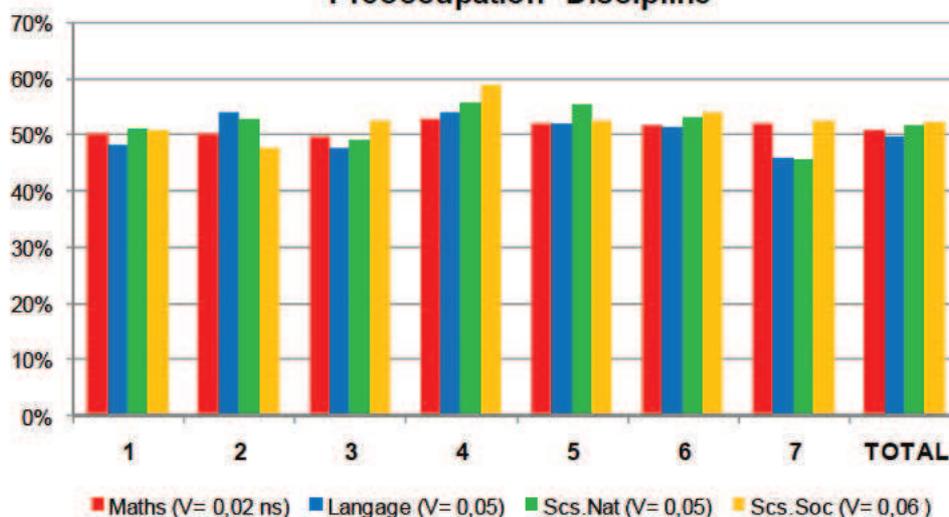
Nous avons considéré les déclarations des enseignants vis-à-vis des sujets les plus fréquemment traités dans des réunions formelles avec le personnel directif, les collègues, les parents et les élèves. A notre avis, ceci peut être pris comme une mesure indirecte des sujets les plus préoccupants pour chaque enseignant. Le graphique B29 nous présente la distribution des quatre sujets considérés, qualifiés ici d'une manière binaire. La question de base a été « *Quelles sont les thématiques les plus fréquentes traitées dans les réunions formelles avec les personnes suivantes ?* » Ils étaient invités à marquer deux sujets pour chaque cas parmi les options suivantes : Chef d'Unité Technico-Pédagogique ; Conseil des Professeurs ; Chef de Département ; D'autres enseignants de la discipline ; Parents

d'élèves ; Elèves. Les sujets étaient : Résultats académiques (Rendement) ; Discipline (i.e. Indiscipline) ; Contenus d'apprentissage ; Méthodologies d'enseignement/apprentissage ; Formation rapportée aux Valeurs.

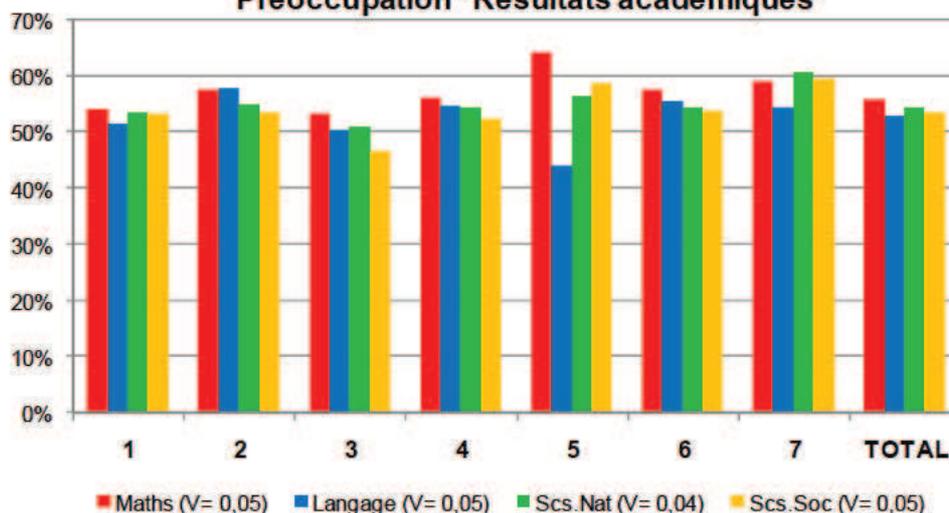
Graphique B29

Pourcentage d'enseignants/classes qui parlent formellement de discipline et de résultats académiques avec d'entités significantes à chaque égard.
Par matière et type d'établissement. (n=9 412 classes) (8^{ème} grade - 2004).

Préoccupation "Discipline"



Préoccupation "Résultats académiques"



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). La préoccupation est appréhendée dans chaque cas comme le fait de « parler dans des réunions formelles du sujet ... », avec les Parents d'élèves dans le cas de la discipline et avec le Conseil de Classe dans le cas des résultats académiques. Autrement dit, les % ici notés correspondent aux enseignants qui font ces respectives déclarations par rapport à une classe donnée.

Nous avons pris alors les cinq aspects proposés dans les questionnaires respectifs (i.e. discipline, méthodes, contenus, valeurs et résultats académiques) mais nous n'avons retenu que deux (discipline et résultats académiques). En effet, la perspective du phénomène qui a été retenue est celle de penser en termes qualitatifs les préoccupations signalées. C'est-à-dire, de choisir, pour nos deux sujets d'intérêt et par discipline, les personnes au sein de la communauté scolaire avec lesquelles le fait d'en parler représenterait davantage une préoccupation spéciale, plus poussée et/ou plus en mesure de refléter des véritables problèmes à cet égard.³⁵

Dans ce cadre, nous avons décidé que, étant données les corrélations existantes entre les personnes déclarées au sein des sujets (i.e. pour un sujet donné, quels étaient les couples de choix les plus reliés), nous avons pris celles qui étaient les plus proches, mais qu'en même temps, de par sa position dans la structure scolaire, représentaient un point particulier pour chaque sujet. En termes concrets cela se traduit par les faits suivants. Premièrement, la discipline des élèves a été appréhendée comme un sujet de préoccupation de la part d'un enseignant (dans chacune des quatre disciplines) qui déclare en parler aux parents.³⁶ Autrement dit, en parler aux les parents montre une tendance à le faire aussi avec d'autres acteurs. Bien entendu, notre choix privilégie le fait que les parents sont, en dernière instance, une ressource hors les murs de l'école vis-à-vis de ce phénomène. Deuxièmement, dans le cas de la préoccupation vis-à-vis des résultats académiques des élèves, nous avons décidé de prendre comme indicateur concret le fait d'en parler au sein du Conseil de professeurs. Etant donné qu'il s'agit d'une instance formelle, qui réunit tous les enseignants d'une classe, nous pensons que le fait d'en parler indique une préoccupation particulière.³⁷ Notons qu'un avantage relatif de ces indicateurs est qu'ils sont distribués plutôt de manière semblable parmi les différents types d'établissement (cf. graphique B30).

³⁵Conséquemment, un futur indice reliant l'opinion des enseignants de quatre disciplines et visant à cerner un climat dit général de chaque classe ne sera pertinent que pour explorer des variables allant au delà des résultats scolaires académiques. En effet, la mesure de l'appréciation de l'enseignant sur le climat de la classe à l'égard de ces deux facteurs (discipline et résultats) reste contingente aux matières. Et cela est d'autant plus valable que nous allons tester l'impact potentiel de ces facteurs sur les acquis scolaires. Il est clair, par exemple, que ces deux sujets restent fortement attachés aux caractéristiques de l'enseignant et à ses capacités à construire ce que la littérature indique comme la relation pédagogique concrète avec ses élèves (cf. Barr et Dreeben, 1983 ; Felouzis, 1994 ; Dubet et Martucelli, 1996 ; Duru-Bellat, 2001 ; Wilkinson et al., 2002). Bref, tout ce qui lui permet de « faire la classe » et plus largement, de participer comme protagoniste à l'aménagement d'un certain climat au sein de celle-ci. Enfin, du côté des élèves, la situation reste semblable : leurs comportements, dont se dérivent les préoccupations des enseignants que nous essayons de cerner, dépendent des rapports majoritaires vis-à-vis de chaque discipline et de chaque enseignant avec ses pratiques et caractéristiques.

³⁶Cette variable est, d'ailleurs, la plus corrélée avec le fait d'en parler avec les élèves (Tau de Kendall de 0,31 à 0,39, significatifs à $p < 0.001$; $n=7061$), mais elle est aussi bien corrélée avec toutes les autres possibilités. Notons, enfin, que toutes les corrélations entre choix, par sujet, sont presque les mêmes, autant pour notre échantillon de travail principal ($n=7061$), que pour le total des classes ($n=9412$).

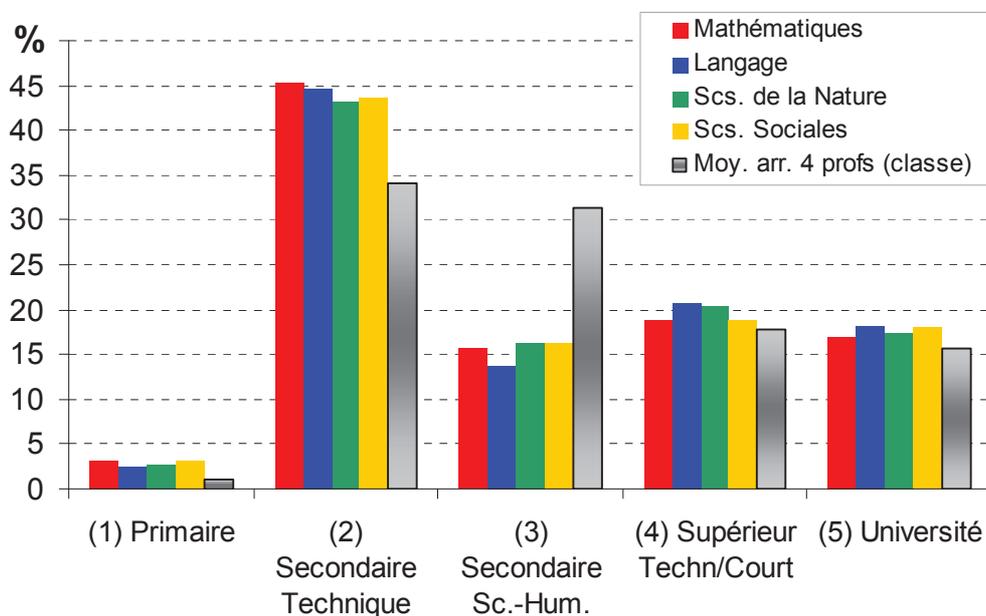
³⁷De plus, cette condition reste bien corrélée aussi avec le fait d'en parler aux parents et aux élèves du même sujet (coefficients Tau de Kendall autour de 0,21 et 0,19, significatifs à $p < 0.001$; $n=7061$). D'ailleurs, notons que certains acteurs tels que le chef de département, voire même les chefs d'UTP (Unité Technico-Pédagogique) n'existent pas chez tous les établissements.

Les attentes des enseignants vis-à-vis de leurs élèves

Graphique B30

Attentes des enseignants selon la discipline par rapport au niveau d'étude le plus haut qu'ils croient que la plupart de leurs élèves attendront, par classe.

(n=9412 classes) (8^{ème} grade - 2004).

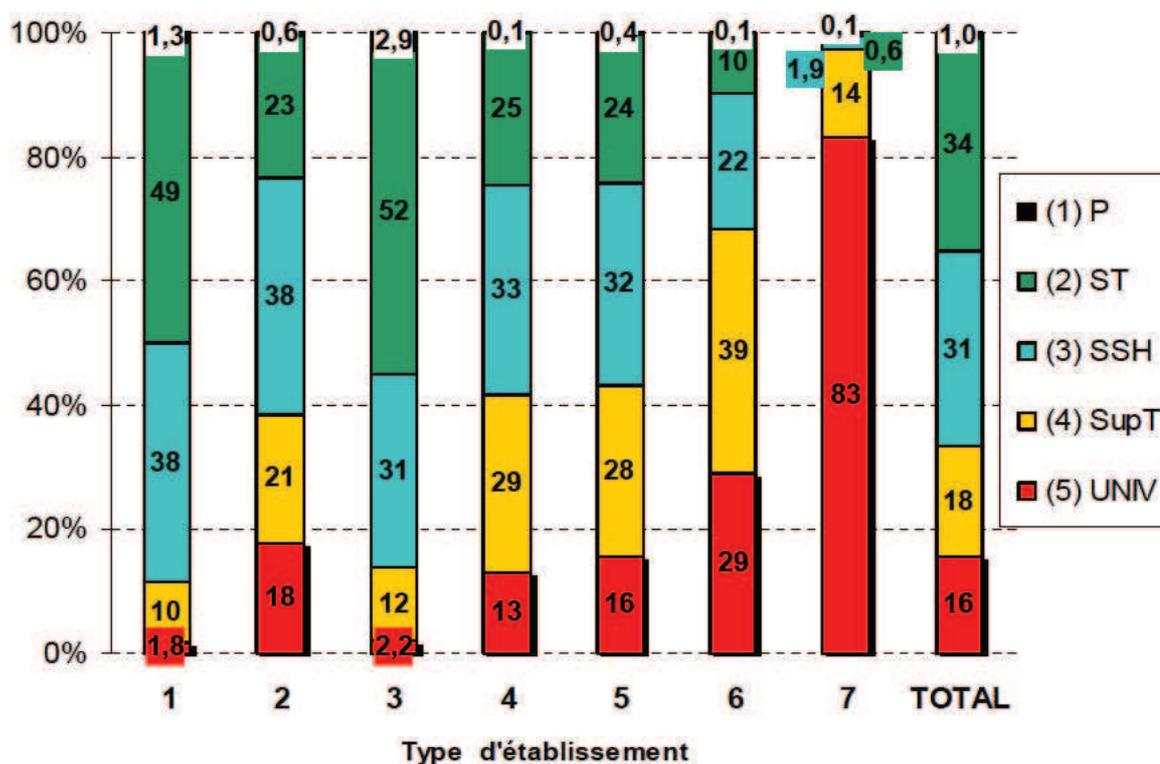


Source : élaboration de l'auteur. Lecture : près de 45% d'enseignants en mathématiques croient que leurs élèves attendront un niveau d'études équivalent au secondaire technique. Note : la dernière barre de l'histogramme correspond à la moyenne arrondie de l'opinion des enseignants des quatre disciplines pour chaque classe.

Sur le graphique B30 ci-dessus nous pouvons apprécier, d'abord, que les opinions des professeurs tendent à être très semblables d'une discipline à l'autre. Les corrélations entre opinions, d'une discipline à l'autre, sont toutes importantes et significatives (r de Spearman entre 0,65 et 0,66). Remarquons tout de même que les enseignants de mathématiques semblent avoir tendance à être plus sévères vis-à-vis de leurs attentes déclarées, tandis que les enseignants de langage seraient légèrement plus souples, sans pour autant manifester des différences très significatives. Le graphique B30 illustre aussi la moyenne arrondie des opinions des quatre enseignants pour chaque classe. Elle reflète le degré d'unanimité et de diversité des attentes des enseignants face à chaque classe. Si ces opinions tendent à être concordantes, notons qu'il y a un déplacement vers le haut au moment de considérer toutes les opinions ensemble. En effet, seulement près d'un tiers (35,2%) des classes restent dans les deux positions les plus basses tandis que 66% des opinions moyennes s'inclinent pour penser que leurs élèves arriveront au moins à finir le secondaire non-technique, dont près de 15% pensent qu'ils iront à l'université. Bien que nous ayons eu déjà un aperçu sur le graphique B20, un regard sur l'ensemble de cette opinion composite selon le type d'établissement est présenté sur le graphique B31 (i.e. moyenne arrondie sur cinq niveaux, par classe).

Graphique B31

Moyenne arrondie des attentes des enseignants (toutes disciplines confondues par classe) selon type d'établissement. Exprimée en pourcentages par niveau et de manière descendante. (n=9 412 classes) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture Niveau Scolaire : (P) Primaire ; (ST) Secondaire Technique ; (SSH) Secondaire Scientifique-Humaniste ; (SupT) Supérieur Technique et/ou Court ; (UNIV) Université.

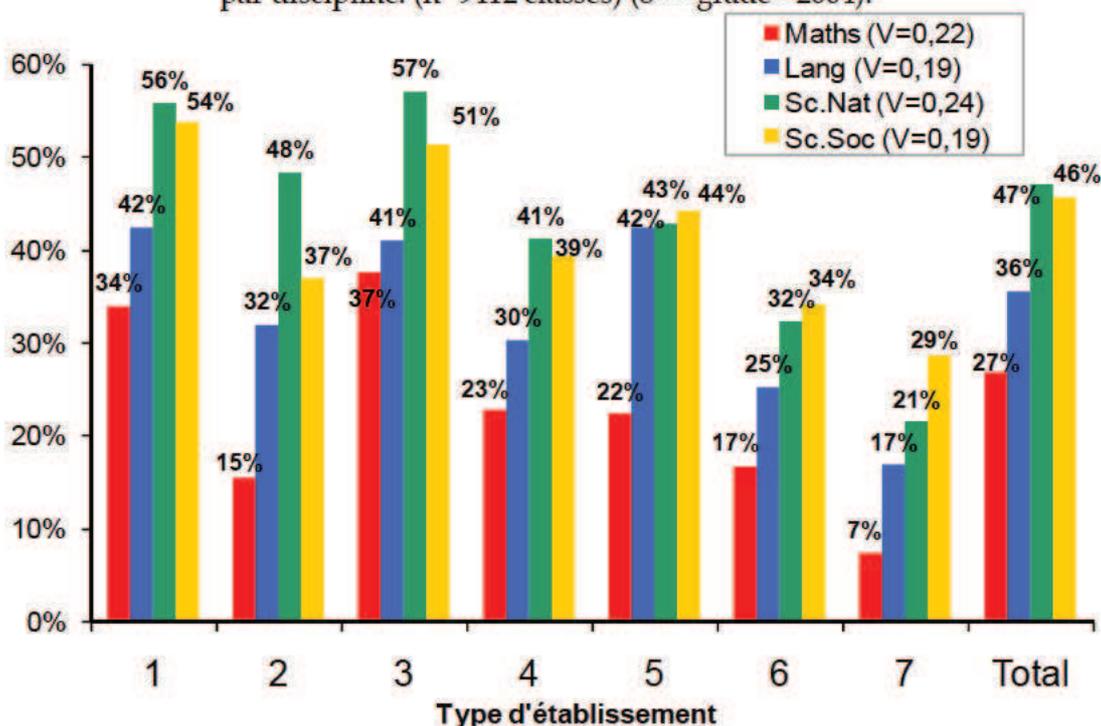
L'opinion des enseignants sur leur propre préparation

La préparation des enseignants pour enseigner le contenu des programmes a été mesurée et construite à partir des déclarations réalisées sur une échelle ordinale à trois niveaux (« pas assez préparé » ; « préparé » ; « très bien préparé »). Ceci en répondant à la question : « Comment vous vous sentez préparé pour enseigner les contenus suivants de 8^{ème} grade » (i.e. le programme officiel). Pour chacune des quatre matières chaque professeur a été invité à se prononcer sur un nombre variable des contenus, spécifiquement identifiés et nommés. S'il y avait 22 items en mathématiques, il n'y en avait que 12 en langage et sciences de la nature, tandis qu'il n'y en avait que 5 en sciences sociales. Bien entendu, les différences par rapport au nombre d'items de contenu appellent à une certaine prudence dans la manière de saisir ces réponses pour construire nos variables. En effet, plus important est le nombre d'items, plus performante et fine peut être la saisie du problème concernant comment les

enseignants se sentent préparés. Ainsi, nous avons décidé de maintenir l'échelle à trois niveaux et de la prendre, momentanément, comme une échelle continue, nous permettant d'estimer la moyenne (et alternativement la somme) de chaque enseignant face à l'ensemble d'items demandés. Après, nous avons fixé deux seuils de coupure en valeur absolue et non pas, par exemple, en proportions relatives du type 33% ou 25% le plus bas, haut, etc. Le seuil inférieur (le plus important pour nous) tient compte de fait qu'un professeur dont la moyenne des réponses est inférieure à 2,49 sur une échelle de 1 (bas) à 3 (haut) ne peut pas être tenu comme en situation de bonne préparation auto-perçue. Certes, cela reste arbitraire, mais pour l'instant, nous pensons que ce critère est adéquat, notamment si l'on considère la tendance à sur-déclarer et/ou à acquiescer. Bref, nous avons préféré fixer un seuil plutôt exigeant. En ce qui concerne le seuil supérieur, nous avons considéré que tous ceux dont la moyenne des réponses dépassait 2,75 pouvaient être pris comme exprimant un niveau de « haute préparation » auto-perçue.

Graphique B32

Proportion d'enseignants (en %) déclarant avoir une préparation plutôt basse pour enseigner les contenus de 8^{ème} grade de leur discipline. Selon type d'établissement et par discipline. (n=9412 classes) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

Le graphique B32 montre les pourcentages, par discipline et selon type d'établissement, situés sous le seuil de 2,49 (entre 1 et 3). Notons d'abord que, pour les sciences de la nature et sociales ces proportions restent, au total, assez élevées frôlant la moitié des enseignants (respectivement, 47% et 46%). En langage cette proportion est de 36% et en mathématiques elle descend à près d'un quart (26%). Du côté des catégories d'établissement, il existe des différences significatives entre les écoles gratuites et non sélectives (privé et public, n°1 et n°3) et tout le reste. En revanche, le public sélectif s'écarte de cette situation avec des proportions beaucoup plus faibles et, sauf en sciences de la nature, en dessous des pourcentages du total. Le graphique B32 nous permet de voir que les enseignants se perçoivent eux-mêmes très différemment selon le type d'établissement. Il serait même possible de penser à l'existence d'une certaine échelle de prestige.

Nous pourrions postuler alors que cette variable permet d'approcher la réalité du marché de travail en termes de l'embauche des maîtres de par leur qualité, ce qui reste, toutefois, discutable. Sous ce point de vue, les meilleurs enseignants préféreraient aller dans des établissements à public social et/ou académiquement trié. D'ailleurs, ils pourraient choisir où aller de fait d'être bien côtés, quoi que notre mesure ne soit pas du tout objective. Autrement dit, soit, cette variable s'approche de la qualité réelle des enseignants en termes de maîtrise des contenus disciplinaires ; soit, elle montre une perception subjective auto-réalisatrice du type « je travaille ici, donc, je suis bon » ; soit, les deux situations à la fois. Hélas, nous ne pouvons pas pour l'instant bien démêler cette problématique. De toute façon nous resterons dans l'idée que, dans une certaine mesure, cette variable capture la qualité des enseignants en termes de maîtrise relative des contenus. Nous ne pouvons donc pas savoir à quel point un enseignant est meilleur qu'un autre mais nous pouvons quand même les classer dans des niveaux qui ne devraient pas s'éloigner trop de la réalité.

Certainement cette variable – l'auto perception de préparation – pourrait être liée à l'âge/expérience des enseignants et à leur formation. Effectivement, pour les quatre matières les professeurs qui se perçoivent eux-mêmes comme ayant une préparation plutôt basse se différencient des autres, de manière légère, étant significativement plus âgés. Les deux niveaux de préparation restants ne se différencient pas entre eux vis-à-vis de l'expérience/âge. De son côté, la formation (avoir un post-titre ou un post-grade) joue aussi un rôle différenciateur vis-à-vis du niveau de préparation dans les quatre disciplines. Comme il était possible de s'attendre, les enseignants avec un diplôme de post-titre ou de post-grade se sentent en moyenne davantage préparés que les autres, et même entre eux à la faveur des post-gradués. Remarquons quand même, que ce rapport est moins fort en maths (R^2 de 0,7%) que pour les restantes trois disciplines (R^2 de 2,1% à 2,6%) car il reste davantage associé au lieu de travail (type d'établissement) de ces enseignants. De même, l'expérience ne semble pas si importante pour expliquer la préparation quand on contrôle la formation et le type d'établissement. En définitive, des modèles considérant à la fois l'expérience et la formation montrent une association significative entre le type

d'établissement – négative pour le privé subventionné et le public non-sélectifs, et positive pour tout le reste – et le niveau de préparation déclaré pour les quatre matières. Cela d'autant plus qu'il s'agit des établissements sélectifs et où l'on paye. De plus, ces rapports ne changent pas si l'on ajoute la variable diplôme (i.e. anciennes écoles normales v/s universitaires). Au total, il n'est pas possible de passer à côté du fait que les établissements qui accueillent une population à plus faible capital culturel sont ceux qui accueillent aussi davantage de professeurs se sentant moins préparés. Cette situation peut faire partie des facteurs explicatifs des différences importantes entre moyennes des scores aux tests. Au total, il faut retenir que les élèves des catégories d'établissements plus performantes en moyenne, jouissent des enseignants qui déclarent se sentir plus préparés que les autres, résultats en phase avec la littérature (cf. chapitre n°1).

Les pratiques des enseignants

Les contacts entre collègues

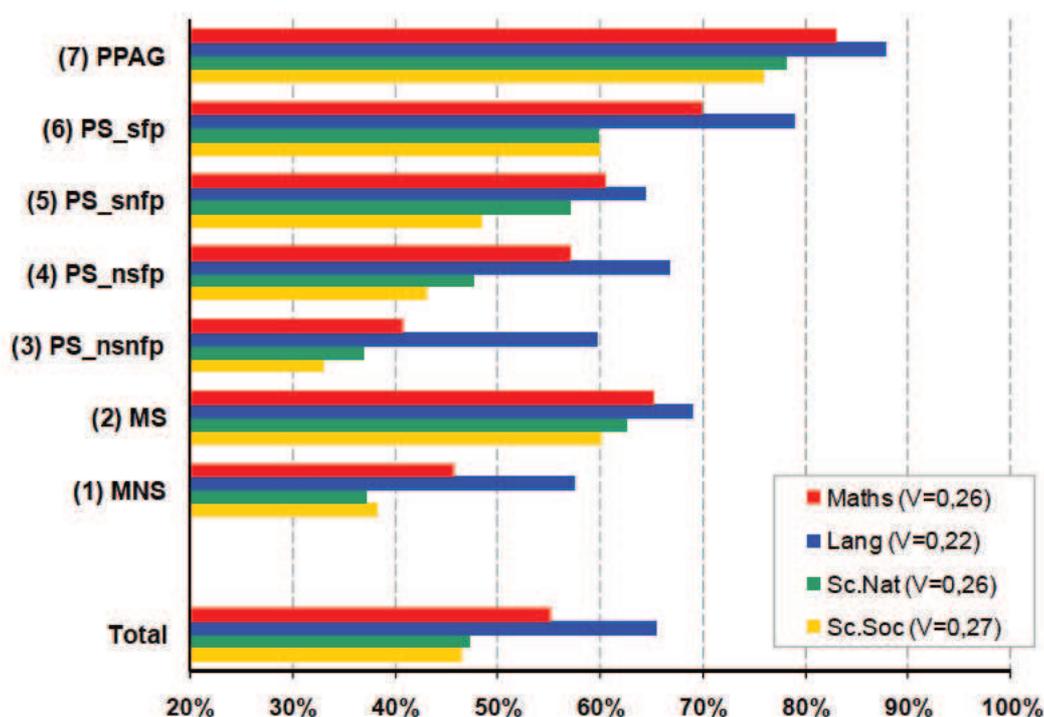
Une autre variable qui nécessite d'être illustrée tient à une approximation du travail en équipe de la part des enseignants par discipline. Nous pouvons, en fait, espérer même qu'elles aient une relation avec les apprentissages des élèves (e.g. Opdenakker et Van Damme, 2000). Bien entendu, cela touche des paramètres tels que la socialisation professionnelle ou le prestige relatif des disciplines, mais aussi l'organisation générale des écoles. En effet, certains établissements possèdent des départements par discipline bien structurés, tandis que d'autres pas. Nous avons recueilli, pour chaque déclaration d'enseignant (i.e. par classe) les réponses sur la fréquence des réunions formelles avec différentes personnes ou responsables au sein de chaque établissement. La question posée était : « *Dans cet établissement, avec quelle fréquence avez-vous des réunions formelles avec les personnes suivantes ? : Chef d'établissement ; Chef d'Unité Technico-Pédagogique ; Conseil des Professeurs ; Chef de Département ; D'autres enseignants de la discipline ; Parents d'élèves* ». Les alternatives de réponse étaient : « *Jamais ; Une fois par an ; Une fois au semestre ou trimestre ; Deux fois ou plus au semestre ou trimestre* ».

Nous avons décidé de considérer la fréquence déclarée des rapports formels entre collègues d'une même discipline, pour nous approcher ainsi du degré de travail en commun entre eux. En effet, nous pouvons supposer que chaque enseignant s'approche des collègues pour mettre au point des sujets tels que l'avancée et la profondeur du programme, les évaluations, les matériels pédagogiques, la jonction entre grades, etc. Cette variable nous rapproche donc des mécanismes, qui au sein de chaque établissement et par discipline, sont mis en œuvre pour coordonner et réaliser les tâches d'enseignement. On pourrait penser que cela nous parle aussi, d'un certain esprit de travail. De cette façon, il est possible de supposer que ces mécanismes – même si inconnus *stricto sensu* – sont moins

présents au fur et à mesure que l'on réalise moins de réunions avec les collègues pendant l'année.

Graphique B33

Pourcentage d'enseignants déclarant avoir deux ou plus de deux réunions formelles au semestre/trimestre avec leurs collègues de la même discipline au sein de leur établissement. Selon type d'établissement et par discipline.
(n=9 412 classes) (8ème grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC ou MUNICIPAL Non Sélectif ; (2) PUBLIC ou MUNICIPAL Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). NOTE : il s'agit des classes et non des individus, donc, ces chiffres s'approchent de la réalité mais ils ne la représentent pas fidèlement.

Le graphique B33 présente une vision de la quantité de contact entre enseignants par discipline, selon le type d'établissement et au total. Ceci considérant le pourcentage déclarant réaliser « deux ou plus de deux réunions pendant le trimestre/semestre ». Nous pouvons y apprécier des différences, autant du côté des disciplines que du côté des catégories d'établissement. Tandis que le total d'enseignants en sciences sociales et en sciences de la nature semblent moins disposés à la haute fréquence des réunions entre collègues (47% pour les deux cas), les enseignants de langage s'échappent vers le haut avec près de deux tiers dans cette catégorie (65%). Ceux de mathématiques ne sont que 55% dans ce niveau. Ces différences assez marquées pourraient aussi être en train de refléter une utilisation plus ou moins efficace du mécanisme de travail en commun liée aux traditions disciplinaires, par exemple, dans le cas des matières souvent prises comme secondaires (e.g. sciences de la nature et sciences sociales). Autrement dit, peu de réunions

dans l'année suffiraient à bien coordonner le travail, ou bien, cela reflète une moindre disposition à l'organisation du travail enseignant (i.e. plus d'individualisme). Toujours est-il que l'hypothèse opérationnelle d'un usage plus efficace des réunions, et qui de ce fait justifierait un nombre réduit de celles-ci, reste non-testable pour l'instant.³⁸ D'un point de vue des catégories d'établissement, les différences entre disciplines semblent se reproduire. Pourtant, les établissements à profil sélectif et/ou payant ont tendance à concentrer davantage de réponses dans la modalité haute des réunions formelles entre collègues.

La couverture des programmes par classe et par discipline

Une variable assez importante dans la littérature correspond aux différences d'une classe à l'autre et entre établissements, rapportées à la couverture des programmes scolaires officiels par discipline.³⁹ En effet, cette variable touche au cœur du problème des opportunités d'apprentissage offertes aux étudiants au sein d'un même système scolaire, d'une même commune ou académie, voire, d'un même établissement (cf. chapitre n°1). D'ailleurs, cette problématique ne semble pas éloignée du cas chilien (cf. chapitre n°6).

Dans l'enquête sur laquelle nous travaillons, les enseignants ont été invités à répondre par classe et par discipline – presque à la fin de l'année scolaire –, sur le niveau d'aboutissement dans la passation des contenus officiels respectifs correspondants au 8^{ème} grade en 2004. Ainsi, les réponses à la question « *Parmi les contenus suivants, combien avez-vous enseigné aux élèves de ce cours ?* », ont considéré trois niveaux – (3) « *complètement* » ; (2) « *partiellement* » et (1) « *je ne l'ai pas enseigné* » –. Ceci pour un ensemble de propositions extraites des programmes officiels du Ministère de l'éducation considérant entre 22 et 5 items selon la discipline. Malgré ces dissimilarités, nous allons essayer d'en extraire des variables cohérentes. Le graphique B34 illustre, pour chaque matière, la somme des réponses, en trois niveaux, pour l'ensemble des classes et après remplissage des données manquantes.⁴⁰

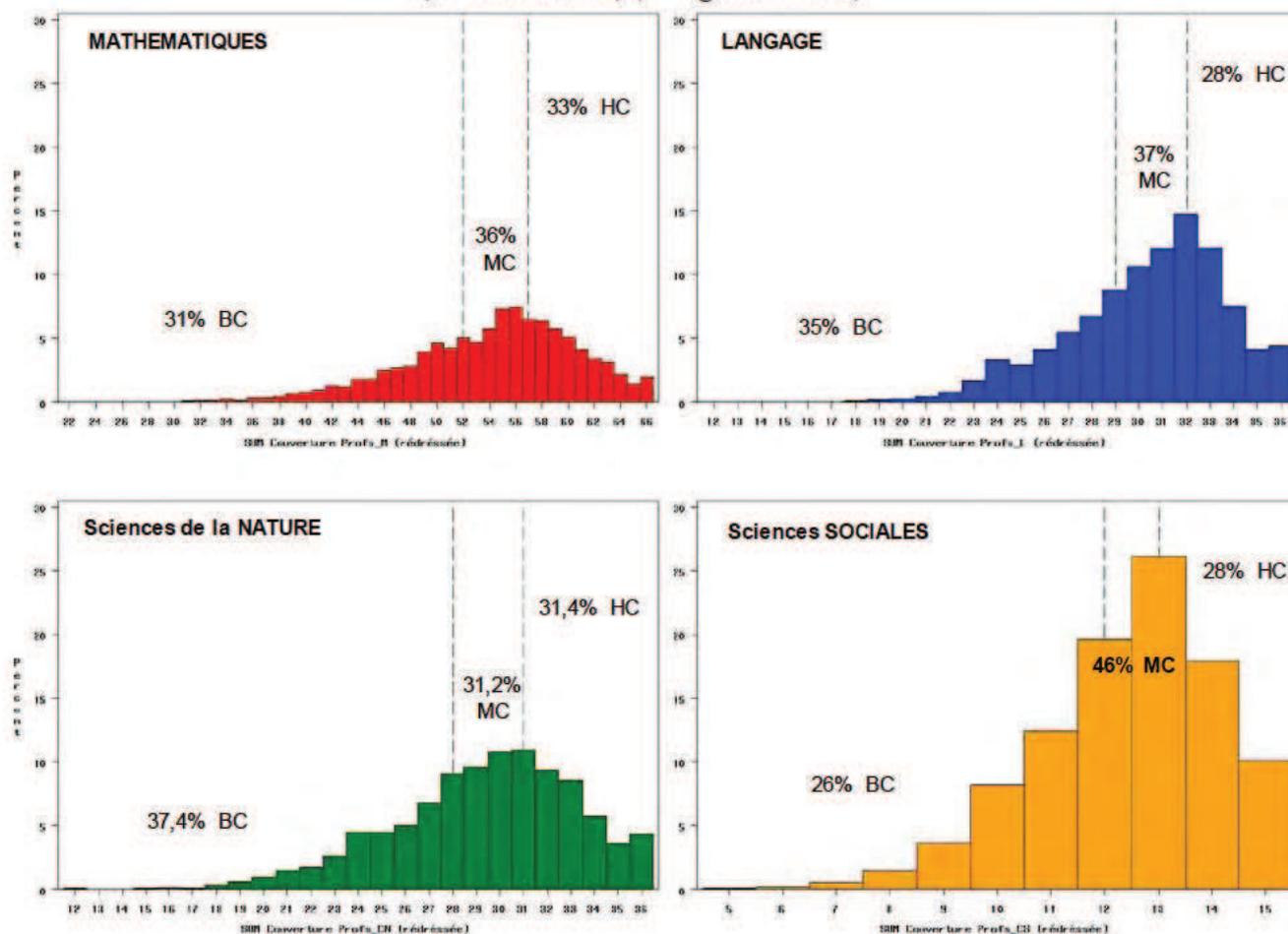
³⁸ Notons que le niveau le plus bas (« *jamais de réunion* ») comptabilise autour de 26% des répondants en sciences sociales et en sciences de la nature. A l'opposé, ils ne sont que 16% et 18% parmi les professeurs de langage et mathématiques, respectivement.

³⁹ Décret du Ministère d'Éducation n°232 du 2002 sur les objectifs fondamentaux et les contenus minimaux et obligatoires du niveau d'enseignement de base. Notons tout de même que la loi chilienne permet aux établissements de présenter des programmes spécifiques et propres.

⁴⁰ La proportion d'enseignants ayant répondu 100% des contenus présentés était de 79,7% en maths (22 items), 85,2% en langage (12 items), 82,4% en sciences de la nature (12 items), et de 90,1% en sciences sociales (5 items). Ces proportions augmentent jusqu'à 95% si l'on raisonne en termes de ceux qui ont répondu au moins deux tiers des propositions faites. La proportion de ceux qui n'ont rien répondu varie entre 3% et 4%. Nous avons donc entrepris un remplissage des données manquantes considérant à la fois de multiples critères. Tout d'abord, en nous appuyant sur les réponses des élèves vis-à-vis aussi de la couverture des contenus généraux (comme opinion moyenne par classe). Ceci, selon le degré d'unanimité de leurs réponses (i.e. un CV, ou coefficient de variation, de la classe \leq à la valeur du 3^{ème} quartile de l'ensemble des CV vérifiés pour les 9 412 classes, était considéré comme « unanime »). Puis, en considérant les rapports par discipline entre les réponses face à chaque proposition (i.e. face à chaque contenu) et un ensemble d'autres

Graphique B34

Histogrammes représentant la somme des réponses par discipline vis-à-vis de la couverture des contenus des programmes scolaires, ainsi que la division et la proportion de chacune pour Basse, Moyenne et Haute Couverture. (n=9412 classes) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : (HC) Haute Couverture ; (MC) Couverture moyenne ; (BC) Basse couverture. Chaque histogramme correspond à la somme des niveaux de couverture par sujet (totale=3, partielle=2, non-vu=1) déclarés par les enseignants face à l'ensemble de propositions faites reflétant les contenus du programme.

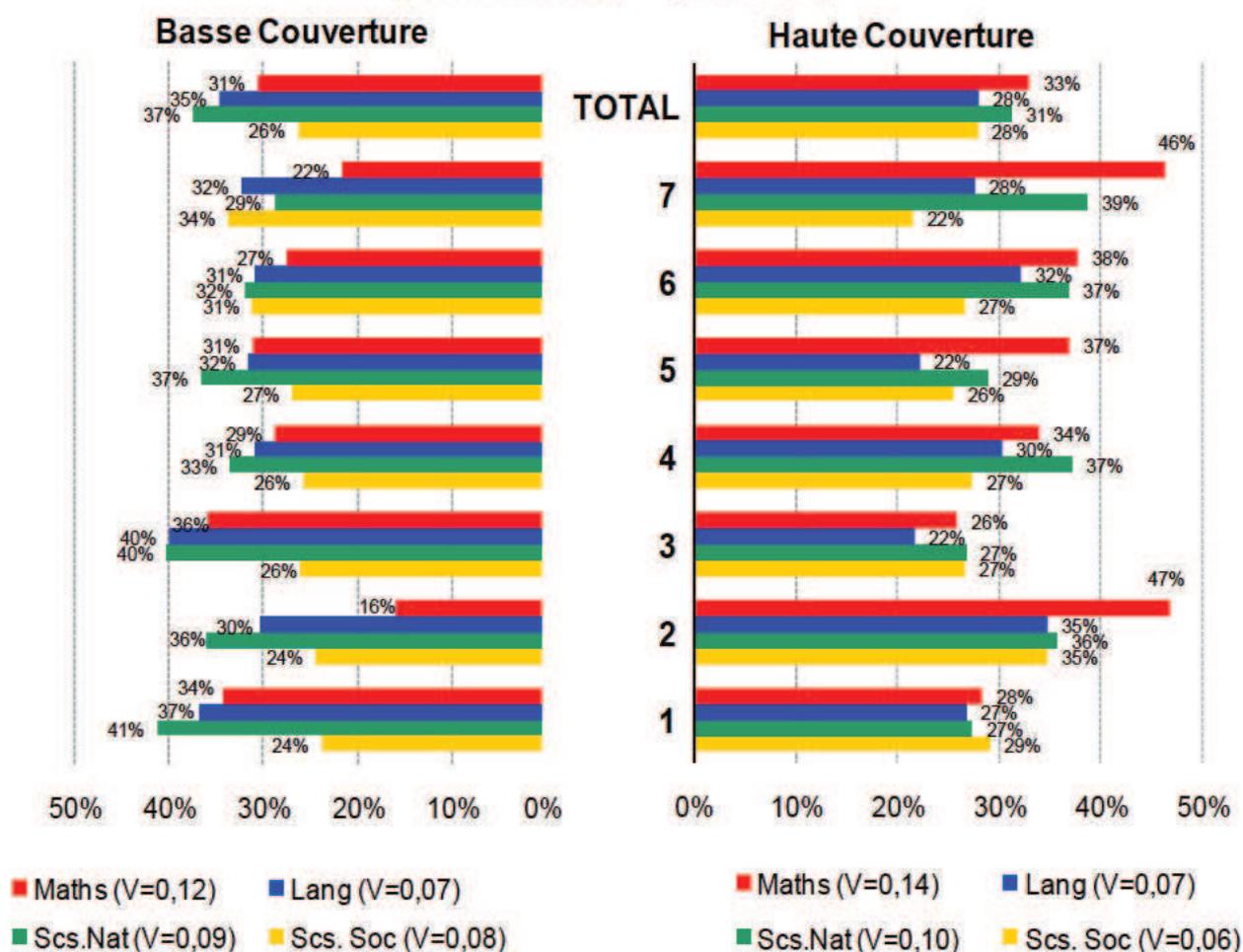
Certainement, il serait possible de considérer des critères absolus pour définir les niveaux de couverture réussis par chaque classe et selon la discipline (par exemple : tous ceux qui seraient en bas de la valeur-niveau « 2 » se placeraient en basse couverture). Pourtant, le

variables extraites des réponses des enseignants eux-mêmes ainsi que des caractéristiques associées aux établissements. Ainsi, nous avons construit 102 analyses de régression (51 avec les opinions « unanimes » des élèves et 51 pour les non-unanimes). Ces régressions ont été faites à partir des groupes d'enseignants ayant tout répondu. Enfin, avec les liens trouvés (i.e. coefficients), nous avons procédé à remplir certains segments des données manquantes, notamment ceux qui avaient des lacunes importantes ou qui n'avaient rien répondu. D'autres segments ont été pénalisés par supposition de basse couverture (i.e. les non-réponses s'interprétaient dans ces cas, comme s'il s'agissait du niveau « je ne l'ai pas enseigné », notamment lorsque la question a été plutôt bien remplie par l'enseignant). D'autres, encore, ont été simplement remplis avec les valeurs des moyennes générales. Notons qu'une procédure semblable a été appliquée dans le cas des items rapportés à la préparation déclarée des enseignants, notée plus haut.

critère des trois tiers approximatifs, mais relatifs entre eux, nous semble plus juste car il reflète en quelque sorte l'état de la couverture des programmes sur l'ensemble du pays, pour l'année et le grade concernés. Autrement dit, la question derrière cette problématique serait davantage le « pour quoi certaines classes arrivent-elles, d'après la déclaration des enseignants, à couvrir une partie plus importante des programmes tandis que d'autres n'arrivent pas à en faire autant ». Si cette question peut se voir en termes absolus comme dans d'autres cas (i.e. le niveau de préparation), cette fois ci, elle devient plus satisfaite si l'on considère la réalité de l'ensemble (illustrée par les histogrammes à allure de tendance normale sur le graphique B34). De plus, les déclarations des enseignants ne sont qu'une mesure relative du problème de chaque enseignant lui-même.

Graphique B35

Proportion des classes, par discipline et selon type d'établissement, en situation de Basse et de Haute couverture des programmes scolaires.
(n=9412 classes) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

Rappelons que notre intérêt porte fondamentalement sur l'impact de la couverture sur les apprentissages des élèves. Le caractère plus ou moins complet de la couverture des programmes scolaires, à la lumière de nos catégories d'établissements, est illustré sur le graphique B35, avec les proportions en basse et en haute couverture, par discipline et pour chaque type d'école. A première vue, il n'est pas aisé de distinguer un patron unique ni par type d'école, ni par discipline. Pourtant, il existe des différences significatives et importantes dans certains cas, notamment en mathématiques (V de Cramer de 0,12 pour la BC et de 0,14 pour la HC). S'il n'est pas facile de dégager une tendance incontestable et nette pour toutes les disciplines concernées en faveur d'un type particulier d'établissement, nous voyons qu'en mathématiques, en langage et dans une moindre mesure en sciences de la nature, nous pouvons apprécier qu'en termes relatifs, il existe de plus fortes proportions de classes en basse couverture chez les établissements gratuits et non-sélectifs. En même temps, la situation inverse s'avère vraie dans le cas de la haute couverture. Pour ces mêmes disciplines, les établissements privés payants et notamment la catégorie public sélectif, s'installent à l'opposé avec une meilleure couverture générale, c'est-à-dire, avec moins de classes en situation défailante et plus en situation de haute couverture, et au pire, avec des situations équivalentes à la moyenne nationale. La seule discipline qui échappe à ce patron est les sciences sociales, où les deux catégories d'établissements publics (n°1 et n°2) obtiennent des meilleures performances en couverture que le privé payant (n°7) et sa plus proche catégorie, le privé subventionné avec FP et sélectif (n°6).⁴¹ Finalement, les établissements sélectifs autant d'un point de vue académique que socio-économique, sembleraient avoir tendance à avancer davantage dans la passation des programmes scolaires. Ils offriraient, de ce fait, de meilleures opportunités d'apprentissage à leurs élèves. Cependant, une telle situation n'est pas totalement marquée, ni entre catégories ni pour toutes les disciplines.

Vision générale des pratiques déclarées par les enseignants

Chaque enseignant a été aussi consulté sur la fréquence avec laquelle il employait, dans chaque classe, un ensemble des méthodologies et/ou des stratégies d'enseignement (9 propositions en mathématique et 11 pour les autres disciplines). La fréquence était mesurée en quatre niveaux, depuis « *toujours ou presque* » à « *jamais ou presque* ». A partir de ces déclarations nous avons construit des « clusters » ou typologies d'enseignants par discipline (cf. Annexes méthodologiques). Notre intérêt premier n'était pas d'établir une catégorisation exhaustive mais plutôt de dégager les enseignants ayant davantage recours au travail en groupe des élèves. En effet, nous pensons que ces enseignants, de par leurs

⁴¹Une explication plausible à cet égard peut considérer le nombre trop réduit des contenus demandés (n=5), donc, une sensibilité moindre de la variable en question pour saisir les différences. Néanmoins, ces distinctions peuvent avoir trait aux différences d'orientation entre le privé et le public. En effet, bien que les programmes soient établis par décret, les établissements privés prennent certaines libertés à l'heure de les appliquer, notamment pour des matières sensibles, telles que l'interprétation des faits historiques et sociaux.

pratiques, poussent leurs élèves à se contacter entre eux et donc à s'influencer mutuellement, autrement dit, en favorisant des effets de pairs plus forts et/ou plus évidents. Néanmoins, il nous intéresse surtout de contrôler ce facteur. De cette manière, nous pensons qu'un enseignant qui, par exemple, demande souvent à ses élèves de faire des rapports bibliographiques ou des exposés oraux a en effet (plus de chances d'avoir) un penchant envers les pratiques de type groupal en ce qui concerne le travail concret des élèves. Certes, nous ne savons pas exactement si, dans ces deux exemples, il s'agit d'un travail fait à deux ou à plus élèves, mais cela reste fortement probable. De même, d'autres pratiques telles que l'exposition verbale des contenus de la part du professeur, supposent un travail groupal moins probable. Nous avons donc opté pour utiliser l'ensemble des propositions disponibles de manière à dégager des profils minimaux des pratiques, y compris celles à forte composante groupale. Le tableau B23 ci bas illustre les réponses des enseignants vis-à-vis de ces différentes propositions, en termes de somme des pourcentages des deux fréquences les plus hautes.

Tableau B23

Proportion d'enseignants déclarant pratiquer dans leurs cours, différentes options méthodologiques avec une fréquence de « toujours ou presque » ou bien « fréquemment ». Par discipline. (9412 classes) (8^{ème} grade - 2004)

QUESTION : <i>Avec quelle fréquence employez-vous, pour les élèves de cette classe, les stratégies ou méthodologies suivantes ?</i>	% d'enseignants déclarant... « toujours ou presque » + « fréquemment »			
	Maths	Lang	Sc. Nat	Sc. Soc
PROPOSITION				
Travail Groupale des élèves	72,0	85,4	87,5	85,2
Travail Individuel des élèves	89,8	89,8	79,9	83,3
Exposition verbale des contenus du cours	64,7	80,0	74,0	75,9
Organisation du cours en base de Question / Réponses	74,9	76,6	77,9	74,3
Utilisation de Guides écrites d'apprentissage	89,6	83,8	77,2	74,6
Utilisation de Questionnaires Brefs	51,0	67,8	74,0	72,8
Exposés Oraux de la part des élèves	32,6	71,4	64,7	64,8
Recherches Bibliographiques et rapports écrits	24,8	67,5	67,0	65,8
Dramatisations ou montages de pièces théâtrales	-	42,9	11,6	18,1
Réalisation des Débats	-	42,0	26,3	32,3
Réalisation des Visites Guidées hors de l'établissement	6,1	12,5	10,9	9,8

Source : élaboration de l'auteur. Lecture : en langage, 90% des enseignants/classes déclare pratiquer en cours, toujours ou fréquemment, le travail individuel des élèves. Les propositions sont notées dans l'ordre original.

Les tendances vis-à-vis des pratiques se ressemblent d'une matière à l'autre. De même, certaines d'entre elles semblent moins importantes ou moins utilisées que d'autres. C'est notamment le cas des visites guidées, avec près de 13% des enseignants se déclarant dans les catégories « fréquemment » ou « toujours ». D'autre part, les mathématiques s'écartent

assez des trois autres matières en ce qui concerne les recherches bibliographiques (25% contre plus de 65%), l'utilisation des questionnaires brefs (51% contre près de 70%), ou encore, le travail groupal des élèves (72% contre plus de 85%). En revanche, l'utilisation des guides d'apprentissage semble une méthode assez fréquente pour les enseignants de maths (90%), bien que les autres le fassent aussi (75% en sciences sociales, 77% en sciences de la nature et 84% en langage). Globalement, nous postulons que ces pratiques reflètent davantage un travail de la part des élèves, tantôt individuel, tantôt groupal. Cette appréciation *a priori* est reflétée par les propositions en gris pour le travail groupal et en blanc pour le travail plutôt individuel, sachant qu'un même enseignant peut avoir recours à plusieurs d'entre elles à la fois.⁴²

Bref, pour arriver à un classement des enseignants à partir de cet ensemble de pratiques, nous avons eu recours à différentes analyses successives, en essayant d'y voir plus clair. De fait, notons que celles-ci sont, en général, positivement corrélées entre elles et qu'il est donc difficile de dégager des tendances opposées suffisamment claires. Nous avons donc décidé de conduire une analyse factorielle (extraction et rotation des axes) de façon à distinguer si les variables se prêtaient effectivement à un tel regroupement comme celui proposé *a priori*.

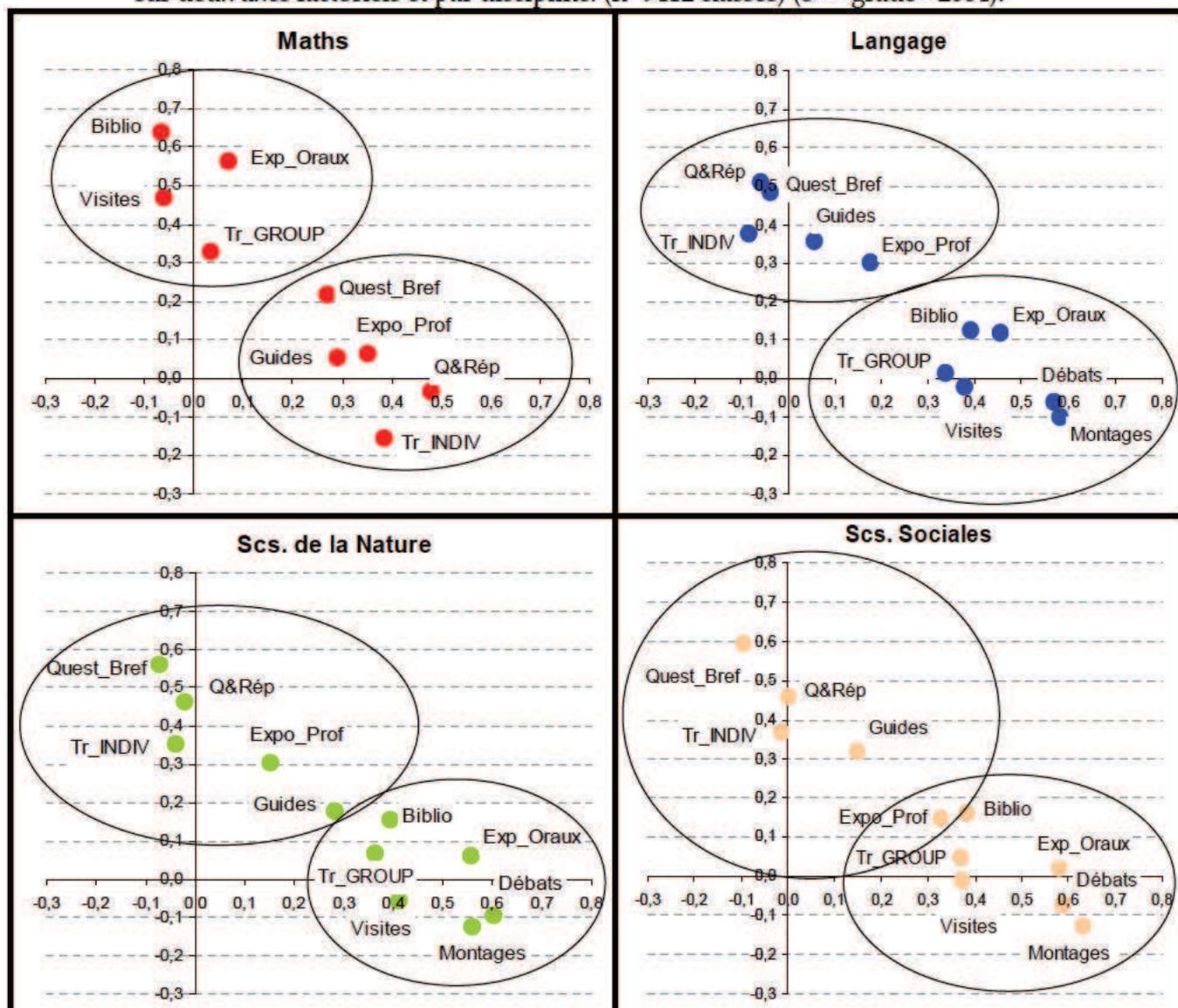
Ainsi, le graphique B36 illustre la position des propositions sur le tableau B23, prises comme des variables sur un plan factoriel à deux dimensions. Pour chacune des disciplines, les deux dimensions présentées sur le graphique B36 permettent de confirmer nos perceptions initiales sur l'existence d'une tendance au rassemblement des pratiques en deux grands groupes. En toute logique, la séparation (i.e. l'indépendance) entre ces deux groupes de pratiques n'est pas totale. Bien au contraire, toutes les corrélations entre les axes sont positives (et de l'ordre de 0,5) ce qui signifie que le fait de manifester un certain type de pratiques est associé à manifester aussi l'autre type. D'ailleurs, nous pouvons voir des pratiques telles que l'utilisation des guides d'apprentissage ou même l'exposition verbale des contenus que tendent à se placer dans une zone intermédiaire nous indiquant ainsi qu'elles sont liées aux deux groupes de pratiques à la fois (e.g. sciences de la nature et sciences sociales).

C'est à cause de cette relative confusion, mais aussi des distances parfois faibles entre groupes distincts de variables, que l'extraction des coordonnées factorielles reste illustrative mais qu'elle ne suffit pas pour construire une typologie. Ainsi, pour obtenir des catégories davantage identifiables, nous avons finalement eu recours à la méthode de regroupement dite de « clusters » basée sur l'algorithme de nuées dynamiques (cf. annexes méthodologiques).

⁴² Notons que la construction de la question semble nuire à une saisie plus claire du problème, limitant ainsi la différenciation entre profils.

Graphique B36

Vision des pratiques en cours déclarées par les enseignants, sur deux axes factoriels et par discipline. (n=9412 classes) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Note : La nomenclature des pratiques est une réduction des intitulés du tableau précédant. Lecture : Les corrélations partielles des variables avec chaque axe extrait par l'analyse factorielle correspondent aux coordonnées sur chaque plan. Plus une variable est proche d'un axe (angle ≈ 0) plus cette variable explique l'axe ou dimension concernée. Par exemple, en maths l'axe horizontal, ou première dimension, regroupe les variables décrivant des pratiques plutôt individualistes vis-à-vis du travail des élèves (e.g. *travail individuel* (0,39 ; -0,16)), tandis que le deuxième axe regroupe les pratiques plutôt groupales (e.g. *travail groupal* (0,04 ; 0,33)). La proportion de variance initiale associée à chaque sous-graphique, oscille entre 45% (maths et langage) et 42% (les autres).

Ainsi, les quatre tableaux suivants (B24 à B27) nous présentent les enseignants par classe et les groupes de classification construits pour chaque discipline. Les enseignants en mathématiques et en sciences de la nature ont été découpés en trois sous-groupes chacun. Les autres, ont été classifiés sur quatre profils.

Tableau B24
Fréquence moyenne des pratiques déclarées par les enseignants en
MATHÉMATIQUES selon regroupement par catégories (3 clusters).
(9412 classes) (8^{ème} grade - 2004)

<i>R</i> ² moyen = 0,26	ENSEMBLE	Pratiques peu définies ou absence d'elles	Pratiques plutôt individuelles	PLUSIEURS PRATIQUES (individuelles et groupales)
<i>n</i>	9412	3481	3269	2662
%	100%	37,0%	34,7%	28,3%
Travail Groupale des élèves	2,9	2,7	2,7	3,3
Travail Individuel des élèves	3,3	2,9	3,7	3,4
Exposition verbale des contenus	2,8	2,4	2,9	3,2
Question / Réponses	3,0	2,4	3,3	3,3
Guides écrites	3,4	2,9	3,7	3,7
Questionnaires Brefs	2,5	2,0	2,5	3,1
Exposés Oraux de la part des élèves	2,2	1,8	1,9	3,0
Recherches Bibliographiques	2,0	1,8	1,7	2,8 (*)
Visites Guidées hors l'établissement	1,4	1,2	1,1	1,9

Source : élaboration de l'auteur. Lecture (*) : les enseignants/classes de mathématiques classés dans le groupe ou cluster « *Plusieurs pratiques (individuelles et groupales)* » sont 2 662 (28% du total). Ils déclarent une fréquence moyenne (entre 1 et 4) de 2,8 pour la pratique « *Recherches bibliographiques et rapports écrits de la part des élèves* ». Ils dépassent d'une manière importante la moyenne générale des enseignants en maths vis-à-vis de cette même pratique (i.e. 2,0). Rappelons que les valeurs de l'échelle considérée correspondent à : (1) « *Jamais ou presque* » ; (2) « *Occasionnellement* » ; (3) « *Fréquemment* » ; (4) « *Toujours ou presque* ». Il faut rappeler aussi que, *stricto sensu*, il s'agit de pratiques de l'enseignant pour chaque classe.

Notons que les coefficients R^2 moyens, communiqués pour chacune des disciplines, nous illustrent à quel point les catégories construites rendent compte de la variabilité dans les fréquences d'usage des pratiques considérées. Nous sommes, en fait, en face d'un instrument (i.e. la question posée) pas assez discriminant. Cependant, nous pensons que l'arbitrage entre parcimonie et difficultés inhérentes aux données, nous donnent une solution plutôt satisfaisante pour rendre compte des différences dans les pratiques en salle de classe.

Tableau B25
Fréquence moyenne des pratiques déclarées par les enseignants en
LANGAGE selon regroupement par catégories (4 clusters).
(9412 classes) (8^{ème} grade - 2004)

<i>R² moyen = 0,29</i>	ENSEMBLE	Pratiques plutôt individuelles	PLUSIEURS PRATIQUES (indiv. et groupales)	Pratiques plutôt GROUPELES	Pratiques peu définies ou absence d'elles
<i>n</i>	9412	4693	1849	1763	1107
<i>%</i>	100%	49,9%	19,6%	18,7%	11,8%
Travail Groupale des élèves	3,1	3,0	3,5	3,3	2,7
Travail Individuel des élèves	3,3	3,4	3,5	3,2	2,8
Exposition verbale des contenus	3,1	3,1	3,5	3,0	2,3
Question / Réponses	3,0	3,2	3,4	2,9	2,1
Guides écrites	3,2	3,3	3,7	2,9	2,3
Questionnaires Brefs	2,9	3,1	3,5	2,2	1,9
Exposés Oraux de la part des élèves	2,9	2,7	3,6	3,3	2,0
Recherches Bibliographiques	2,8	2,6	3,5	3,2	2,0
Dramatisations ou montages	2,4	2,1	3,4	2,6	1,8
Réalisation des Débats	2,4	2,1	3,3	2,8	1,7
Visites Guidées hors l'établissement	1,7	1,5	2,5	1,6	1,3

Source : élaboration de l'auteur. Lecture : Idem tableau antérieur.

Tableau B26
Fréquence moyenne des pratiques déclarées par les enseignants en
SCIENCES de la NATURE selon regroupement par catégories (3 clusters).
(9412 classes) (8^{ème} grade - 2004)

<i>R² moyen = 0,23</i>	ENSEMBLE	Pratiques plutôt traditionnelles	PLUSIEURS PRATIQUES (indiv. et groupales)	Pratiques peu définies ou absence d'elles
<i>n</i>	9412	4410	2581	2421
<i>%</i>	100%	46,9%	27,4%	25,7%
Travail Groupale des élèves	3,2	3,2	3,6	2,7
Travail Individuel des élèves	3,1	3,1	3,2	2,8
Exposition verbale des contenus	2,9	3,0	3,4	2,4
Question / Réponses	3,0	2,8	3,5	3,1
Guides écrites	3,0	2,9	3,6	2,8
Questionnaires Brefs	3,0	2,7	3,5	3,0
Exposés Oraux de la part des élèves	2,8	2,8	3,5	2,0
Recherches Bibliographiques	2,8	2,9	3,4	2,1
Dramatisations ou montages	1,6	1,4	2,2	1,2
Réalisation des Débats	2,1	1,9	2,8	1,5
Visites Guidées hors l'établissement	1,6	1,5	2,1	1,3

Source : élaboration de l'auteur. Lecture : Idem tableau antérieur.

Tableau B27

Fréquence moyenne des pratiques déclarées par les enseignants en SCIENCES SOCIALES selon regroupement par catégories (4 clusters).
(9412 classes) (8^{ème} grade - 2004)

<i>R² moyen = 0,24</i>	ENSEMBLE	PLUSIEURS PRATIQUES (indiv. et groupales)	Pratiques plutôt INDIVIDUELLES	Pratiques peu définies ou absence d'elles	Pratiques plutôt GROUPELES
<i>n</i>	9412	4604	2186	1617	1005
<i>%</i>	100%	48,9%	23,2%	17,2%	10,7%
Travail Groupale des élèves	3,1	3,3	2,7	2,9	3,3
Travail Individuel des élèves	3,1	3,4	3,2	2,6	2,7
Exposition verbale des contenus	3,0	3,2	2,9	2,2	3,1
Question/ Réponses	3,0	3,3	3,0	2,2	2,5
Guides écrites	3,0	3,4	2,6	2,5	2,6
Questionnaires Brefs	2,9	3,4	2,5	2,8	2,2
Exposés Oraux de la part des élèves	2,8	3,2	2,2	2,2	3,2
Recherches Bibliographiques	2,8	3,2	2,2	2,5	2,8
Dramatisations ou montages	1,8	2,1	1,3	1,3	2,7
Réalisation des Débats	2,2	2,5	1,7	1,7	2,9
Visites Guidées hors l'établissement	1,6	1,7	1,3	1,3	2,0

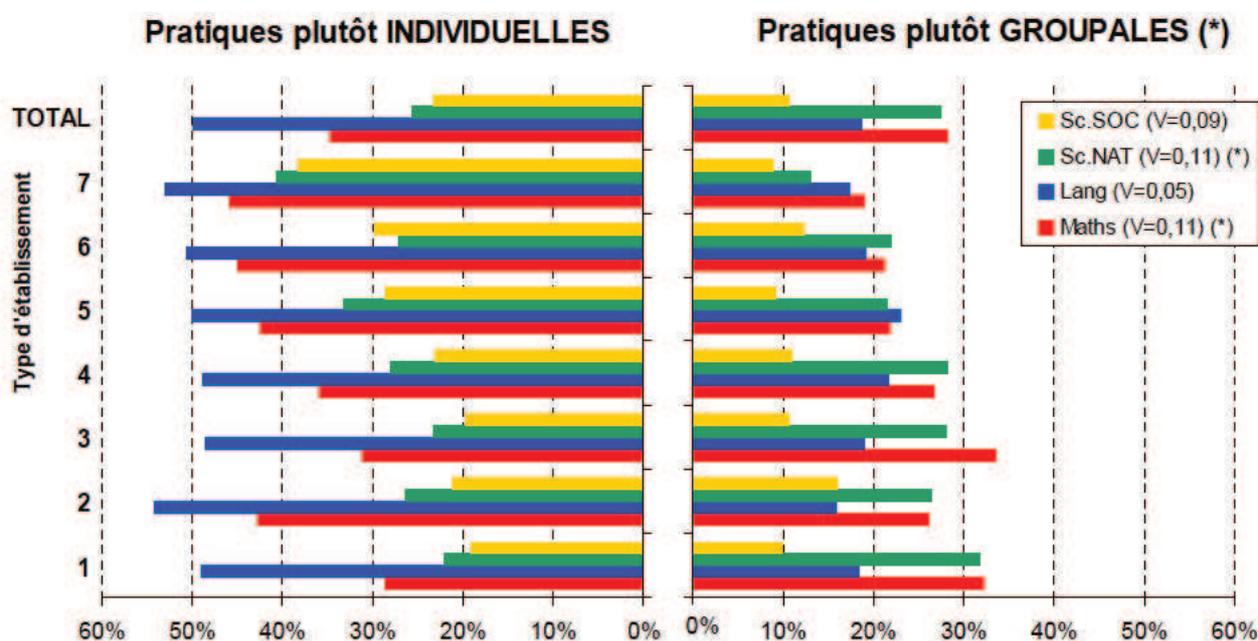
Source : élaboration de l'auteur. Lecture : Idem tableau antérieur.

Enfin, notons qu'il existe des différences entre les proportions d'enseignants par profils de pratiques, selon les catégories d'établissements. Ceci est illustré par le graphique B37. Notons que sont seulement pris en compte les regroupements ou les clusters qui représentent des pratiques soit fortement individuelles, soit plutôt groupales.⁴³Nous pouvons y voir que les pratiques à caractère plus individualiste – de la part du travail des élèves et selon l'appréciation des enseignants – sont, en général, plus importantes en proportion que les pratiques fortement groupales ou ayant un recours élevé à celles-ci. Ceci à l'exception de sciences de la nature (26% contre 27%). Néanmoins, cela semble modulé quelque peu par le type d'établissement et selon les disciplines. En effet, il se dessine une légère tendance à la différenciation entre les catégories privées à public plus aisé (payantes et sélectives, avec ou sans subvention) et les catégories gratuites. Ces dernières tendent à concentrer des proportions plus importantes des pratiques plutôt groupales, pour toutes les disciplines sauf en langage où la différenciation est moins nette.

⁴³Notons qu'en maths et en sciences de la nature, les clusters correspondent à « Plusieurs Pratiques », car, parmi les trois sous-divisions dans ces disciplines ne se distinguent pas des groupes spécifiques à pratiques groupales. Dans le cas des pratiques plutôt individuelles, le cluster choisi est celui des pratiques « Pas bien définies », davantage éloigné des pratiques groupales (cf. tableaux B24 et B26).

Graphique B37

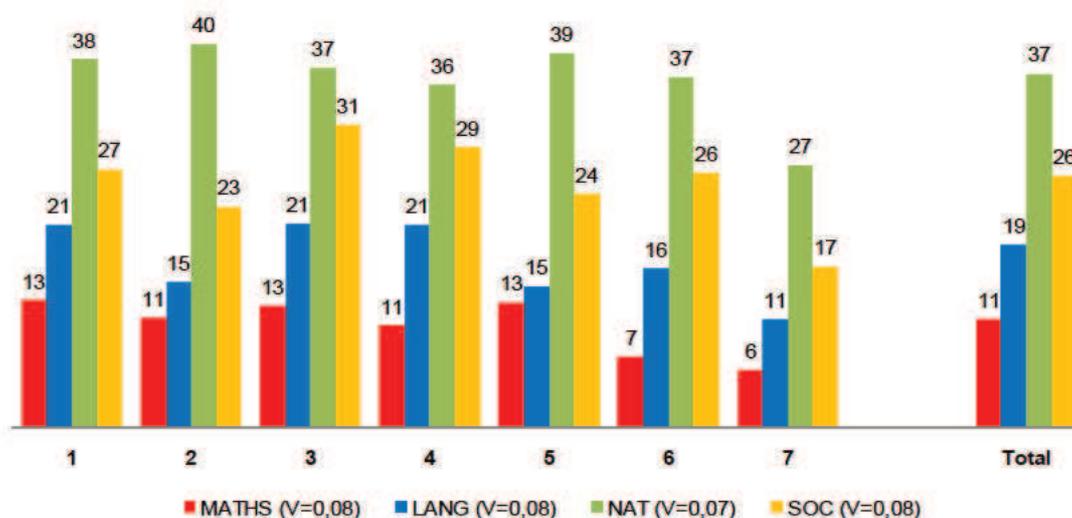
Proportions d'enseignants (classes) qui déclarent avoir des pratiques en salle plutôt individuelles ou plutôt groupales. Par discipline et selon type d'établissement.
(n=9 412 classes) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

Graphique B38

Proportions des classes qualifiées comme étant à pratiques groupales à partir des perceptions majoritaires des élèves. Par discipline et selon type d'établissement.
En pourcentage. (n=9 412 classes) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture Type d'établissement, idem tableaux antérieurs.

Enfin, une perspective complémentaire est apportée par les déclarations des élèves vis-à-vis des pratiques enseignantes (cf. graphique B38). Si dans l'ensemble, les différences ne semblent pas très marquées, nous trouvons, encore une fois, que le secteur privé payant s'éloigne de l'usage (cette fois-ci perçu par les élèves) des pratiques à caractère plutôt groupal pour le travail des élèves en cours. En effet les classes dans les établissements à public favorisé (catégories n°6 et n°7) semblent travailler significativement moins sous de telles modalités, notamment en maths et langage.

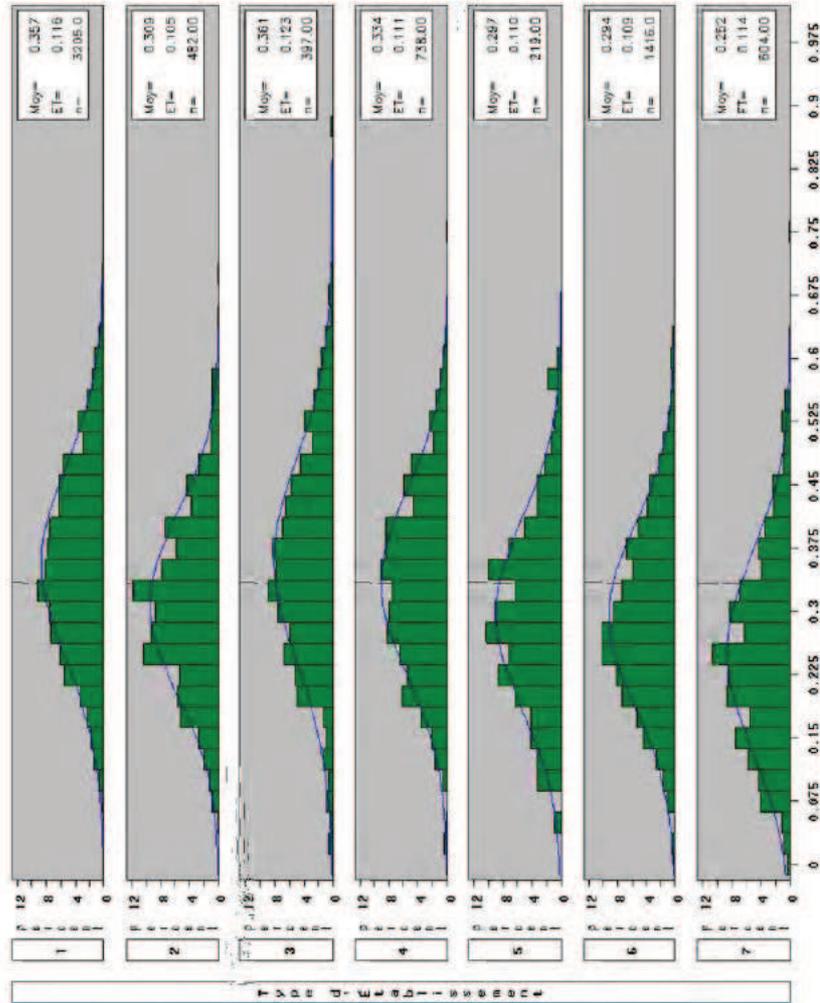
Les proportions des élèves par classe en conditions de Bas BEE et de Haute CSS

Un dernier regard sur le niveau classe est apporté par deux variables que nous avons construites en essayant de mieux contrôler des potentiels phénomènes reliés à la notion de « climat scolaire » (cf. chapitres n°1 et n°5). En effet, nous avons décidé de renvoyer le détail des analyses sur ces deux facteurs pour des travaux ultérieurs. Pour le moment nous voudrions juste souligner que la distribution des proportions par classe de ces variables maintient une allure normale par type d'établissement, tout en illustrant les différences de moyenne entre ceux-ci. En effet, nous retrouvons les distances par catégorie soulignées au niveau individuel (cf. graphique B4 et tableau B12), mais désormais vues à partir de valeurs agrégées sur le graphique B39.

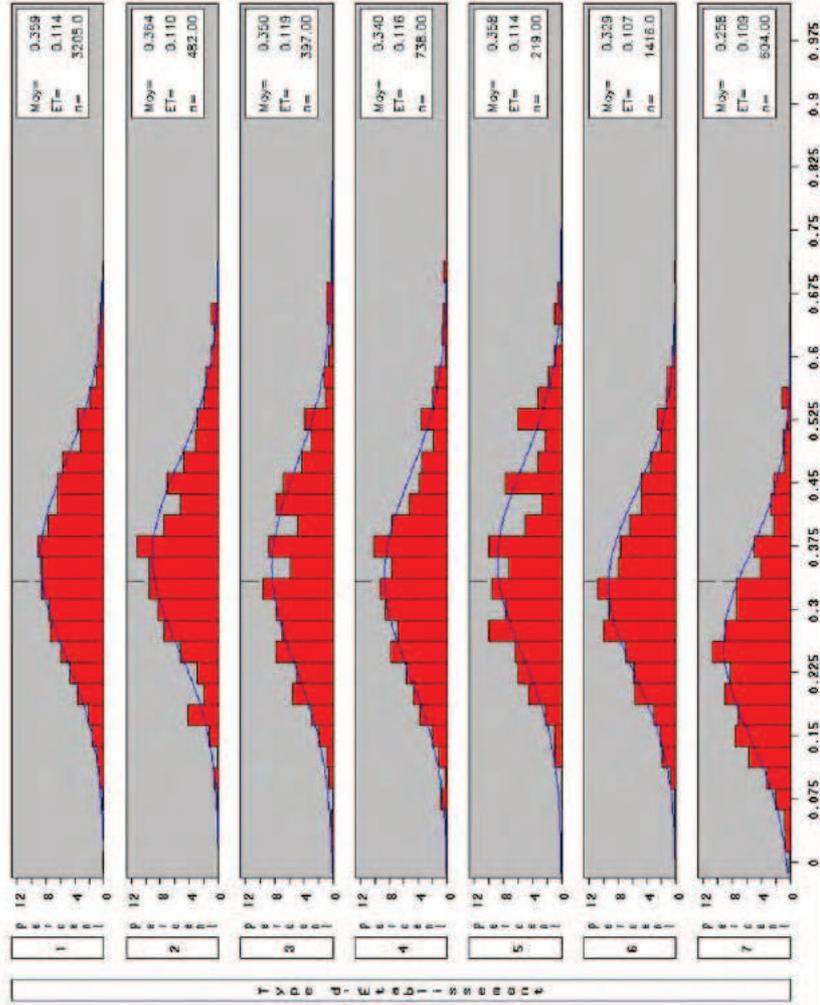
Graphique B39

Histogrammes des distributions des classes selon leur proportion d'élèves en Bas BEE ou en Haute CSS. Par type d'établissements, (n=7 061 classes urbaines à >19 élèves) (8^{ème} grade - 2004)

% par Classe en condition de Bas BEE



% par Classe en condition de Haute CSS



Source : élaboration de l'auteur. Lecture type d'établissement : Idem tableaux antérieurs. Les respectives moyennes générales (i.e. 33%) sont notées par la ligne à traits.

VARIABLES GENERALES DE NIVEAU INDIVIDUEL

Facteurs sociodémographiques et socio-scolaires

Nous avons déjà commenté le genre et la condition de redoublement des élèves. Maintenant, nous allons considérer certaines caractéristiques de la famille où l'enfant évolue, même si cela se fait sur une coupure transversale. En effet, après quelques analyses nous avons décidé de considérer deux aspects sur le groupe familial de chaque élève. D'abord, le fait que l'élève ait des frères, dont le nombre, le sexe et la position dans la fratrie, restent, hélas, inconnus. Cette variable nous semble être un contrôle significatif vis-à-vis des effets de pairs. De fait, la possibilité que les frères aident les élèves pour leur travail scolaire représente un facteur dont l'impact peut se confondre avec celui des pairs d'école. D'autre part, la sujétion à un même environnement familial par des frères placés dans le même grade et/ou la même école, représente un phénomène qu'il serait souhaitable de contrôler.⁴⁴ Nous avons considéré, ensuite, que les élèves dont les foyers ne sont pas constitués d'un couple étaient dignes d'attention particulière car ils profitent, vraisemblablement, de moindres opportunités de soutien parental. De plus, le sujet reste d'importance dans la mesure où, au Chili, il existe un pourcentage assez significatif de foyers monoparentaux, notamment sous la tutelle des femmes. Cette réalité serait davantage prégnante dans les secteurs défavorisés urbains.

Enfin, un troisième facteur assez souligné par la littérature est rapporté à l'enseignement préscolaire des élèves (e.g. Mingat et al., 1992⁴⁵ ; Caille, 2001⁴⁶).⁴⁷ Si les écarts sociaux à l'école semblent avoir une « *genèse précoce* » (Duru-Bellat, 2003a, p. 55-59) il n'est pas moins vrai que certaines inégalités « devant l'école », devraient montrer une influence modulatrice sur les acquis scolaires. Nous tentons donc de saisir l'impact probable sur les acquis qui découle du fait d'être passé tôt ou non par le préscolaire.

Ainsi, le graphique B40 illustre la distribution des trois variables ici commentées : préscolaire, fratrie et type de famille.

⁴⁴C'est le cas de la recherche menée par McEwan (2003) sur le Chili. Il essaye d'isoler, par approximation, les frères, de façon à identifier un même environnement familial. Une telle démarche nous ne l'entreprendrons pas ici, bien qu'elle reste une possibilité pour le futur.

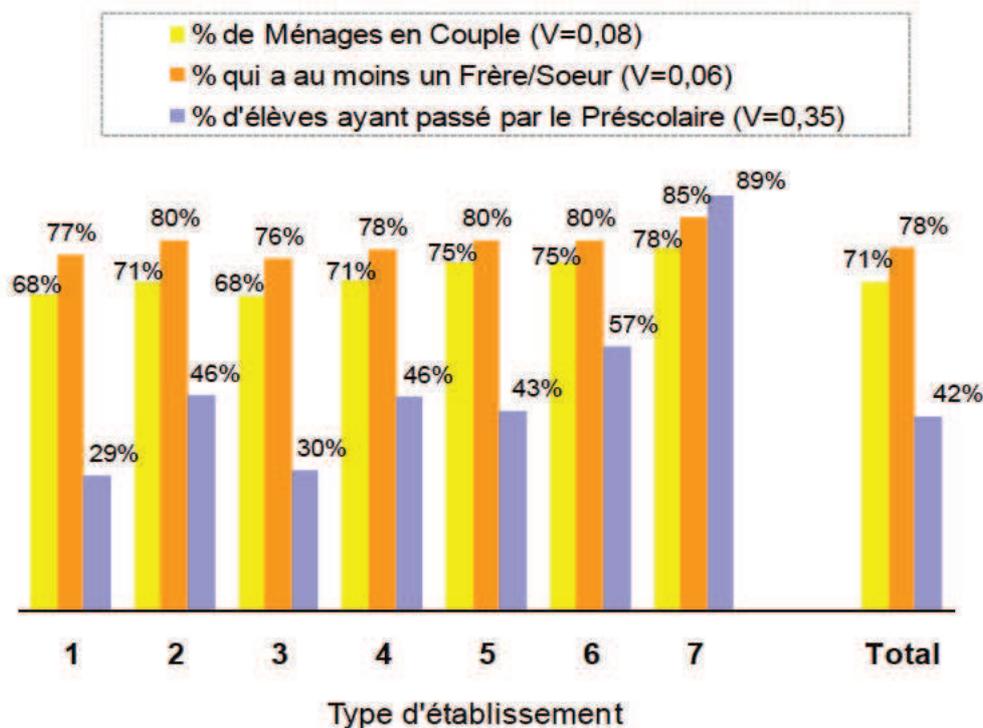
⁴⁵ Mingat, A. ; Jarousse, J.P. ; Richard, M. ; 1992 ; *La scolarisation maternelle à deux ans : effets pédagogiques et sociaux* ; Education et Formations, n°31 ; pp. 3-9

⁴⁶ Caille, J.P. ; 2001 ; *Scolarisation à deux ans et réussite de la carrière scolaire au début de l'école élémentaire* ; Education et Formations, n°60 ; pp. 7-18

⁴⁷ Faute d'autres informations, nous avons considéré les réponses à la question « *l'élève, est-il allé à 'pré-kindergarten' ?* » (i.e. 3 à 4 ans) comme source pour la construction de cette variable. D'autres enquêtes nous ont donné raison (cf. CASEN).

Graphique B40

Proportions, par type d'établissement, des trois différentes caractéristiques sociodémographiques des élèves. (n=278 476) (8^{ème} grade - 2004).



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture : 30% des élèves inscrits chez le privé subventionné non-sélectif et sans FP (n°3) sont passés par le préscolaire, au moins dès le niveau pré-kindergarten (≈4 ans).

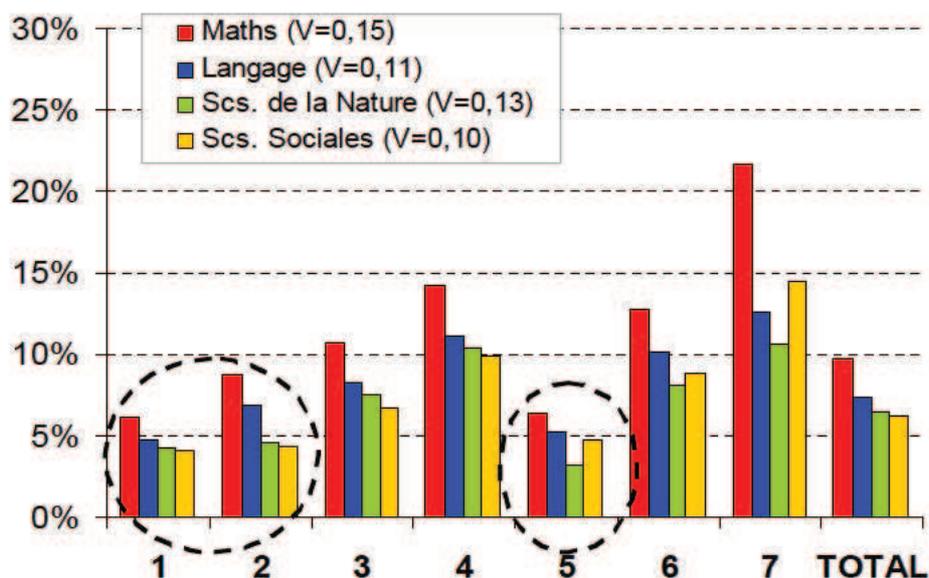
Avoir ou non le livre du cours par discipline

Le graphique B41 illustre les proportions d'élèves qui déclarent ne pas posséder de livre du cours de l'année dans chaque discipline. Certes, au total, la grande majorité des élèves a cet outil. Pourtant, il existe des différences importantes entre types d'établissement (e.g. V de Cramer $\geq 0,1$). Cette variable est issue des déclarations directes des élèves sur leur travail avec les textes scolaires respectifs, autant à la maison que dans la classe. Le constat le plus intéressant est que les élèves du secteur public (n°1 et n°2), mais aussi, ceux du secteur privé subventionné sélectif sans FP (n°5), indiquent les plus bas pourcentages d'élèves déclarant ne pas avoir le livre du cours pour les quatre matières. En revanche, le secteur privé payant s'écarte de la moyenne nationale et ses élèves déclarent davantage ne pas posséder ces outils (entre 8% et 14% selon la discipline). Notons que dans le total, mais aussi pour chacune des catégories, la possession du livre de mathématiques est moins répandue qu'ailleurs. A l'opposé, en sciences de la nature et sciences sociales cet outil

semble davantage valorisé et présent, au point que seulement 6,6% des élèves du pays ne l'avaient pas, en 2004 et pour le 8^{ème} grade.

Graphique B41

Pourcentages d'élèves par discipline et selon type d'établissement qui N'AURAIENT PAS le livre du cours. (n=278 476 élèves) (8^{ème} grade - 2004)



Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous).

En outre, notons que les différences entre catégories peuvent se rapporter à des politiques explicites de la part des établissements. Dans les secteurs du privé pour lesquels le paiement devient un enjeu important (n°6 et n°7), il se peut que ces politiques se passent autrement et/ou que des pratiques pédagogiques (non observables) compensent ce moindre équipement relatif. Notons tout de même que, d'une part, la question posée aux élèves ne différencie pas le type de livre utilisé (officiel v/s autre, par exemple) ; et que, d'autre part, le niveau d'équipement national semble assez important (entre 90% et 94% des effectifs selon la discipline).

L'indice de Niveau Socioéconomique (NSE) construit

Un ensemble de variables qui nous a fortement occupé pour ce travail est celui des facteurs reflétant le statut socioéconomique des familles, et par là, l'origine sociale des élèves. Nous avons déjà commenté en détail la scolarité des deux parents (cf. graphiques B8 et B9). Maintenant, nous partons du constat de l'existence de fortes corrélations entre la scolarité des mères et celle des pères. Puis, des fortes liaisons entre ces deux variables et les revenus par ménage et individuels (i.e. revenu du ménage divisé par sa taille), de même qu'avec la possession de certains biens durables à la maison. Nous avons donc décidé d'entreprendre la construction d'un Indice composite du Niveau Socioéconomique (Indice de NSE) attaché à chaque élève.⁴⁸ La matrice de corrélations entre les variables constitutives de cet indice est notée dans le tableau B28 suivant.

Tableau B28

Matrice des Corrélations de Pearson entre différentes variables reflétant le capital économique et humain disponible pour chaque élève. (278476 élèves - 8^{ème} grade - 2004)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	NSE
(1) Scolarité Mères	-											0,85
(2) Scolarité Pères	0,70	-										0,87
(3) Rev. Indiv. (LOG)	0,58	0,60	-									0,83
(4) Chauffe-eau	0,38	0,38	0,42	-								0,51
(5) Machine à laver	0,30	0,31	0,35	0,39	-							0,42
(6) Réfrigérateur	0,19	0,19	0,18	0,32	0,29	-						0,24
(7) Téléphone fixe	0,35	0,37	0,41	0,46	0,35	0,26	-					0,50
(8) Enregistreur Vidéo	0,34	0,36	0,42	0,30	0,29	0,19	0,32	-				0,53
(9) Four Micro-ondes	0,36	0,37	0,43	0,40	0,35	0,23	0,39	0,40	-			0,55
(10) TV Câble	0,30	0,32	0,39	0,21	0,22	0,13	0,26	0,30	0,32	-		0,48
(11) Automobile	0,38	0,39	0,50	0,31	0,28	0,18	0,31	0,38	0,38	0,33	-	0,59
Téléphone portable	0,16	0,16	0,22	0,17	0,22	0,21	0,05	0,20	0,21	0,15	0,24	0,25

Source : élaboration de l'auteur. Lecture : entre la scolarité des pères et le logarithme naturel du revenu par tête du foyer il existe une corrélation de +0,6. Les corrélations de chaque variable avec l'Indice de NSE construit sont illustrées à droite. Toutes les corrélations sont significatives (à $p < 0,000$)

Cet Indice est donc voué à diminuer la duplication d'information tout en profitant, à la fois, du maximum disponible de celle-ci. Celui-ci devrait nous permettre de différencier *in fine* chaque élève en termes de niveau de capital économique et humain disponible. Ainsi, le double avantage de cette échelle différenciatrice, capable de synthétiser beaucoup d'informations, nous permettra par la suite de diviser la population en groupes de taille homogène (de 20% et de 10%), tout en disposant d'une mesure continue utilisable selon le

⁴⁸ Ceci sachant que d'autres recherches sur des données chiliennes (cf. Mizala et al., 2005 ; Valenzuela et al., 2008) ont fait pareil, mais sans pour autant spécifier les caractéristiques générales des variables. De plus, ces démarches ont fait souvent appel à l'extraction voie ACP (cf. annexes méthodologiques).

type et les objectifs des analyses. En outre, la construction de cet indice nous a permis d'estimer quelles variables, parmi les douze mises à contribution, sont susceptibles de mieux refléter les autres, autant d'un point de vue théorique qu'empirique.

Ainsi, pour construire le dit Indice de NSE nous avons, d'abord, rempli les données manquantes à l'aide de nos propres données, notamment en régressant la scolarité des pères et des mères sur d'autres données sociodémographiques disponibles. Puis, en utilisant les informations, désagrégées par commune et région, fournies par l'Enquête nationale de ménages (CASEN) réalisée en 2003 (c'est-à-dire un an avant l'enquête SIMCE sur laquelle nous travaillons). De même, nous avons considéré les estimations du PNUD au Chili, pour obtenir certaines variables par commune (i.e. revenu, scolarité féminine et masculine, Indice du Développement Humain ou IDH). Nous avons, enfin, comparé toutes ces données avec les nôtres (i.e. les revenus moyens des ménages par commune, les années de scolarité, etc.). Nous avons trouvé que les corrélations de Pearson étaient, au niveau des régions, respectivement de 0,92 et de 0,93 pour la scolarité des mères et des pères respectivement, et de 0,45 au niveau de communes. Ces valeurs étaient aussi de 0,95 pour le revenu moyen par commune. Nous avons donc considéré que, en cas d'impossibilité de trouver un meilleur estimateur de remplissage, nous pouvions utiliser ces autres données pour combler nos lacunes.⁴⁹

Une fois le remplissage satisfait, nous avons estimé le revenu individuel disponible mensuellement pour chaque ménage, puis, calculé le Logarithme naturel de ce chiffre. Cette première transformation provient du fait que nos analyses préliminaires nous amènent à considérer une relation quadratique entre les scores aux tests d'acquis scolaires et le revenu des ménages. Cette relation serait donc mieux capturée par le logarithme de cette variable, tout en continuant à avoir une seule mesure du phénomène (et non pas deux). En même temps, parmi les douze biens durables à la maison enquêtés par le questionnaire aux parents, nous en avons sélectionné neuf (cf. tableau B28).⁵⁰ Finalement, avec ces trois sources : années de scolarité, revenu individuel par ménages et biens à la maison, nous avons construit et comparé différentes analyses factorielles sous quatre méthodes différentes (cf. annexes méthodologiques) en vue d'obtenir une solution consistante. Nous avons estimé aussi un Alpha de Cronbach équivalent à 0,85, ce qui est

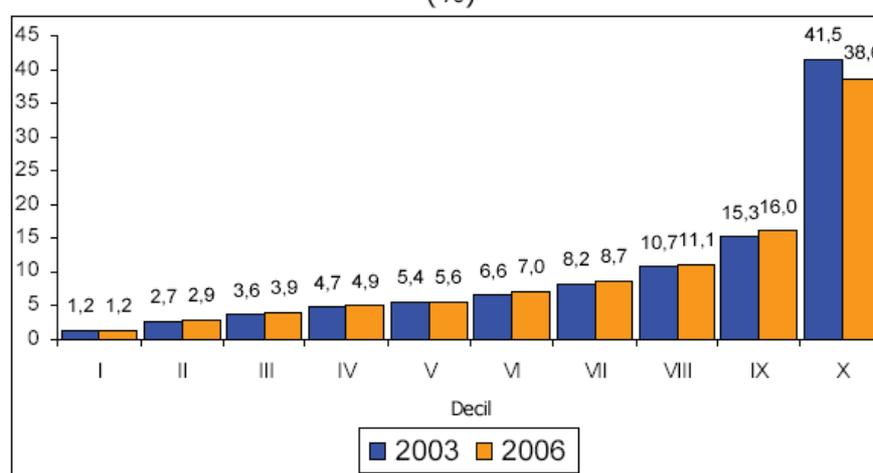
⁴⁹ L'ampleur de ces lacunes était, à l'origine et respectivement, de 25% et 27% pour la scolarité des mères et des pères ; tandis que les données manquantes pour le revenu du ménage n'atteignaient que 4%. Pourtant, le questionnaire aux parents fournissait beaucoup d'autres informations, notamment le type d'emploi du chef de famille, le type de sécurité sociale et d'assurance retraite, les frais de scolarisation directs reversés aux écoles (avec 4% de données manquantes), les dépenses indirectes de scolarisation (8% de manquantes), les biens à la maison (8%), la commune d'habitation, etc. Ceci parmi d'autres données assez utiles pour construire des approximations au profil réel des familles n'ayant pas répondu à toutes les questions principales.

⁵⁰ Les biens non pris en compte pour l'Indice de NSE correspondent à « Ordinateur », « Accès à Internet » et « Imprimante ». En effet, nous avons considéré ceux-ci comme des biens de capital culturel à l'état matériel (Bourdieu, 1977) et donc, par cohérence avec les choix faits vis-à-vis du capital culturel des familles (i.e. reflété par les nombre de livres à la maison), nous avons placé ces biens à part des autres richesses.

plutôt satisfaisant comme indicateur de la pertinence d'un amalgame.⁵¹ Ainsi, nous avons entrepris nos AF obtenant, d'emblée, un MSA de 0,90 ce qui confirme la liaison et à la fois la singularité des variables choisies. Notons que la solution factorielle était stable d'une méthode à l'autre, donc, nous avons fait notre choix en regardant la valeur de *communautés* et la valeur des corrélations entre les facteurs obtenus et les variables originales. Dans cette démarche, nous avons assigné plus d'importance aux scolarités des parents et au revenu par tête des ménages. Finalement, nous avons choisi l'un des axes extraits par la méthode ULS.⁵²

Graphique B42

Distribución del ingreso autónomo según decil de ingreso
Autónomo per cápita del hogar, 2003, 2006
(%)



Fuente: Mideplan, en base a Encuestas CASEN 2003 y 2006

Source: Ministère de Planification du Gouvernement chilien (2007 ; p. 6).⁵³[Traduction du Titre : « Graphique N°1 - Distribution du revenu autonome selon décile de revenu autonome per capita du ménage, 2003, 2006 (%) »]

Conséquemment, nous avons utilisé cette échelle ou dimension factorielle pour diviser la population en cinq et dix groupes égaux en taille, et rangés des plus défavorisés aux plus favorisés. Nous avons voulu comparer cette division avec son équivalente pour l'enquête des ménages CASEN 2003, pour savoir si elle était en phase avec la réalité montrée par d'autres données nationales. Cette comparaison peut s'apprécier sur les graphiques B42 et B43. Le premier a été extrait des rapports officiels du Ministère de Planification. Il compare

⁵¹ Notons toutefois que, de même qu'avec les corrélations de Pearson notées plus haut (cf. tableau B27), nous avons ici pris la licence de considérer les biens à la maison (i.e. des variables binaires muettes) comme des variables continues. Dans le cas des analyses factorielles, la même licence a été considérée.

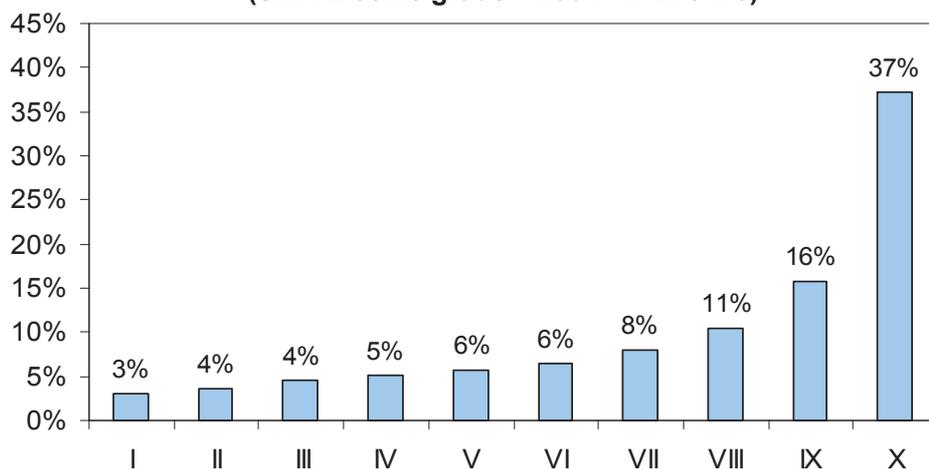
⁵² Ceci sachant que cette méthode cumule quelques avantages déjà décrits (cf. annexes méthodologiques) tout en trouvant des corrélations de 0,98 et à 0,99 (r de Pearson), respectivement, avec les axes extraits par les méthodes Alpha et ML. De même, ce résultat factoriel est corrélé à 0,94 avec le facteur extrait voie ACP. Autrement dit, nous pouvons postuler que le choix fait n'est pas gênant pour la suite du travail.

⁵³ MIDEPLAN ; 2007 ; *Documento N°2: Distribución del Ingreso e Impacto Distributivo del Gasto Social* ; Serie Análisis de resultados de la Encuesta de Caracterización Socioeconómica Nacional, CASEN 2006 ; 44 pp.

les résultats des enquêtes CASEN en 2003 et en 2006. L'autre graphique, a été construit à partir de notre propre division de la population nationale faite à partir de notre Indice de NSE.

Graphique B43

**Distribution en % du Revenu par tête
des ménages, selon décile de NSE construit.
(SIMCE 8ème grade – 2004 - n=278476)**



Source : élaboration de l'auteur à partir du Questionnaire des parents (SIMCE 2004).

Ces deux graphiques considèrent leurs populations respectives rangées par déciles. Ils nous montrent, pour ce démembrement, la distribution des revenus par tête des ménages autonomes (i.e. sans compter des subsides divers ou le service domestique). Certainement, il existe entre eux des différences entre les proportions affichées, mais celles-ci restent plutôt légères. A notre avis, cela tient au caractère approximatif des mesures faites à partir du questionnaire aux parents SIMCE (i.e. nos données), où les familles ont été invitées à déclarer leur niveau de revenu mensuel parmi quinze options ou classes. De plus, dans cette enquête il ne s'agit que d'un échantillon : celui des familles avec des enfants en 8^{ème} grade pendant l'année 2004. Néanmoins, la correspondance entre les niveaux est presque parfaite, ce qui est bon signe si l'on considère qu'il s'agit de deux enquêtes totalement distinctes mais qui aboutissent à des résultats semblables en ce qui tient à la façon dont le revenu est distribué parmi les Chiliens. Certes, cela ne dit rien par rapport au fait que les individus soient bien ou mal classés, mais cela reste un défaut de toute enquête de ce genre.⁵⁴ Nous postulons donc, que nos résultats et les classements que nous avons construits ne s'éloignent pas trop de la réalité mise en évidence par d'autres données et à des moments proches.

⁵⁴Bien sûr nos estimations sont entachées d'erreurs de mesure ainsi que de phénomènes de sur déclaration et de sous déclaration, ce que l'enquête CASEN peut corriger en faisant appel à des sources d'information multiples mais inaccessibles pour nous.

En outre, nous avons évalué la composition de la population scolaire (8^{ème} grade, 2004) rapportée à notre typologie d'établissements. Ceci à partir des déciles estimés par notre Indice de NSE. Notons que d'autres études ont aussi construit des tableaux similaires mais en ne considérant que les trois secteurs classiques dans la littérature consacrée au système scolaire chilien (cf. chapitre n°6). Quoiqu'il en soit, nos estimations sont très semblables à celles préexistantes, fixées aux dates 1990 et 2000 (Mizala et al., 2005).

Tableau B29
Distribution d'élèves par décile de NSE estimé, selon type d'établissement. En pourcentage. (8^{ème} grade - 2004)

Décile NSE	Type d'Établissement							TOTAUX	
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	%	n
I	78,2 (*)	2,8	11,4	4,0	1,3	2,2	0,03	100	27853
II	75,5	3,8	9,7	5,2	2,0	3,8	0,03	100	27811
III	70,4	4,6	9,9	6,9	2,5	5,8	0,04	100	27878
IV	66,3	5,5	7,9	8,2	3,2	8,8	0,11	100	27840
V	56,7	7,2	7,0	10,9	4,2	13,9	0,19	100	27854
VI	46,4	8,0	6,2	13,3	4,6	21,1	0,33	100	27849
VII	38,4	8,9	5,3	15,0	4,4	27,4	0,63	100	27848
VIII	27,3	10,1	3,4	15,3	4,1	37,0	2,7	100	27847
IX	15,5	10,2	2,1	14,3	3,2	44,0	10,7	100	27848
X	4,3	4,9	0,8	5,8	1,4	25,5	57,2	100	27848
TOTAL	47,9	6,6	6,4	9,9	3,1	19,0	7,2	100	278476

Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture (*) : 78,2% d'élèves appartenant au Décile de NSE « I », étudient leur 8^{ème} grade en 2004 dans des établissements municipaux non-sélectifs. Seulement un 0,03% de ce décile était inscrit dans des écoles privées payantes au même moment.

La forte segmentation socioéconomique reflétée dans le système scolaire s'apprécie très clairement dans les concentrations sur les extrêmes du tableau B29, mais aussi, sur le tableau B30. Par exemple, près de 60% du décile de NSE le plus favorisé (des effectifs de 8^{ème} grade en 2004), était inscrit dans des établissements privés payants, contre 4,3% du même décile qui le fait dans des établissements publics non-sélectifs. Par ailleurs, 16% des élèves parmi le total d'effectifs des établissements publics non sélectifs (n°1) provient du décile le plus défavorisé. Ce chiffre est de 18% pour leurs homologues privés. En revanche, 80% des élèves des établissements privés payants (de ce grade et cette année), proviennent du décile des plus favorisés, tandis que seulement 1,3% dans ces établissements est issu des trois premiers déciles. En effet, tous les établissements académiquement sélectifs et/ou

avec la modalité de financement partagé ont tendance à accueillir moins d'un élève sur cinq en provenance des trois déciles les plus défavorisés.

Tableau B30
Distribution des élèves par type d'établissement
selon décile de NSE estimé. En pourcentage. (8^{ème} grade - 2004)

Décile NSE	Type d'Établissement							
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	
I	16,3	4,3	17,9	4,0	4,3	1,2	0,03	
II	15,7	5,7	15,2	5,3	6,5	2,0	0,04	
III	14,7	7,0	15,5	6,9	8,0	3,0	0,06	
IV	13,8	8,4	12,4	8,2	10,4	4,6	0,15	
V	11,8	10,9	10,9	11,1	13,5	7,3	0,26	
VI	9,7	12,1	9,8	13,5	14,8	11,1	0,46	
VII	8,0	13,5	8,4	15,1	14,3	14,5	0,87	
VIII	5,7	15,3	5,3	15,5	13,3	19,5	3,7	
IX	3,2	15,4	3,3	14,5	10,3	23,2	14,9	
X	0,9	7,5	1,2	5,9	4,7	13,5	79,5	
TOTAUX	%	100	100	100	100	100	100	
	N	133417	18361	17721	27557	8623	52777	20020

Source : élaboration de l'auteur. Lecture : Type d'établissement : (1) PUBLIC Non Sélectif ; (2) PUBLIC Sélectif ; (3) PRIVÉ SUBV. Non Sélectif sans Financement Partagé ; (4) PRIVÉ SUBV. Non Sélectif avec Financement Partagé ; (5) PRIVÉ SUBV. Sélectif sans Financement Partagé ; (6) PRIVÉ SUBV. Sélectif avec Financement Partagé ; (7) PRIVÉ PAYANT (tous). Lecture : 17,9% d'élèves des écoles privées subventionnées non sélectives et sans FP, proviennent du Décile de NSE I.

En définitive, et à partir de ce que l'on a vu ici, nous postulons que notre Indice de NSE garde d'importantes propriétés, mais aussi qu'il présente des avantages absolus d'utilisation qui le rendent pertinent et appréciable à l'heure de chercher un bon compromis entre parcimonie et qualité des analyses. Pourtant, comme nous l'avons indiqué dans les chapitres n°11, n°12 et n°13, nous allons nous servir surtout de la scolarité des mères. Ceci, par souci de comparabilité des résultats avec la littérature, et ayant à l'esprit que cette variable exprime, de manière importante, les informations véhiculées par l'Indice de NSE (r de Pearson=0,85), plus encore, si on les regarde agrégées au niveau classe (r de Pearson=0,98). Cependant, nous nous servirons dudit indice et de ses dérivés à plusieurs reprises, notamment dans des buts exploratoires.

Les relations générales entre les variables de contexte au niveau classe

La matrice de corrélations illustrée dans le tableau B31 montre plusieurs mesures prises de manière agrégée au niveau classe. Quelques-unes le sont à titre d'exemple et d'autres ont été déjà décrites et seront insérées dans différentes analyses. Au regard de cette matrice, notons la très forte relation entre différentes manières de prendre le niveau socioéconomique moyen (par classe). En effet, nous y voyons que la corrélation (r de Pearson) entre la moyenne de l'Indice composite de NSE (voir annexes plus haut) et la moyenne d'années de scolarité des mères dans la classe est de +0,98 tandis qu'elle est un peu plus légère pour des mesures telles que la proportion d'élèves issus des 20% des plus favorisés (+0,92) ou la proportion des mères ayant plus de douze ans d'études (+0,95).⁵⁵

Quand nous regardons cette même variable et sa corrélation avec des indices agrégés de capital culturel, nous apprécions aussi une relation très forte (+0,88 ou -0,81 selon le choix). De même, ces indices entretiennent des liens forts (i.e. $\geq |0,5|$) avec la proportion de redoublants par classe, ce que ne devrait pas nous étonner compte tenu de ce que l'on a vu sur la distribution sociale du redoublement. Enfin, la relation entre tonalité sociale des classes et moyenne de résultats académiques (prise ici de manière agrégée entre les quatre disciplines) est aussi très significative (+0,81).

En même temps, nous voyons que la relation entre hétérogénéité sociale (i.e. écart-type de la scolarité des mères par classe) et niveau socioéconomique moyen est négative et d'une taille pas du tout méprisable ($r = -0,34$). Cela revient à dire que, plus le niveau social moyen des classes est élevé, plus elles sont homogènes de ce point de vue. Mais cette relation serait davantage quadratique, montrant que les classes à bas niveau socioéconomique moyen sont aussi de tendance homogène. En tous cas, ce rapport négatif est confirmé, quoi que de manière plus atténuée, par la relation entre diversité académique dans la classe et niveau académique moyen de celle-ci. Toutefois, cette relation est davantage forte quand on la regarde discipline par discipline, ayant aussi la même forme quadratique. Ces deux mesures d'hétérogénéité (sociale et académique) restent par ailleurs corrélées mais pas très fortement (+0,11) ce que souligne que la diversité académique au sein des classes est largement plus importante que la diversité sociale (cf. Ramírez, 2007 ; chapitre n°6). Nous nous servirons des deux mesures à la fois pour certaines analyses, de même que des écarts-type des scores dans la classe par discipline.

D'autres variables entretiennent une relation, entre elles, quelque peu étrange à première vue mais qui confirment ce que l'on a vu sur les profils des élèves.

⁵⁵Ce dernier indice est utilisé, par exemple, par Bellei (2009).

Tableau B31

Matrice des Corrélations de Pearson pour différentes mesures possibles reliées à la composition du public par classe (7061 classes urbaines à ≥ 20 élèves) (8^{ème} grade - 2004)

<i>Variable</i>	<i>Moy.</i>	<i>E-T</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Moyenne Années d'étude (Mères)	10.4	2.55															
(2) Écart-type Années d'études (Mères)	2.54	0.5	-0.34														
(3) % de Mères >11 ans d'étude	0.39	0.3	0.95	-0.28													
(4) Moyenne Indice de NSE	0.09	0.8	0.98	-0.36	0.93												
(5) % d'élèves du Quintile V (favorisés)	0.22	0.3	0.92	-0.35	0.93	0.95											
(6) % en BAS Capital Culturel	0.29	0.2	-0.81	0.28	-0.75	-0.81	-0.69										
(7) % en HAUT Capital Culturel	0.31	0.2	0.88	-0.32	0.83	0.90	0.85	-0.82									
(8) % en Bas BEE	0.33	0.1	-0.38	0.09	-0.36	-0.36	-0.32	0.36	-0.34								
(9) % en Haute CSS	0.34	0.1	-0.22	0.11	-0.23	-0.23	-0.27	0.12	-0.20	-0.13							
(10) % de Redoublants	0.16	0.1	-0.56	0.13	-0.54	-0.53	-0.45	0.55	-0.48	0.39	-0.03						
(11) % de Filles	0.50	0.2	0.02	-0.02	0.02	0.02	0.01	-0.05	0.02	-0.08	0.00	-0.13					
(12) % qu'Interagit	0.70	0.1	-0.25	0.09	-0.24	-0.25	-0.23	0.21	-0.22	0.00	0.12	0.10	0.15				
(13) % déclarant qui Aime étudier	0.83	0.1	-0.69	0.32	-0.65	-0.71	-0.71	0.53	-0.66	0.05	0.40	0.28	-0.01	0.22			
(14) Niveau moyen de compétences estimé	0.03	0.6	0.81	-0.24	0.79	0.80	0.76	-0.70	0.76	-0.47	-0.14	-0.63	0.04	-0.18	-0.53		
(15) Écart-type niveau de compétences	0.04	0.9	-0.02	0.11	-0.04	-0.04	-0.09	-0.04	-0.06	0.04	0.05	-0.06	-0.07	0.01	0.06	-0.08	

Source : élaboration de l'auteur. Lecture : entre la variable Moyenne d'années d'études des mères par classe et la moyenne par classe de l'Indice de NSE il existe une corrélation de Pearson équivalente à $r=+0,98$. Sous fond gris sont marquées les variables prioritaires de composition à être utilisées dans les analyses ultérieures sur les acquis scolaires. La variable « % qu'Interagit » correspond à la proportion d'élèves, par classe, qui n'est pas classée sous le profil général « Solitaire ». La variable « % déclarant qu'il Aime étudier » est issue des raisons données pour aller à l'école réunissant les options « très d'accord » et « d'accord ». Le niveau moyen de compétences estimé n'est mis ici qu'à titre illustratif montrant la forte corrélation entre les variables de NSE et les résultats aux tests (ici agrégés). Certaines moyennes sont plus élevées que dans le total (9 412 classes) du fait qu'on a éliminé les établissements ruraux. Il faut rappeler aussi qu'il s'agit, pour la plupart, des valeurs moyennes à partir des moyennes. Les chiffres en italique ne sont pas significatifs (à $p<0,05$).

Par exemple, la proportion déclarant aimer étudier est négativement corrélée avec le niveau social par classe (-0,69), mais aussi avec la proportion d'élèves n'étant pas qualifiée sous le profil « solitaire » (-0,25) (i.e. ceux qu'interagissent). En fait, cela ne fait que refléter ce que l'on a vu en regardant les valeurs moyennes de ces pratiques ou de ces déclarations par décile de NSE au niveau individuel (cf. chapitre n°10). Un point intéressant est que l'hétérogénéité sociale est faiblement et positivement corrélée avec la proportion d'élèves qu'interagissent (+0,09). Il en est de même avec l'hétérogénéité académique (+0,15). Ces deux mesures sont aussi bien reliées, en termes agrégés, avec le pourcentage de ceux qui déclarent qu'ils aiment étudier. Il n'est pas sans intérêt de noter que la relation entre niveau académique de la classe et proportion qui interagit est négative (-0,18). Ce rapport est toutefois mieux représenté par une courbe convexe qui montrerait des légers gains moyens pour des proportions pas trop élevées d'interactions.⁵⁶

Ces constats de niveau agrégé, peuvent tout simplement cacher, encore une fois, des pratiques d'étude de la part des élèves et/ou des pratiques enseignantes assez différenciées. Ceci dans un système où l'on sait que les écoles (et leurs classes) restent fort distinctes en termes socio-scolaires. Autrement dit, la proportion de solitaires par discipline s'accroîtrait là où les élèves sont plus autonomes et savent mieux comment faire face aux défis scolaires, jouissant d'outils plus développés pour ce faire. Ceci expliquerait donc la performance moyenne supérieure, par discipline, pour des basses proportions d'élèves interagissant. Une autre piste explicative est que les interactions entre élèves et les différentes formes d'entraide (ex : prêts de cahiers) restent en partie une réponse à certaines défaillances des enseignants (ex : bas niveaux vis-à-vis de la préparation pour enseigner et/ou de la couverture des programmes). Cela vaudrait dire que ces relations négatives pourraient exprimer, non pas une mauvaise influence des interactions par elles-mêmes, mais le fait que celles-ci seraient plutôt en train de traduire des contextes scolaires moins performants, ou encore, d'exprimer le côté négatif de l'entraide que certains auteurs ont pointé (cf. Van Zanten, 2001). Cela reste un sujet à explorer ultérieurement.

Pour sa part, la proportion des filles ne montre pas de fortes relations avec la majorité des autres variables, sauf dans le cas du redoublement (-0,13) où l'on confirme les constats déjà faits (cf. graphique B12 et tableau B16). Cette variable agrégée entretient aussi un rapport avec le pourcentage d'élèves en condition de faible bien-être à l'école (-0,08) qui pourrait découler du même cadre que la situation antérieure. En effet, ce premier aperçu invite à croire que les filles pourraient être, en moyenne, plus heureuses à l'école, ce qui conforte la littérature. Par ailleurs, aucune relation de niveau classe ne semble se dessiner entre la proportion par genre et

⁵⁶Pourtant, cette perspective change quand on la regarde de manière désagrégée par discipline. En effet, le rapport continue à être négatif et devient plus marqué pour toutes les disciplines sauf en maths, où l'on se dessine une légère courbe concave ascendante.

la proportion d'élèves en haute confiance en soi scolaire. En fait, cette dernière variable entretient des rapports pas trop prononcés avec aucun autre facteur (i.e. $<|0,3|$). Comme attendu, ceux-ci sont négatifs, notamment avec les indices de niveau socioéconomique et culturel, mais aussi avec les résultats scolaires (cf. chapitre n°10). Cela peut cacher le fait que les élèves des établissements à public défavorisé, ayant les moins bons résultats aux tests, déclarent des plus hauts niveaux de confiance en soi scolaire que les autres, de par des mécanismes à explorer (cf. chapitre n°4).

Finalement, notons qu'il existe une corrélation négative entre l'indice de niveau académique ou de compétences des élèves par classe et la proportion d'élèves en condition de faible bien-être à l'école (-0,47). Cela semble cohérent dans une certaine mesure, sachant que la littérature a montré un lien circulaire entre ces deux types de variables et qu'il va dans le sens ici escompté (cf. chapitre n°5). Pourtant, nous voudrions souligner que cela pourrait aussi indiquer que cette variable contextuelle (i.e. la proportion en Bas BEE par classe) cache la relation existante entre scores individuels et scores moyens par classe. En effet, sur une analyse transversale, cette relation risque d'être forcément polluée (i.e. *reflectionproblem*, Manski, 1993 ; cf. chapitre n°2). Malgré cela nous pensons que cette variable mérite de rester dans les analyses. D'abord par les raisons déjà citées plus haut, puis parce que la corrélation ici notée est inférieure à $|0,5|$, ce qui indique que la variable entraîne aussi des informations (i.e. des variations) singulières. Enfin, parce que les rapports entre cette proportion et, par exemple, la moyenne de scolarité des mères, reste semblable dans toutes les catégories d'école.

Après cette brève révision des variables prises de manière agrégée, remarquons qu'une partie de la littérature (e.g. Hoxby et Weingarth, 2006 ; Thrupp et al., 2002) plaide pour un contrôle simultané de plusieurs mesures associées aux groupes (genre, niveau socioéconomique, niveau scolaire, appartenance ethnique, etc.). Ceci semble tout à fait cohérent et souhaitable, notamment lorsqu'on dispose de coupures longitudinales bien établies. Toutefois, nous avons considéré plus judicieux de ne pas trop augmenter le nombre de mesures de composition, d'autant plus que certaines d'entre elles semblent véhiculer des informations très semblables comme le montrent les corrélations les plus élevées ($\geq|0,50|$) du tableau B30. De fait, ce recouvrement ne serait que le reflet de la forte segmentation socio-scolaire du système chilien (cf. chapitre n°6). Nous pensons toutefois que d'autres examens postérieurs pourraient porter sur la comparaison et/ou le remplacement des indicateurs, voire l'incorporation d'autres non utilisés ici. Par exemple, la proportion de redoublants constitue sans doute une mesure du niveau académique des groupes. Ou bien, le capital culturel (livres à la maison) pourrait être pris comme une mesure alternative ou concomitante à la scolarité des mères (e.g. Bellei, 2009).

ANNEXE «C »

**QUESTIONS
et
CHOIX MÉTHODOLOGIQUES**

ANNEXE MÉTHODOLOGIQUE N°1

1. La valeur générale de la modélisation

Notre travail tente de recouvrir les trois niveaux des questions scientifiques : la description, l'analyse causale, et l'analyse interprétative. Nos propos se trouvent, donc, à mi-chemin entre ces catégories que Boudon (2002)⁵⁷ nomme comme la sociologie camérale ou descriptive, destinée à décrire et aider les prises de décisions, et la sociologie scientifique, vouée à éclairer les phénomènes sociaux, leurs mécanismes, sources et significations à travers des explications logiquement et empiriquement contrôlables. Ainsi, pour donner forme à ces objectifs, nous avons été contraints de transformer les réflexions théoriques en stratégies d'analyse empirique basées sur des modèles mathématiques d'analyse des données. Ceci sachant qu'en sciences sociales « *la modélisation multivariée constitue un substitut de l'expérimentation* » (Duru-Bellat et Suchaut, 2005 ; p. 17).

Dans une synthèse attentive de la valeur de la modélisation statistique en sciences sociales, Bressoux (2008)⁵⁸ note que parmi les traits distinctifs d'un modèle scientifiquement valable on peut signifier qu'il correspond à une « *représentation schématique qui s'efforce de reproduire les propriétés fondamentales d'un objet complexe* » (p. 17). Le modèle ne peut donc être qu'une « *simplification de la réalité* » (p. 18).⁵⁹ Puis, cet auteur ajoute qu'un modèle se doit d'être exprimé « *dans un langage aussi rigoureux que possible* » (p. 18) sachant qu'il est forcément un intermédiaire sur qui « *nous déléguons la fonction de connaissance* » (Bachelard, 1979 cité par Bressoux, 2008, p. 18). Au total, le rôle d'un modèle, c'est de spécifier un ensemble de relations hypothétiques entre variables, en exprimant des relations causales ou simplement corrélationnelles entre elles. Sa formalisation est aussi appelée spécification

⁵⁷ Boudon, R. ; 2002 ; *Sociology That Really Matters* ; European Sociological Review, Vol.18, n°3 ; pp. 371-378

⁵⁸ Bressoux, P. ; 2008 ; Modélisation statistique appliquée en sciences sociales ; De Boeck, Paris ; 443 pp.

⁵⁹ Notons que, pour le présent travail, le choix d'une approche voie fonction de production en éducation reste lié à trois grands facteurs qui sont, comme on l'a dit, théoriques et pratiques à la fois. D'abord, l'usage des données à grande échelle (Boudon et Bourricaud, 1982), permettant de tirer de conclusions plus ou moins généralisables. Puis, le fait que certaines problématiques, telles que celles liées aux interactions entre élèves et à leurs perceptions subjectives n'ont pas été abordées souvent sur des grandes populations (cf. chapitre n°3). Enfin, l'intérêt de faire, si possible, un apport singulier mais en phase avec la littérature dominante au Chili. Pourtant, à l'opposé de certains types d'études économétriques qui assurent que la réalité et les comportements humains peuvent être véritablement réduits à une équation, nous n'assignerons à ce type d'outil qu'une valeur analytique des relations probables entre facteurs. Conséquemment, suivant Duru-Bellat (2007b), disons que si l'approche quantifiée est un moyen parmi d'autres pour construire des connaissances en sociologie, il peut aussi servir à se préserver des prises de positions trop idéologiques. Ceci sachant que des telles analyses seront toujours limitées et portées par leur mode de construction (Duru-Bellat et Suchaut, 2005).

(Gujarati, 2004).⁶⁰Certainement, une telle démarche globale a des limitations importantes dérivées, d'une part, de la manipulation de données toujours imparfaites et entachées des erreurs à des étapes successives, et d'autre part, découlant des risques encourus de par la réduction de la réalité et de par la difficulté à saisir le sens et l'ampleur des liens entre facteurs.

1.1 Les modélisations utilisées sous forme de fonction de production

Dans la recherche en éducation, nous sommes habituellement confrontés à des modèles qui prennent la forme de l'équation suivante :



$$Y_{ijs(t1)} = \alpha + \beta_1 E_{i(t1)} + \beta_2 F_{i(t1)} + \beta_3 CE_{js(t1)} + \beta_3 C\bar{P}_{js(t1)} + \beta_4 S_{s(t1)} + \varepsilon_{ijs(t1)}$$

235094 élèves

7061 classes

3565 écoles

Où, ici, Y correspond au score en langage de l'élève i placé dans la classe j de l'école s au moment $t1$; et où E , F , CE , CP et S correspondent respectivement à des vecteurs contenant des variables liées à l'élève ; à sa famille ; aux caractéristiques de la classe est de l'enseignant qu'il fréquente ; aux caractéristiques (moyennes) du groupe des pairs dans la classe ; et aux caractéristiques de son école. Enfin $\varepsilon_{ijs(t1)}$ représente le terme d'erreur du modèle (i.e. la part des scores inexpliquée par les variables incluses). Au côté gauche de Y est représentée une distribution des scores au sein de la population d'élèves étudiée et pour la discipline concernée et qui nous rappelle que le score de Y prédit par le modèle correspond à l'élève se situant dans la moyenne de cette distribution.

Nous avons voulu noter cette équation générale d'estimation d'une fonction de production des acquis scolaires (à un moment donné), car celle-ci nous permet d'illustrer les différents problèmes auxquelles on doit faire face à l'heure de penser à la valeur explicative de nos résultats. D'abord, le fait que nous avons des unités (élèves, classes et écoles) qui diffèrent en nombre et qui sont, à la fois, des niveaux distincts d'analyse. Puis, le fait que certains groupes de variables sont liées à d'autres : par exemple, le choix d'école n'est pas du tout étrange aux caractéristiques des familles des élèves, le groupe de pairs non plus d'ailleurs. Enfin, le fait que des modèles comme celui représenté par cette équation ne travaillent qu'avec un moment de la

⁶⁰Gujarati, D. ; 2004 ; *Économétrie (Trad. de la 4^{ème} Ed. américaine par Bernard Bernier (Basic Econometrics))* ; De Boeck, Bruxelles ; 1009 pp.

distribution des résultats (la moyenne), laissant forcément de côté les autres « types d'élève ». Notons enfin que l'approche temporelle n'était pas véritablement disponible, en dépit des gains qu'elle représente pour toute estimation scientifique. De ce fait, nous allons rester sous une perspective transversale (i.e. au temps $t1$) sans pouvoir ajouter des variables préalables à ce moment (i.e. provenant du temps $t0$). De cette manière, à partir des trois constats notés, nous allons choisir et justifier différentes solutions possibles parmi les outils d'analyse disponibles dans la littérature. Ces différents choix seront adressés dans les parties et les sous-parties des annexes méthodologiques ici déployés.

1.2 Les phénomènes dits de contexte et la modélisation multiniveaux : association incontournable

Nous avons déjà noté que l'un des problèmes majeurs soulignés par la littérature en sociologie (cf. Bressoux, 2008 ; Duru-Bellat et Suchaut, 2005 ; Dreeben, 2000) ou en économie (cf. Manski, 2000 ; Brock et Durlauf, 2001 ; Hanushek et al., 2003) consiste à mettre en relation, à des fins analytiques, les caractéristiques du contexte et les *outcomes* individuels, soient celles que l'on peut penser comme mobiles, tels que les influences sociales mutuelles entre individus (Manski, 2000) ; soient celles correspondant à des données environnementales plus stables dans le temps.⁶¹ Comme cela est le cas pour toute modélisation sous forme de fonction de production, c'est la présence de co-variations entre variables de ces deux niveaux qui est à la base des analyses possibles.⁶² Bien sûr, on ne répètera jamais assez que corrélation (i.e. l'existence d'une co-variation) ne veut pas dire, *per se*, causalité. Ceci reste ainsi l'un des problèmes majeurs de la modélisation mathématique des phénomènes sociaux puisque les éléments qui font supposer des relations causales doivent être apportés par le chercheur en

⁶¹ Si le premier type peut être exemplifié par le cas des élèves dont l'enseignant ou les camarades changent d'une année sur l'autre, le deuxième type serait mieux représenté, par exemple, par les moyens dont une école dispose. Cette distinction, d'après l'instabilité temporelle de situations contextuelles, peut être fort utile d'un point de vue empirique mais elle est aussi théoriquement importante, notamment sur des données diachroniques.

⁶² Ces co-variations sont alors la base de ce qu'on connaît comme études corrélationnelles. Celles-ci fondent leur pouvoir explicatif notamment sur des analyses de variance et dans la mesure où nous rencontrons dans la réalité des situations où il existe de la variabilité. Toujours est-il que la réalité peut nous montrer aussi des situations importantes à comprendre mais avec très peu de variabilité (Meuret, 1995). Le raisonnement de base est simple : si l'on observe que deux variables co-varient on peut donc émettre trois hypothèses : (a) que l'une est la cause de l'autre, mais que la deuxième n'a aucune influence sur la première ($A \rightarrow B$) ; (b) que l'une agit sur l'autre, et donc, qu'elle peut être supposée comme cause, mais que cette deuxième agit aussi sur la première ($A \leftrightarrow B$) ; et (c), qu'il existe une troisième variable qui agit à la fois sur les deux autres ($C \rightarrow A : B$). Dans le cas des analyses multivariées on cherche donc à dégager l'effet net et/ou la relation de variation nette entre variables en considérant plusieurs facteurs à la fois (i.e. en contrôlant par ces facteurs) pour s'approcher ainsi de la réalité. Ce type de raisonnement dit *toutes choses égales par ailleurs* (*othertingsbeingequalou ceterisparibus*) tente d'évaluer l'impact d'une variable sur une autre, mais affranchi de toutes les relations que la première entretient avec les autres variables présentes dans un modèle qui est supposé capturer le monde réel.

dehors de résultats empiriques qu'il trouve (Duru-Bellat et Mingat, 1997).⁶³ Pourtant, l'un des principaux problèmes qui se posent au moment de modéliser et d'entreprendre l'opérationnalisation du contexte – à part celui fondamental de le définir et de le justifier en termes d'unités d'analyse – correspond à ce qu'on l'on connaît comme l'erreur écologique ou encore, le biais d'agrégation (cf. Bressoux, 2008 ; Blakely et Woodward, 2000⁶⁴).⁶⁵ Bressoux (2008) note que pour bien maîtriser ce type de phénomène, différents chercheurs ont prôné, depuis le début des années quatre-vingt, le besoin d'analyses capables de rendre compte d'une multiplicité de niveaux de manière simultanée. Cette quête aurait mené, entre autres, à l'émergence d'un type de modèle dès lors appelé « multiniveaux ».⁶⁶ Ces modèles sont également connus en français comme des modèles hiérarchiques linéaires ou encore comme des modèles à coefficients aléatoires (p. 290). Pour leur part, Raudenbush et Bryk (1992) prennent l'appellation *hierarchical linear models* (HLM) (introduite par Lindley et Smith en 1972, cités par Raudenbush et Bryk, 1992 ; p. 3-4) fait de l'importance capitale, pour différents domaines de recherche, de cette caractéristique structurelle des données (*i.e.* l'emboîtement hiérarchique). Ray (2004)⁶⁷ ajoute qu'ils sont aussi appelés modèles mixtes et note qu'en économétrie, on parle surtout des modèles à effets fixes et à effets aléatoires (p. 206).

⁶³D'autres problèmes à considérer sont les propriétés mathématiques des variables qui autorisent ou pas certains traitements (type d'échelle, normalité ou pas, caractère aléatoire, etc.). De même, en considérant la distinction entre variables à expliquer et variables explicatives et/ou de contrôle, il est nécessaire de faire la distinction entre variables exogènes et endogènes (Gujarati, 2004). Si les premières sont prédéterminées, c'est à dire qu'elles prennent des valeurs ou des catégories en fonction des facteurs qui se trouvent en dehors du modèle, les deuxièmes peuvent être définies comme des variables dont les valeurs et caractéristiques dépendent ou se trouvent déterminées par, au moins, une autre variable présente dans le modèle. C'est notamment le cas du choix de l'école (Voir Annexe sur ce sujet). Il faut avouer aussi que dans la réalité des phénomènes éducatifs, les variables ne sont que très rarement tout à fait indépendantes les unes des autres.

A son tour, ces distinctions correspondent à ce qu'on peut appeler des variables observées. Toutefois, il existe aussi des variables dites latentes qui ne sont pas directement observées ni observables, mais dont on suppose que les mesures disponibles les reflètent. C'est bel et bien le cas des échelles que nous avons construit pour exprimer le Bien-être à l'école ou la Confiance en soi scolaire. Mais c'est aussi le cas de la prédisposition à choisir un type d'école donné et que nous avons tenté de capturer avec l'Inverse du Ratio de Mills. Au total, nous avons trois principales sources d'endogénéité: a) les erreurs de mesure sur les variables explicatives ; b) la causalité dans les deux sens lorsque la variable explicative agit sur la dépendante et inversement ; et c) le biais de variable omise lorsqu'une variable non incluse dans le modèle est corrélée avec au moins une des variables explicatives et agit alors à la fois tant sur celle-ci que sur la variable à expliquer.

⁶⁴ Blakely, T.A.; Woodward A.J.; 2000 ; *Ecological effects in multi-level studies* ; Journal of Epidemiol Community Health, Vol.54, n°5 ; pp367-374

⁶⁵Suivant Blakely et Woodward(2000) qui citent à Diez-Roux et al. (1997) on rencontre des erreurs d'analyse du type écologique, lorsque : (a) l'unité d'analyse est le groupe mais l'inférence est faite au niveau individuel ; (b) l'unité d'analyse est l'individu mais l'inférence est faite au niveau groupal ; (c) l'unité d'analyse est l'individu mais l'on ne prend pas en compte les variables groupales importantes ; (d) l'unité d'analyse est le groupe mais on exclu les variables individuelles importantes. Parmi les différents types d'erreur d'inférence, c'est le premier qui nous intéresse davantage. Il correspond à ce que l'on connaît comme *ecological fallacy* (erreur écologique), terme popularisée d'après Robert (1950, cité par Hox, 2002).

⁶⁶ Pour une discussion sur l'orthographe « multiniveaux » (par opposition à « multiniveau »), voir Ray (2002).

⁶⁷ Ray J.-C.; 2004 ; *Transferts sociaux et gains d'activité des jeunes adultes européens. Une application des modèles multiniveaux* ; Économie et Prévision, Vol.3-4, n°164 ; pp. 189-210

Bref, de nos jours et au-delà des nuances entre dénominations et usages, cet outil semble incontournable pour tout travail portant sur l'école.

Pour sa part, Bickel (2007)⁶⁸ voit dans cette évolution une conséquence logique du développement permanent des outils visant à dépasser les limites des modèles MCO (Moindres Carrés Ordinaires, ou *OLS*, *Ordinary Least Squares*), dont le problème du traitement des contextes reste l'un des enjeux majeurs. Ainsi et en termes plus concrets, des nombreuses recherches, notamment à tonalité sociologique (e.g. Raudenbush et Bryck, 1992⁶⁹; Bressoux et al, 1997⁷⁰; Monseur et Lafontaine, 2005⁷¹; Bressoux, 2008), ont mis en évidence sous différents angles l'inadaptation des modèles MCO pour rendre compte des situations où différents niveaux d'analyse et de regroupement de données se superposent.⁷²Hox (2002)⁷³ parle même de *multilevelresearch* (p. 1) pour se référer à ces problèmes.

Parmi les principales caractéristiques qui expliquent l'intérêt des modèles multiniveaux par rapport à d'autres techniques statistiques, nous pouvons citer :

1.2.1 L'amélioration des estimations au niveau individuel par le biais d'un contrôle adéquat des variables environnementales (Raudenbush et Bryk, 1992), mais surtout d'un bon établissement des niveaux d'analyse. Autrement dit, les MMN dépassent les modèles MCO dans le sens où ces derniers ne sont pas capables de contrôler la structure hiérarchique des données (*i.e.* des élèves placés dans des classes appartenant à des écoles à leur tour placées dans des quartiers, etc.) et supposent une seule équation mono-niveau qui mélange des données d'environnement (*i.e.* les caractéristiques des classes ou des écoles qui sont communes à plusieurs élèves à la fois), avec des données individuelles (*i.e.* les caractéristiques associées aux élèves eux-mêmes). En effet, à la différence des modèles MCO ou encore, des modèles Moindres Carrés Généralisés (MCG ou LSG), les MMN « *utilisent simultanément toute l'information disponible à plusieurs niveaux d'analyse* » (p. 13 ; Arrègle, 2003).⁷⁴

⁶⁸ Bickel, R.; 2007 ; *Multilevel analysis for applied research: It's just regression !* ; Guilford Press, NY ; 355 pp.

⁶⁹ Raudenbush S.W. ; Bryk A.S. ; 1992 ; *Hierarchical Linear Models. Applications and data analysis methods* ; Coll. *Advance Quantitative Techniques in the Social Sciences Series* ; Sage Publications, London ; 265 pp.

⁷⁰ Bressoux, P. ; Coustère, P. ; Leroy-Audouin, C. ; 1997 ; *Les modèles multiniveau dans l'analyse écologique : le cas de la recherche en éducation* ; *Revue Française de Sociologie*, Vol.38 ; pp. 67-96

⁷¹ Monseur, C ; Lafontaine, D. ; 2005 ; *Methodological Issues Raised by Equity Indicators Derived from Multilevel Analyses* ; Working Paper ; 24 pp.

⁷² Pour une synthèse des techniques possibles et parfois encore appliquées pour rendre compte des phénomènes contextuels dans le cadre des modèles MCO, voir Bressoux et al, 1997.

⁷³ Hox, J.J. ; 2002 ; *Multilevel analysis. Techniques and applications*. LEA, London ; 304 pp

⁷⁴ Arrègle J.-L. ; 2003 ; *Les modèles hiérarchiques: principes et illustration* ; *M@n@gement*, Vol. 6, n°1 ; pp. 1-28

En revanche, les modèles MCO tendent à sous-estimer la variance des coefficients d'impact des variables contextuelles (et donc, les écarts types qui servent de base au calcul de leur signification statistique). Ils multiplient donc artificiellement les données contextuelles pour les placer au niveau du nombre d'individus, avec à la clé une véritable « *multiplication miraculeuse des unités d'analyse* » (Bressoux, 2008 ; p. 285). Hox (2002) résume le problème ainsi : « *standard statistical tests lean heavily on the assumption of independence of the observation. If this assumption is violated (and a multilevel data this is almost always the case) the estimates of the standard errors of conventional statistical tests are much too small, and this results in many spuriously 'significant' results* » (p. 5). En outre, les estimateurs de base utilisés par les MMN (i.e. à partir de la méthode de Maximum de Vraisemblance ou *Maximum Likelihood*) sont plus efficaces que ceux issus des MCO : « *the estimates for the regression coefficients are generally unbiased, for OLS, GLS and Maximum Likelihood (ML) estimations (...). OLS estimates are less efficient because they often have a larger sampling variance* » (Hox, 2002 ; p. 173). Selon Hox, « *an advantage of the Maximum Likelihood estimation method is that it is generally robust, and produces estimates that are asymptotically efficient and consistent. With large samples, ML estimates are usually robust against mild violations of the assumptions, such as having non-normal errors* » (p. 37-38).⁷⁵ Toutefois, il pointe que l'on peut faire appel à deux fonctions pour estimer la vraisemblance : le FML (*full maximum likelihood* nommée ML sur SAS®) et le RML (*restricted maximum likelihood*). Bressoux (2008 ; p. 317) explique que la première considère tant la partie fixe comme la partie aléatoire du modèle et se centre sur l'ensemble de l'échantillon. Pourtant, cette approche biaise un peu l'estimation des composantes de variance (partie aléatoire). En effet, Hox (2002) pointe que « *estimates of the residual error at the lowest level are generally very accurate. The group level variance components are sometimes underestimated (...). It is clear that with increasing sample sizes at all levels, estimates and their standard errors become more accurate* » (p. 174). La deuxième approche (RML) est centrée sur les résidus de l'échantillon et les composantes de variance en corrigeant les défauts de la première, mais en perdant un peu de précision dans les estimations (Bressoux, 2008). De plus, le choix d'une approche ou de l'autre n'est pas sans conséquences sur l'interprétation des statistiques exprimant la qualité totale d'un modèle et permettant de le comparer avec d'autres (i.e. la statistique de déviance). Pourtant, Bressoux (2008) note que la littérature semble converger vers le fait que si le nombre des groupes est important, les biais de l'approche FML tend à disparaître.⁷⁶

1.2.2 Une autre caractéristique des MMN découlant du point antérieur, c'est qu'ils peuvent rendre compte des corrélations entre unités appartenant à un même espace ou un même

⁷⁵ Par consistance, il faut comprendre que l'estimation a la propriété de converger vers la vraie valeur dans la population totale au fur et à mesure que l'échantillon s'accroît vers l'infini.

⁷⁶ Hox (2002) ajoute que, « *if there is strong interest in cross level interactions, the number of groups should be larger (...). If there strong interest in the random part, the variance and covariance components and their standard errors, the number of groups should be considerably larger* » (i.e. ≥ 100) (p. 175).

environnement. De fait, comme le note Bressoux (2008), dans une situation d'emboîtement « *les observations ne sont pas indépendantes les unes des autres* » (p. 276). En revanche, les modèles MCO ne peuvent pas contrôler, *a priori*, ce phénomène. Celui-ci reste une source importante de hétéroscédasticité (*ex* : le caractère non-aléatoire des résidus ou erreurs d'estimation par régression linéaire, condition nécessaire à la robustesse des estimations). Ce manque de homocédasticité ou présence de hétéroscédasticité est attribuable, en grande partie, aux vraisemblances manifestes dues au regroupement des élèves en clusters ou groupes bien précis (e.g. avoir le même prof, les mêmes infrastructures, etc.).

Cependant, il est intéressant de noter que des nombreuses recherches en économie de l'éducation font appel à une correction des écarts types des coefficients estimés par régression MCO. Cette correction, nommée « écarts-type robustes » est proposée spécialement par le logiciel STATA® et elle vise à mieux contrôler la signification statistique des effets estimés. Ces écarts-type robustes, connus aussi comme estimateurs de variance « Huber/White », permettent de considérer les résidus individuels non indépendantes à l'intérieur des clusters choisis (i.e. classes ou écoles) (pp. 336-341 ; STATA, 2003).⁷⁷ Ainsi, les observations au niveau des individus sont assumées indépendantes entre clusters mais pas forcément au sein de ceux-ci. Cela constitue, bien entendu, un raisonnement du type multiniveau du moment où lesdits clusters représentent des unités qui regroupent de manière hiérarchique aux unités-individus. On parle alors d'estimations MCO « robustes » (Leblond, 2003).⁷⁸ Mais si cette approche permet de mieux contrôler la signification statistique des effets estimés par le biais de pénaliser les écarts-types des variables contextuelles (i.e. liés à une unité supérieure), elle ne touche pas les coefficients eux-mêmes.⁷⁹ Nous avons utilisé cette technique à des fins exploratoires et comparatives, de fait qu'elle est très répandue dans les recherches en économie de l'éducation au Chili.⁸⁰

Cependant, Hox (2002), en se référant à l'intérêt des corrections de la variance par clusters et l'utilité des modèles MMN indique que « *some of the corrections procedures developed for cluster*

⁷⁷STATA ; 2003 ; Base Reference Manual, Vol.3 (N-R) Release 8 ; Stata Press Publication, Texas ; 464 pp.

⁷⁸ Leblond, S. ; 2003 ; *Guide d'économétrie appliquée pour Stata* ; Université de Montréal ; 50 pp.

⁷⁹ La méthode d'estimation des variances des coefficients robustes (erreurs standards au carré) est nommée aussi d'Eicker-White et son effet pratique est celui de gonfler cette variance (Leblond, 2003) changeant ainsi la signification des coefficients.

⁸⁰Consignons enfin qu'il existe d'autres méthodes dans le cadre des MCO pour donner de la robustesse aux estimations. Notamment l'identification et le contrôle des données-individus dits *outliers* ou aberrants et qui exercent de ce fait une influence particulière sur les estimations. On parle alors des Moindres Carrés Pondérés ou WLS (*Weight Least Squares*) qui contrôlent aussi l'impact de l'hétéroscédasticité sur la signification des estimations. Tant le logiciel SAS® que STATA® proposent des procédures pour y parvenir. Toujours est-il que la correction de robustesse par le biais des écarts-types ajustés au niveau classe ou école (mais pas les deux en même temps) chez STATA®, nous semble mieux adaptée à nos propos car celle-ci tient compte de la nature emboîtée des données si l'on est contraint de travailler avec des modèles MCO.

and other complex samples are quite powerful (...). Actually, in principle these correction procedures could also be applied in analyzing multilevel data, by adjusting the standard errors of the statistical tests. However, in general the intraclass correlation and hence the effective N is different for different variables. In addition, in most multilevel problems we have not only clustering of individuals within groups, but we also have variables measured at all available levels » (p. 6). Toutefois, cet auteur pointe aussi qu'en présence des variables dépendantes trop éloignées d'une distribution normale, la correction par « écarts-types robustes » est conseillée au sein des modèles multiniveaux (cf. Chap. 11, partie 2).⁸¹

1.2.3 Un troisième avantage des MMN sur les modèles plus couramment utilisés correspond à la distinction théorique, mais aussi opérationnelle, entre *effets fixes* et *effets aléatoires*. Si les premiers sont des effets « *non stochastiques qui révèlent d'un ensemble fini de modalités d'un facteur* » à étudier (p. 286 ; Bressoux, 2008), les seconds peuvent être compris comme « *un vaste ensemble (...) dont un échantillon seulement a été observé* » (p. 287).⁸²Falissard (1998)⁸³ indique que dans toute analyse de variance, un facteur « *peut être à effet fixe ou à effet aléatoire* » (p. 87) dont le premier correspond à une variable qui ne change pas « *quand l'expérience considérée est réalisée à plusieurs reprises* » (p. 87).⁸⁴Autant Bressoux (2008) que Falissard (1998) soulignent que le

⁸¹ Notons que la Proc MIXED de SAS® propose l'option « *empirical* » pour que les écarts-types des effets fixes du modèle soient estimés selon la correction Hubert/White (SAS User's Guide, 2004 ; p. 2676).

⁸²La régression conventionnelle donne seulement des estimations ou prédictions d'une variable Y pour un individu (i) placé à un certain niveau (j) de la forme suivante : $Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$ où ε_{ij} possède une variabilité constante (i.e. égal pour tout i quel que soit j) et où β_0 et β_1 représentent des estimations moyennes (i.e. sur l'ensemble de la population) de l'impact moyen de X_{ij} sur Y_{ij} (i.e. qui produit \bar{Y}_{ij}) et pensées comme étant fixes à travers le total de cette même population. Pourtant, si les sujets i sont assignés de manière aléatoire aux différents niveaux j de X , alors on pourrait devoir évaluer la part aléatoire et non fixe de l'impact de X sur Y pour tout sujet i placé en j .

La formulation d'un tel modèle peut s'écrire comme suit : $Y_{ij} = (\beta_0 + a_0) + (\beta_1 + a_1)X_{ij} + \varepsilon_{ij}$ où a_0 et a_1 représentent respectivement la déviation du sujet ij par rapport à la constante et la pente β_0 et β_1 commune à tous les ij . Cela peut se reformuler comme $Y_{ij} = (\beta_0 + \beta_1 X_{ij}) + (a_0 + a_1 X_{ij} + \varepsilon_{ij})$ ce qui permet de visualiser plus clairement la partie fixe ou commune à tous les individus i quelque soit le niveau j de X , et la partie des coefficients aléatoires qui varie selon les j ou les i . (p. 355 ; Panik, 2009).

Enfin, on assume que ε_{ij} est indépendante pour tout i et distribué de manière égale à travers tous les j (distribution $N(0, \sigma_\varepsilon^2)$). Pour leur part, les coefficients a_0 et a_1 sont assumés être issus aléatoirement d'une distribution de possibles coefficients. On peut assumer aussi que la variation des pentes et des constantes est indépendante (Panik, 2009 ; p. 355) ou bien qu'elle exprime des co-variations qui prennent la forme d'une matrice de variance-covariance du type bloc-diagonale (Bressoux, 2008 ; p. 297). Ainsi, les modèles mixtes voient les possibles regroupements ou clusters des observations comme étant une déviation aléatoire d'une population infinie d'équations de régression. En revanche, les effets fixes sont supposés constants. Ceci donne deux sets de paramètres alors utilisés pour déterminer la probabilité de distribution des données.

⁸³Falissard, B.; 1998 ; Comprendre et utiliser les statistiques dans les sciences de la vie (2^{ème} Ed.) ; Masson, Paris ; 344 pp.

⁸⁴Si l'on parle de l'école, une autre manière d'exprimer la différence entre effets fixes et aléatoires tiendrait au fait que les premiers ne modifient pas leur sens, magnitude et signification quel que soit l'emplacement des élèves. Ainsi, par exemple, dans n'importe quelle classe ou école, l'impact du genre de l'élève sur ses apprentissages en maths ou en langage serait un impact demeurant plutôt inchangé. En revanche, si nous postulons que dans

chercheur doit faire attention en attribuant à chaque facteur sa position, soit comme effet fixe, soit comme effet aléatoire. Dans le cas des premiers, on est « *intéressé par des moyennes, tandis que dans le cas des effets aléatoires on est intéressé par des variances. Or, on ne fait pas les mêmes inférences (...)* » (Bressoux, 2008 ; p. 288).

Raudenbush et Bryk (1992) soulignent la capacité des MMN à modéliser et à tester l'existence des effets d'un niveau sur l'autre (*cross-level effects*). Ainsi, les MMN permettent, de plus, d'analyser les interactions entre variables de différents niveaux. En effet, les MMN opèrent sur la base que chaque coefficient estimé pour une variable possède (potentiellement) une partie fixe ou commune à tous les individus exposés à celle-ci et des parties aléatoires attachées à des niveaux supérieurs d'emboîtement. Ces parties s'expriment donc comme une erreur ou écart à l'effet moyen ou général dont nous pouvons qualifier l'importance et la signification. Par exemple, l'impact d'une pratique donnée de la part des enseignants aurait un effet fixe commun à toutes les classes avec des enseignants déclarant ladite pratique, et une partie attachée à chaque classe (ou à chaque enseignant) qui viendrait s'ajouter à l'antérieure. Ceci admet donc des possibles variations d'un environnement à l'autre, des variations qu'il est possible d'identifier en les assignant une magnitude et une signification statistique pour ensuite les interpréter. Plus encore, ces variations peuvent être modulées par d'autres variables. La pratique des enseignants déjà commentée peut varier son impact selon la composition de la classe (sociale, de genre, etc.) illustrant une interaction à un même niveau d'analyse (la classe). Plus encore, l'impact de cette pratique peut interagir avec les caractéristiques personnelles des élèves, se montrant plus efficace pour certains que pour d'autres, illustrant une interaction entre niveaux.⁸⁵ Pourtant, dans la pratique des analyses on ne prend que quelques variables sous la perspective d'un effet aléatoire (Bickel, 2007) ; autant par souci de parcimonie que parce que certaines variables ne varient guère d'une unité à l'autre, et que l'ajout des pentes aléatoires complexifie les analyses.

Consignons enfin, qu'une autre méthode aussi utilisée par la littérature économique en éducation est celle dite à « effets fixes », toujours dans le cadre des régressions MCO (cf. McEwan, 2003). Cela consiste à considérer comme des variables muettes rapportées à une référence l'ensemble des écoles, circonscriptions, etc., de manière à dégager les effets des autres variables de tout ce qui peut tenir aux spécificités des écoles, académies, régions, etc.

certaines classes ou écoles, cet effet peut varier significativement de fait des caractéristiques attachées à la classe ou à l'école (i.e. qu'il change de sens et/ou de magnitude ou signification) nous serions alors en présence d'un effet aléatoire.

⁸⁵Rappelons que dans le cadre des modèles MCO simples, tous les effets sont, par construction, supposés être communs à toutes les classes et enseignants et à toutes les écoles. La correction par écarts-types robustes (MCO) ainsi que d'autres alternatives comme la récupération des résidus des MCO (Bressoux et al., 1997) ne permettent pas non plus de démêler ou de saisir ces éventuelles interactions entre les variables des différents niveaux.

McEwan (2003), en travaillant sur des données semblables aux nôtres, met en avant comme une des raisons pour utiliser cette technique, le contrôle des phénomènes d'endogénéité et/ou de variables omises à l'heure d'évaluer des éventuels effets de pairs. Par ce biais, l'auteur s'adresse au problème de la non-indépendance des erreurs, très probablement corrélées avec les caractéristiques des pairs au sein des classes. Il indique : « *let us assume that the families and students who choose a particular school (and peer group) share similar preferences, motivations, and other unobserved characteristics that influence the outcomes of their children. In this case, the error term has a school-specific component that is constant for each individual in a school, and an idiosyncratic component that varies across individuals within classrooms and schools* » (p. 132). Bien entendu, nous sommes en présence, à nouveau, d'un raisonnement qui tente de s'adresser aux problèmes des données hiérarchisées dans le cadre des estimations MCO. En définitive cette démarche revient à prendre les macro-unités comme des facteurs fixes, à modalités finies et uniques. Or, ces niveaux ne peuvent pas être vus comme un facteur fixe quitte à restreindre la généralisation des résultats au seul échantillon analysé. En effet, comme le fait remarquer Bressoux (2008), dans un tel modèle « *seule l'appartenance ou non au groupe est indiquée* » (p. 288), et bien qu'il indique un effet propre à chaque unité – exprimé par sa constante commune à tous les individus y appartenant –, « *l'inconvénient d'une telle spécification c'est qu'elle ne dit rien sur les mécanismes précis qui sont à l'œuvre* » (p. 288). De fait, ce type de spécification interdit toute inclusion des facteurs caractérisant les macro-unités considérées dans le modèle, ce qui revient à ne pas pouvoir accomplir l'objectif de fond de toute analyse contextuelle (i.e. comprendre comment et dans quel degré les variables du niveau supérieur affectent les *outcomes* individuels).⁸⁶ Malgré ce qu'on vient d'indiquer, notons que notre échantillon (cf. annexe « A ») recouvre la presque totalité du pays, ce qui devrait nous annoncer des différences pas très importantes entre les estimations MN et celles MCO.

1.2.4 Finalement, Raudenbush et Bryk (1992) mettent en exergue la possibilité de diviser les composantes de variance et de covariance à partir des données emboîtées, même si elles sont du type non balancé (*nested unbalanced data*). Une telle partition de la variance associée aux différents niveaux pris en compte permet de mieux comprendre les différences associées et/ou dues aux individus (e.g. *within-school or within-classes variance*) et celles placées à des niveaux organisationnels supérieurs (*between-school or between-classes variance*). Cela revient à dire, que, même en ayant des estimations semblables sur les effets des variables entre un MMN et un

⁸⁶ Bickel (2007) insiste sur le fait que cela ne serait pas exclusivement un problème théorique (i.e. prendre les unités d'emboîtement comme des variables de niveau individuel ne l'étant pas en fait), mais qu'il entraîne aussi une perte d'information et de possibilités de jugement sur les éventuels effets liés au contexte organisé par ces unités. D'ailleurs, et mis à part les difficultés d'encodage et le caractère grossier de cette démarche, Bickel (2007, p. 77) elle engendre des problèmes additionnels liés au fait de transformer les unités du niveau supérieur en variables muettes. D'abord, le gonflage artificiel des degrés de liberté pour les estimations, puis, l'émergence des problèmes additionnels de colinéarité. Toutefois, cette technique peut s'avérer tout à fait indiquée dans certains cas (Bickel, 2007).

autre du type MCO, les premiers nous donnent davantage d'information utile. En effet, cette division de la variance en intra-groupe (*within*) et inter-groupes (*between*) peut être qualifiée comme un phénomène intéressant en soi (Chaix et Chauvin, 2002).⁸⁷Celle-ci permet d'associer la variabilité globale à chaque niveau d'analyse et donc, aux variables qu'y sont attachées. Ceci reste impossible dans le cadre habituel des MCO et des analyses de la variance. Certes, on peut regarder les variations du R² selon qu'on ajoute ou retire des variables de différents niveaux, mais jamais de manière précise quand celles-ci sont toutes présentes dans le modèle. Plus encore, la procédure d'analyse attachée aux MMN fournit un cadre de référence à travers un modèle dit vide et qui correspond à une « *analyse de variance du type ANOVA avec effets aléatoires* » (Bressoux, 2008 ; p. 291).

1.3 Le coefficient de corrélation intra-classe

Monseur et Lafontaine (2005) insistent sur le fait qu'en éducation il existe trois niveaux incontournables de variance à saisir et à expliquer : a) entre écoles ; b) entre classes au sein des écoles ; et c) entre élèves au sein des classes. De fait, les classes ne sont pas égales aux écoles, ni en termes opérationnels ni en termes théoriques. Ces auteurs soulignent l'importance du coefficient *Rho* mesurant la corrélation inter-classes ou intergroupes, c'est-à-dire, le degré de variabilité d'une mesure donnée attachée au niveau supérieur choisi (i.e. supérieur au niveau individu). Ainsi, plus cette corrélation est importante (i.e. le *Rho* se situe plus près de « 1 ») plus cela veut dire que les scores varient davantage d'une école à l'autre plutôt qu'entre les élèves au sein des écoles, qui auraient, du coup, des fortes similitudes à l'intérieur de celles-ci.⁸⁸Ce coefficient est souvent utilisé dans les études internationales (cf. PISA) comme une mesure d'équité et de différenciation au sein des systèmes éducatifs (Monseur et al., 2005). Au total, ce démêlage de la variance en niveaux entraîne des fortes répercussions sur les possibilités d'analyse de l'impact de chaque variable, notamment dans la mesure où l'inclusion de celles-ci dans un modèle apporte à réduire la variance inexpliquée de chaque niveau (écoles, classes, individus, etc.). Dans le cas d'une recherche à trois niveaux, Hox (2002) pointe que ces coefficients de corrélation de niveau supérieur peuvent être assumés de deux manières : selon si l'on place dans le numérateur, la variance des deux coefficients correspondant aux unités de niveau supérieur – notamment parce qu'ils sont emboîtés (ex : les

⁸⁷ Chaix B. ; Chauvin P. ; 2002 ; *L'apport des modèles multiniveaux dans l'analyse contextuelle en épidémiologie sociale : une revue de la littérature* ; Revue d'épidémiologie et de santé publique, Vol. 50, n°5 ; pp. 489-499

⁸⁸Postlethwaite (1995, p. 81, cite par Monseur et Lafontaine, 2005) définit le *Rho* comme « *an indicator of variance between clusters relative to the variance within clusters. In education, clusters are usually regions or schools or classes. The most frequent use of rho is for the variance between-schools and within-schools or between-classes and within-classes. The extreme values of rho are (...) 0.0 and 1.0. If one is examining the scores on a reading test, for example, a rho of 0.30 means that 30 percent of all variation is between schools and 70 percent between students within schools. When the rho is high, the difference between schools is high and, in this case, the level of a student's reading score will very much depend on which school he or she attends.* » (p. 2).

classes dans les écoles), ou bien, si l'on place seulement le coefficient du niveau spécifique dont on cherche à estimer leur apport à la variance totale explicable (p. 32).

Le coefficient Rho peut donc être assumé comme suit :

$$\rho(Y_{ijs}, Y_{i'js}) = \left[\frac{(\sigma^2_{vs} + \sigma^2_{\mu j})}{(\sigma^2_{vs} + \sigma^2_{\mu j} + \sigma^2_e)} \right]$$

où :

ρ correspond au coefficient Rho à estimer (corrélation intra-groupe(s))

Y_{ijs} correspond à l'*outcome* de l'élève i placé dans la classe J de l'école S ;

$Y_{i'js}$ correspond à l'*outcome* de l'élève i' placé dans la même classe J et la même école S que l'élève i ;

σ^2_{vs} et $\sigma^2_{\mu j}$ et σ^2_e correspondent respectivement aux variances d'erreur associées au niveau école, classe et élève. La somme de ces trois termes correspond à la variance totale.

Bien évidemment, la variance entre classes (au sein des écoles ou bien en l'absence de ce niveau d'analyse) et la variance entre écoles peuvent être résumées par un seul terme (en y incluant les classes). Nous retiendrons qu'il faut être prudent à l'heure d'interpréter ces coefficients sans oublier l'emboîtement effectivement existant dans la réalité.⁸⁹ Enfin, notons que Snijders et Bosker (1999, cités par Monseur et Lafontaine, 2005) indiquent que « *this parameter (NB: le Rho) is called a correlation coefficient, because it is equal to the correlation between values of two randomly drawn micro-units in the same, randomly drawn, macro-unit* » (p. 2).

1.4 Les choix de nos analyses

Une fois présentées de manière générale les justifications de l'usage prioritaire des MMN pour nos analyses, nous devons expliciter certains choix qui tiennent directement à la mise en opération de cette option. Notons d'abord que la littérature recommande de prévenir du danger qui représentent les énormes possibilités de complexité offertes par les modèles multiniveaux (cf. Bressoux, 2008 ; Bickel, 2007). En guise de règle pratique, « *ne pas multiplier les effets aléatoires, même s'ils sont statistiquement significatifs* » (p. 305 ; Bressoux, 2008). A cet égard,

⁸⁹ Cela revient à dire, par exemple, que si l'on ne prend que les écoles comme deuxième niveau, l'interprétation du Rho sera forcément celle de la corrélation (i.e. la vraisemblance probable entre résultats) entre deux élèves appartenant à la même école mais raisonnant comme si toutes les classes au sein de chaque école n'avaient pas d'effet différenciateur. Consignons toutefois que Hox (2002 ; p. 32) indique qu'actuellement la littérature accepte deux utilisations, l'une portant sur la variance des classes et l'autre des écoles (en y incluant celle des classes). Ceci d'après les buts de l'analyse : soit, départager la variance par niveaux ; soit, établir une mesure de la corrélation attendue entre deux élèves aléatoires mais placés dans la même situation (i.e. la même classe dans la même école).

Hox (2002) conseille de considérer « *higher-order interactions only if there is a strong theoretical justification for their importance, or if an unusually large variance component for a regression slope calls for explanation* » (p. 31). Par ailleurs, Hox (2000) prévient aussi sur l'utilisation de trois ou plus de niveaux d'analyse : « *the resulting models are not only difficult to follow from a conceptual point of view, they may also be difficult to estimate in practice* » (p. 30). Un des problèmes soulignés à cet égard est la réduction des unités à des niveaux les plus élevés. Pourtant, Hox (2002) ajoute que « *intuitively, three-level structures such as pupils in classes in schools (...) appear to be both conceptually and empirically manageable* » (p. 30), ce qui est aussi conseillé par Monseur et Lafontaine (2005). Nous pensons qu'il est cohérent d'estimer des modèles à trois niveaux, compte tenu qu'ils sont plus près de la réalité et de ce fait plus heuristiques.

Toujours est-il que, parmi l'ensemble des possibilités d'analyse offertes par les MMN, la plupart restera, majoritairement hors de notre travail ici. Nous nous attacherons donc principalement à estimer les différences entre les constantes ou scores moyens attachés à chaque classe et à chaque école (i.e., pris comme des effets aléatoires). Puis, nous allons nous concentrer davantage à mieux cerner les effets fixes communs à tous les élèves du pays, au delà de l'école et de la classe qu'ils fréquentaient en 2004. Concernant les choix plus spécifiques et générales des modèles, et sachant que nous avons plus de 3000 établissements avec plus de 6000 classes concentrant plus de 200000 élèves, nous allons utiliser de préférence l'approche d'estimation FML ou Maximum de Vraisemblance complet (Hox, 2002 ; Bressoux, 2008). Ceci va nous permettre de considérer, en général, la statistique de la déviance comme un paramètre de comparaison de la qualité d'ajustement des modèles.⁹⁰ L'autre critère que nous allons prendre est la statistique BIC (*Schwarz's Bayesian Information Criterion*) qui offre la possibilité de comparer les modèles non-emboîtés, autrement dit, ne se contenant pas les uns dans les autres (Hox, 2002 ; p. 45-46).

La littérature conseille le centrage des variables sur leur grande moyenne. Ceci devient une procédure presque incontournable lorsque on va tester des effets d'interaction entre variables de niveau distinct (Hox, 2002 ; Bickel, 2007 ; Bressoux, 2008). De fait, seulement quand on a centré toutes les variables, même celles catégorielles, on peut raisonner véritablement en termes d'individu moyen dans des situations moyennes, car chaque variable catégorielle est rapportée à sa proportion dans la population. Toutefois, nous allons garder pour l'instant les variables dans son état original, à l'exception de celle dépendante et des mesures déjà presque standardisées à la base (e.g. les scores factorielles). Ceci par des raisons de lisibilité, de

⁹⁰ Cette statistique est définie comme $-2 \log V$ (i.e. moins deux fois le logarithme naturel de l'estimation de vraisemblance) ou fonction *likelihood* de convergence (Hox, 2002 ; p. 43). Ceci sachant que la différence entre les statistiques des deux modèles emboîtés suit une distribution χ^2 avec comme degrés de liberté la différence entre le nombre de variables incluses dans chacun (Nakache et Confais, 2003).

comparabilité avec d'autres procédures (cf. régression par quantiles) et parce que nous sommes, surtout, à un état d'usage des MMN en termes de gain de précision, de consistance et de contrôle.⁹¹ En outre, nous allons nous servir des estimations de variance expliquée estimée à partir des coefficients aléatoires par niveau (Hox, 2002 ; p. 64), prenant comme base d'appui les modèles dit vides (Bressoux, 2008 ; p. 291). Ceci sachant que cette estimation devient davantage compliquée lorsque l'on admet, soit des pentes aléatoires, soit des structures de covariance entre celles-ci et les constantes par niveau, ou plus précisément, quand on fait intervenir des effets d'interactions entre niveaux (Bressoux, 2008 ; Hox, 2002). Ce choix momentané va nous permettre de démêler en termes généraux l'importance explicative de certaines variables et de mieux comprendre ce qui se passe à chaque niveau de l'école au fur et à mesure que l'on intègre de variables explicatives. Nous prendrons donc volontiers la notion de « variance modélisée » plutôt que celle de « variance expliquée » *stricto sensu* (Bressoux, 2008), tout en gardant en vue d'autres indicateurs sur le sujet tels que les R² des modèles MCO ou les r et R² entre scores prédits et originaux (Singer et Willet, 2003, cités par Bressoux, 2008).

Globalement, les équations reflétant les modèles MN que nous allons mettre en place sur nos variables continues (cf. tableaux n°11 à n°18, Tome I), peuvent s'exprimer, dans leur forme de base, comme suit :

a) Estimations au Niveau 1 (élève) :

$$Y_{ijs} = \alpha_{0js} + \beta_{100}X + \beta_{200}DimsSUBJ + \beta_{300}Pratiques + \beta_{400}\widehat{CPairs}_{-i} + \varepsilon_{ijs}$$

Où,

Y_{ijs} correspond au score de l'élève i placé dans la classe j de l'école s ;

α_{0js} est le score moyen constant correspondant comme base à tout élève dans la classe j de l'école s ;

$\beta_{100}, \beta_{200}, \beta_{300}$ et β_{400} sont des vecteurs de coefficients exprimant, respectivement, la relation entre les scores et les caractéristiques individuelles socio-scolaires de l'élève (X), les dimensions subjectives qu'il exprime vis-à-vis de sa scolarité ($DimsSUBJ$) et les interactions et pratiques d'étude qui nous intéressent. Ces termes sont considérés comme fixes dans cette équation (i.e. ils ne constituent pas des pentes proposées comme aléatoires) ;

⁹¹ Certes, comme le note Bressoux (2008), « (L)e coefficient de régression standardisé permet (...) de savoir quelles sont les variables qui ont l'impact le plus fort » (p. 135), mais ici ils nous intéressent surtout d'avoir une idée des variations selon les échelles originales et selon si l'on présente ou pas une caractéristique donnée. D'ailleurs on peut toujours revenir à la standardisation des coefficients *a posteriori*. Ceci à l'aide des erreurs standards de ceux-ci et des écarts-type de variables. Comme l'indique Hox (2002) et d'autres cela change les estimations sur la constante.

β_{400} correspond à un vecteur des impacts moyens estimés des caractéristiques du public de la classe entourant à l'élève i (i.e. moyenne et écart-type d'années de scolarité des mères, proportions en bas bien-être et en haute confiance). Rappelons que ces termes devraient, en théorie, avoir été placés dans le « niveau classe » si l'on n'avait pas fait la correction commentée plus loin, visant à exclure l'élève lui-même à l'heure de les estimer ;

ε_{ijs} correspond à un terme d'erreur attaché à l'élève i et supposée aléatoire et constant entre classes et écoles ;

b) Estimations de Niveau 2 (classe) :

$$\alpha_{0js} = \alpha_{00s} + \gamma_{010}CE + \gamma_{020}CrE + \gamma_{030}PrE + \gamma_{040}TC + u_{0js}$$

Où,

α_{0js} correspond au score moyen des classes dans l'école s ;

$\gamma_{010}, \gamma_{020}, \gamma_{030}$ et γ_{040} sont des vecteurs des coefficients exprimant, respectivement, la relation entre les scores moyens des classes par école (ceux qui à leur tour servent de constante pour tout élève dans sa respective classe j) et les caractéristiques des classes, ici divisées en : caractéristiques de l'enseignant (CE), croyances de l'enseignant (CrE), pratiques de l'enseignant (PrE), et enfin, la taille de la classe (TC).⁹² Ces termes sont ici notés comme fixes (i.e. ils ne constituent pas des pentes pouvant varier significativement d'une école à l'autre) ;

u_{0js} correspond à une variation aléatoire ($\sim N(0, \sigma_u^2)$) attachée à la classe j dans l'école s et supposé constante entre écoles ;

c) Estimations de Niveau 3 (établissement) :

$$\alpha_{00s} = \alpha_{000} + \delta_{001}RE + \delta_{002}TE + \omega_{00s}$$

Où,

α_{00s} correspond à le score moyen de l'école s ;

α_{000} correspond à la constante commune à toutes les écoles analysées ;

δ_{001} et δ_{002} sont des vecteurs des coefficients exprimant, respectivement, la relation entre les scores moyens par école et les variables de niveau établissement reliées à leurs ressources (RE), puis à leur catégorie (TE) ;

⁹² En effet, sur des variables de ce type on voit les limites théoriques et pratiques du fait de penser l'individu séparé des valeurs moyennes qu'il contribue à engendrer. Ici, enlever un élève à chaque groupe classe ne change absolument rien à la globalité et à la distribution de cette variable. Bien entendu cela n'est pas le cas d'autres variables telles que « la moyenne de ». Toujours est-il que la taille des classes reste comme une variable de niveau deux.

ω_{00s} correspond à une variation aléatoire ($\sim N(0, \sigma_\omega^2)$) attachée à l'école s et qui n'est pas capturée par les variables ici considérées ;

Notons que si nous avons accepté que certaines pentes de niveau 1 (élève) ou 2 (classe) puissent varier entre les unités de niveau supérieur (classes ou écoles), on devrait estimer deux équations pour chaque niveau supérieur, l'une pour les constantes (ce qu'on a montré ici) et l'autre sur les pentes choisies. Par exemple, si l'on pose que les pratiques et les interactions déclarées par les élèves peuvent avoir un impact différent selon la classe d'appartenance, nous devrions noter ce terme comme suit :

β_{3j0} **Pratiques** (au lieu de β_{300} *Pratiques*) ; puis, l'expliquer par les variables de deuxième niveau :

$$\beta_{3j0} = \gamma_{300} + \gamma_{310}CE + \gamma_{320}CrE + \gamma_{330}PrE + u_{3js}$$

Où,

γ_{300} correspondrait à un vecteur contenant les effets moyens sur les scores des élèves (i.e. communs à toutes les écoles) de la part des pratiques considérées ;

$\gamma_{310}, \gamma_{320}, \gamma_{330}$ seraient les vecteurs des coefficients exprimant la détermination de ces impacts de par les caractéristiques des classes déjà notées ;

u_{3js} correspondrait à la variation spécifique de ces coefficients attachés à la classe

Pour leur part, les coefficients aléatoires seront supposés, pour la plupart, non corrélés entre eux, ce qui dans le cas des pentes pensées comme aléatoires, deviendra à choisir une structure contrainte d'estimer à zéro les possibles corrélations entre paramètres aléatoires, du moins entre pentes.⁹³ Enfin, consignons que toutes nos estimations sur des MMN seront faites à l'aide la Proc MIXED de SAS® (V.9.1) en prenant comme référence l'article de Singer (1998)⁹⁴ et l'ouvrage de Bressoux (2008). Toutefois, nous allons nous servir aussi de la technique d'estimation par les MCO avec des écarts-types robustes (sur STATA® V.8), notamment par souci de comparaison entre estimations issues de méthodes différentes.

⁹³ En termes concrets, cela revient à choisir l'option « Type=UN(1) » sous SAS®, ce qui signifie que le logiciel n'estime que la variation entre constantes et entre pentes. En même temps les variations conjointes testées le sont de manière, *a priori*, non-structurée. Autrement dit, leur structure est issue des propres données. Ce choix sera utilisé ici dans les cas où l'on admettra des pentes éventuellement aléatoires, compte tenue la haute consommation de temps qui nécessitent des structures plus complexes et qui n'apportent pas assez à nos objectifs.

⁹⁴ Singer, J.; 1998 ; *Using Proc Mixed to fit multilevel models, hierarchical models and individual growth models* ; Journal of educational and behavioural statistics, Vol.24, n°4 ; pp. 323-355

1.5 Modèles Catégoriaux

Pour les modèles catégoriaux nous allons utiliser la Procédure LOGISTIC et la Procédure GLIMMIX (SAS®, V.9.1) selon les recommandations et références du SAS Institute (cf. Schabenberger, 2005).⁹⁵ Ces modèles seront aussi estimés à trois niveaux,⁹⁶ étant équivalents en fait à des régressions logistiques mononiveaux mais avec des constantes acceptées comme aléatoires (Schabenberger, 2005 ; Bressoux, 2008).⁹⁷ Sur les modèles communiqués, par pratique et discipline, nous allons indiquer une série de statistiques d'ajustement. Celles-ci seront divisées en deux groupes selon qu'elles sont issues des modèles multiniveaux ou, de la référence mononiveau respective. Cette démarche devrait permettre d'élargir le jugement qu'on peut en tirer de chacun des modèles. Ainsi par exemple, dans le cas des MMN logistiques nous ferons davantage attention à l'indicateur de dispersion fournie par le ratio entre le Khi carré Généralisé et les degrés de liberté. Dans ce cas, une valeur proche de 1 est signe d'un bon ajustement du modèle aux données. Au contraire, des écarts importants à la valeur 1 peuvent être qualifiés de « *overdispersion* » (>1) ou de « *underdispersion* » (<1) (Schabenberger, 2005 ; Hox, 2002). En revanche, notons que les statistiques des modèles logistiques mononiveaux, pris comme références informatives, s'avèrent plus nombreuses. Dans ce cadre, nous ferons attention au R² de Nagelkerke (SAS Institute, 2005) comme une mesure approximative de la variation prise en compte par le modèle. De même, les variations de la statistique du logarithme de vraisemblance (-2 Log L) seront utiles pour qualifier l'apport de l'inclusion des variables aux modèles. Pour sa part, la statistique « *c* » nous informera sur la pertinence d'usage d'un modèle visant à prédire le comportement étudié. Enfin, la probabilité et la taille du test Khi2 de Hosmer et Lemeshow nous informeront sur la capacité des modèles à reproduire la structure originale des données. Sachant que la statistique Khi2 est sensible aux variations de taille nous regarderons surtout le fait de ne pas rencontrer des valeurs trop élevées (e.g. Khi2>1000) comme indicateur d'un ajustement acceptable. Par ailleurs, nous avons accordé du temps à la comparaison entre coefficients des effets fixes issus des modèles logistiques mono et multiniveaux (i.e. à constantes aléatoires), trouvant que la vraisemblance entre les deux est assez forte. Ces différences, en plus d'être réduites à un nombre restreint de variables pour lesquelles celles-ci s'avéraient significatives, portent exclusivement sur des

⁹⁵Schabenberger, O.; 2005 ; Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models; Paper 196-30 ;SAS Institute Inc., Cary, NC ; 20 pp.

⁹⁶ Soulignons que par des raisons de simplicité dans la mise en opération et dans l'interprétation, et eu égard aussi de la très faible variation des modalités à expliquer au niveau de la classe, nous aurions préféré n'estimer ces modèles qu'à deux niveaux. Pourtant, dans ces cas plusieurs modèles ne convergeaient pas. Donc, nous avons été obligés de prendre l'option des modèles à trois niveaux pour garantir leur convergence respective. Quoi qu'il en soit, pour toutes les estimations à variable dépendante dichotomique, nous prendrons beaucoup d'appui dans les régressions logistiques et leurs indicateurs comme support aux analyses faites sur des modèles mixtes.

⁹⁷ Notons que la Proc GLIMMIX de SAS® utilise le *Predictive Quasi-Likelihood* (PQL) comme méthode d'estimation des paramètres. Pour plus de discussions sur cette méthode et des alternatives, voir Bressoux (2008, pp. 408-409).

facteurs de niveau classe ou école. De ce fait, et compte tenu ce que nous avons déjà indiqué, nous ne communiquerons que les résultats des modèles issus de la Proc GLIMMIX de SAS® (cf. tableaux n°23 à n°30, Tome I).

Finalement et pour plus de précision, rappelons que tout coefficient logit issu de ce type de modélisation des phénomènes dichotomiques équivaut à :

$$\text{Log} \left[\frac{P_{ijs}}{1 - P_{ijs}} \right] = \beta X_{ijs}$$

où le vecteur des coefficients β , associé aux variables contenues dans le vecteur X_{ijs} représentant les caractéristiques de l'élève i placé dans la classe j de l'école s , correspond au logarithme naturel du ratio entre P_{ijs} et son contraire. Ceci sachant que P_{ijs} représente la probabilité que l'élève i placé dans la classe j de l'école s déclare le comportement étudié. Cela permet de calculer aussi les probabilités marginales, ou variations dans la variable dépendante, représentées par les coefficients β estimés par les modèles.⁹⁸

1.6 D'autres aspects à considérer

1.6.1 Les phénomènes de colinéarité

Le degré de colinéarité entre certaines variables au sein d'un modèle de régression peut être pensé comme une indication du degré d'information dupliquée entre mesures. Il ne s'agit pas d'un phénomène regrettable en soi, car ce n'est pas un problème de nature, mais plutôt de quantité (Gujarati, 2004). Ceci compte tenu que nous savons que les variables sont, la plupart du temps, reliées entre elles et que de ce fait, elles véhiculent des informations semblables. L'important est donc de qualifier ce phénomène et de l'avoir sous surveillance pour s'assurer que chaque variable apporte à l'analyse quelque chose en plus.

⁹⁸ Notons que dans le modèle multiniveau à constantes aléatoires, on postule que la variation (l'erreur) dans les estimations des *logits* associés à la probabilité d'occurrence d'un événement est liée au niveau supérieur d'emboîtement. On assume l'existence d'une variation entre ces unités supérieures composée autant d'une partie constante ou commune à toutes celles-ci, que d'une autre partie spécifique à chacune. Dans notre cas, au niveau classe par exemple, cela prend la forme suivante : $\text{Log}(P_{ijs}/(1 - P_{ijs})) = \beta_{0js} + \beta_1 X_{ijs}$ ou $\beta_{0js} = \gamma_{00s} + u_{0j0}$. Remarquons qu'ici la variance par niveau ne suit pas la même logique ni la même interprétation que dans les modèles multiniveaux à variable dépendante continue (Hox, 2002). Il s'agit surtout de contrôler la dispersion des résidus due à l'existence d'emboîtement. En outre, les estimations de la valeur « -2 Res Log Pseudo-Likelihood » (sur SAS®) ne jouissent pas des mêmes propriétés comparatives que la statistique de déviance chez les modèles MN sur des variables continues (Hox, 2002 ; Bressoux, 2008).

Il existe en général trois méthodes pour mesurer la colinéarité. D'abord, le regard des erreurs standard des coefficients estimés (Falissard, 1998) où l'on doit faire attention à sa taille car s'ils sont très grands (coefficients de variation ≥ 100 par ex.), il est possible qu'il y ait de la colinéarité entre variables au sein du modèle. A cette estimation plutôt intuitive s'ajoutent deux indicateurs : le VIF (*Variance Inflation Factor*) et l'Indice d'État (*condition index*). Le premier correspond à une mesure attachée à chaque variable allant entre 1 et ∞ . Si la valeur de cet indice est égale à 1 nous sommes en absence de colinéarité et l'on peut dire que cette variable apporte une information absolument unique ou singulière. En général, on considère qu'il montre une sévère colinéarité s'il est supérieur à 10 (p. 452 ; Panik, 2009).⁹⁹Toutefois, Bickel indique la valeur 4 comme « *usual cutoff* » pour l'indice VIF (p. 78 ; 2007). Pour sa part, Bressoux (2008) note que beaucoup d'auteurs admettent la valeur 5 comme seuil de colinéarité acceptable pour une variable (p. 143), mais que d'autres recommandent la valeur 10 lorsque nous sommes en présence de plusieurs régresseurs (p. 144). Enfin, le deuxième indicateur est préféré par certains auteurs qui critiquent les inconsistances du premier (cf. Gujarati, 2004). Panik (2009) note que des valeurs inférieures à 25 restent faibles. Gujarati signale qu'entre 30 et 100 nous avons de la colinéarité au sein du modèle mais qu'au delà de 100, elle est sévère. De ce fait, des variables deviennent colinéaires si elles représentent autour de 50% pour un Indice d'état plutôt élevé.¹⁰⁰ En termes concrets, nous allons regarder, pour chaque modèle construit, les indices disponibles sur la colinéarité de manière à savoir dans quelles marges on est. Ceci à l'aide d'un modèle OLS égal à nos estimations multiniveau. Toujours est-il que nous avons utilisé ces indices de manière exploratoire, par exemple, pour décider de la non-inclusion, simultanée, de variables telles que la scolarité de la mère et celle du père.

1.6.2 Le calcul des valeurs correspondant au groupe

L'exclusion de l'individu lui-même dans le calcul des valeurs moyennes de son groupe de pairs est appliqué très souvent en économie de l'éducation (cf. Hanushek et al., 2003 ; Hoxby, 2000 ; Hoxby et Weingarth, 2006). D'un point de vue analytique, on doit considérer les problèmes associés à l'endogénéité pour essayer de ne pas mettre, des deux côtés de l'équation, les mêmes variables et valeurs. Du moins, diminuer au maximum ces situations, de même qu'essayer d'isoler au maximum l'individu du groupe de référence (ici, la classe). Ainsi,

⁹⁹ Panik, M.; 2009 ; *Regressions Modeling. Methods, Theory and Computation with SAS* ; CRC Press, NY ; 814 pp.

¹⁰⁰ En effet, cet indice provient d'une analyse en composantes principales qui extrait les valeurs propres de chaque variable dans l'ensemble. Ainsi, « *For each variable (...) the proportion of the variance of the estimate accounted for by each principal component. A collinearity problem occurs when a component associated with a high condition index contributes strongly (variance proportion greater than about 0.5) to the variance of two or more variables.* » (SAS Institute, User's Guide, 2005).

le plus souvent l'estimation des effets de pairs dans un contexte de groupe classe prendrait la forme suivante :

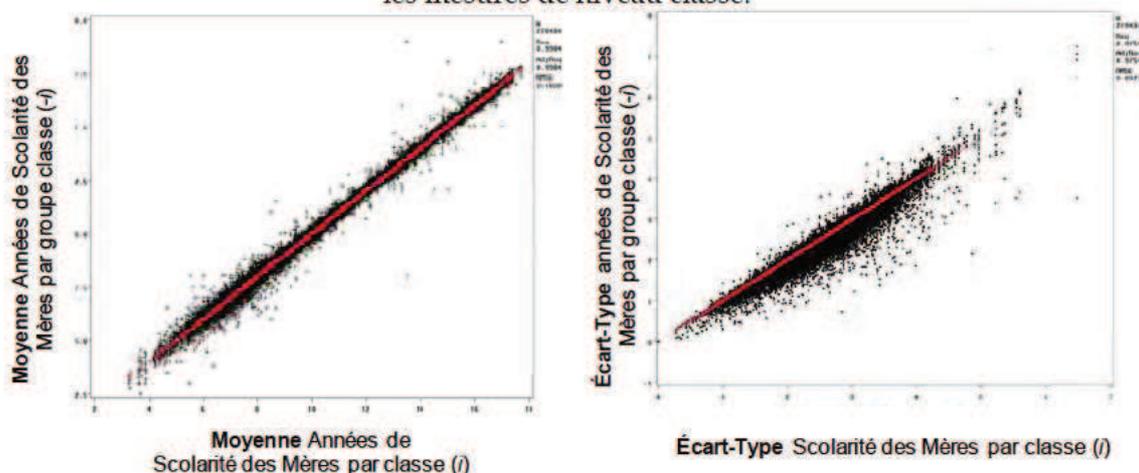
$$Y_{ijs(t1)} = \beta_1 X_i + \beta_2 C_j + \beta_3 C_{js(t1)} + \beta_3 \tilde{P}_{-ijs} + \dots + \varepsilon_i$$

Où $Y_{ijs(t1)}$ correspond aux résultats à un test donné pour l'élève i scolarisé dans la classe j (et l'école s) pendant la période $t1$. Suivant l'équation générale déjà décrite plus haut, X_i correspond à un vecteur des caractéristiques attachées à l'individu i et $C_{js(t1)}$ est un vecteur contenant les caractéristiques de la classe j de l'école s pendant l'année $t1$. Enfin, les effets de pairs sont censés être capturés par le terme noté comme $\tilde{P}_{(-ijs)}$ et qui correspond à un vecteur de caractéristiques moyennes des élèves dans la classe j de l'école s , mais sans y compter l'individu i lui-même (i.e. $-i$).

De cette manière on obtient une valeur différente pour chaque individu/élève (i.e. une moyenne du groupe classe $(-i)$), ce qui est sans doute en phase avec l'utilisation courante en économie des modèles mono-niveau. Réitérons qu'en général la démarche est d'autant plus adéquate que l'on dispose d'une mesure de $\tilde{P}_{(-ijs)}$ au temps $t0$ au lieu du $t1$. Le choix que nous voudrions illustrer va toutefois au delà du phénomène temporel. Il s'agit de l'inclusion, ou de l'exclusion, de l'individu dans les calculs des valeurs moyennes de son groupe d'appartenance. Notons pourtant que la valeur des moyennes $(-i)$ sont très proches, en termes de distribution, des moyennes incluant l'individu $(+i)$. Cela est illustré par le graphique C44 suivant.

Graphique C44

Illustration de rapports entre deux manières d'objectiver les mesures de niveau classe.

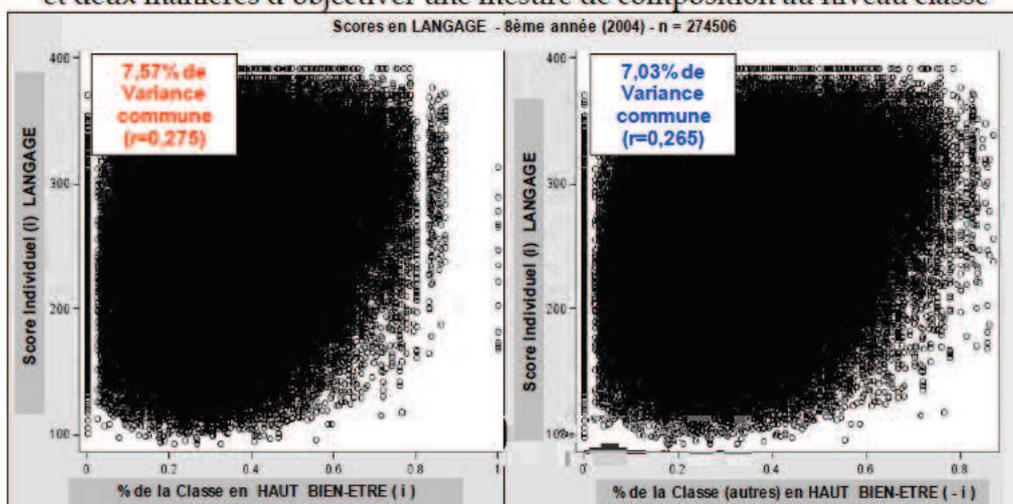


Source : élaboration de l'auteur. Lecture : variation conjointe de deux types de mesures par classe (avec ou sans l'individu). A gauche, la scolarité moyenne des mères et à droite, l'écart-type de cette même variable.

En outre disons que, en plus de sa forte proximité à la moyenne du groupe en entier (e.g. le $R^2=0,998$ entre les deux moyennes de la scolarité des mères sur nos données), il faut noter que théoriquement cette correction reste discutable car elle n'enlève rien à la valeur concrète de ce qui signifie une proportion donnée de telle ou telle caractéristique, par exemple, en termes de climat dans la classe. Mais admettons que ce débat pourrait aller dans les deux sens, autant du côté de ceux qui voient les moyennes des caractéristiques par classe comme des valeurs totalement « autres », autant du côté de ceux qui prônent cette correction. Notons que ceci implique des changements dans la formulation des modèles. A notre avis et malgré sa légitimité, cette correction est reliée au manque de distinction précise, notamment chez les économistes, de ce qui est effet de composition et de ce qui constitue des vrais effets de pairs. Pourtant, comme nous l'illustrons sur le graphique suivant, le fait d'inclure ou pas l'individu dans le calcul des moyennes par classe ou école peut avoir des conséquences, quoi que très légères, sur les estimations. En termes concrets, pour les MMN cela se traduit par un calcul avec des degrés de liberté différents (i.e. attachés à l'individu au lieu de le faire à des niveaux supérieurs), mais tout en obtenant des coefficients fixes et aléatoires presque identiques (à quelques décimales près). Pourtant, les mesures de déviance ont tendance à être plus réduites lorsqu'on considère cette correction.

Graphique C45

Rapports entre les scores des élèves en Langage et deux manières d'objectiver une mesure de composition au niveau classe



Source : élaboration de l'auteur. Lecture : corrélations de Pearson illustrées entre les scores individuels en langage et la proportion par classe en condition de Haut Bien-être à l'école, selon qu'elle a été estimée en considérant l'élève lui-même (i) ou pas (-i). Les scores sont dans leur échelle originale ($\mu=250$ et $\sigma=50$).

Compte-tenu des possibilités de comparaison et de ce qu'on a vu d'un point de vue empirique nous allons nous servir de cette mesure plutôt que des moyennes ou de proportions sur la classe et/ou l'école en entier. Ceci sachant que l'on peut toujours revenir en arrière sur ce sujet au cas où cela serait nécessaire.

1.7 La régression par Quantiles

La méthode de régression par quantiles (*quantile regression*) a été introduite par Koenker et Bassett (1978, cités par Koenker et Hallok, 2001).¹⁰¹ Celle-ci peut être vue comme une « *natural extension of classical least squares estimation of conditional means models to the estimation of an ensemble of models for conditional quantile functions* » (Koenker et Hallok, 2001 ; p. 143). Autrement dit, la régression par quantiles permet d'estimer l'impact des co-variables sur la distribution d'une variable dépendante en tenant compte des différents moments qui la constituent et de la forme de celle-ci, tandis que la régression classique ne s'occupe que d'un seul moment de la distribution : la moyenne. Comme le note Buchinsky (1998)¹⁰² « (...) *potentially different solutions at distinct quantiles may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable* » (p. 89).

Cette méthode présente donc un précieux avantage pour la recherche en sciences sociales. En effet, théoriquement, mais aussi empiriquement, les individus placés en bas ou en haut de n'importe quelle échelle (salaires, scores aux tests, progressions académiques, poids, taille, etc.) ne sont pas influencés de la même manière par certains facteurs que les individus placés au-dessus/au-dessous des moyennes de ces mêmes échelles. De plus, une telle approche devient fortement indispensable quand les niveaux extrêmes de la variable à analyser ou ceux supérieurs à un certain seuil, restent la cible des recherches (Chen, 2005).¹⁰³ C'est notamment le cas des études dans le domaine de la santé et/ou l'environnement (Chen, 2005). Répondre à ces problématiques nécessite donc des outils sophistiqués comme c'est le cas de cette variante de la régression linéaire traditionnelle, dont elle devient tout à fait complémentaire (Koenker, 2005).¹⁰⁴

¹⁰¹ Koenker, R. ; Hallock K. ; 2001 ; *Quantile Regression: An Introduction* ; Journal of Economic Perspectives, Vol.15, n°4 ; pp. 143–156.

¹⁰² Buchinsky, M. ; 1998 ; *Recent Advances in Quantile Regression Models: A practical guide for empirical research* ; Journal of Human Resources, Vol.33, n°1 ; pp. 88-126

¹⁰³ Chen, C. ; 2005 ; *An Introduction to Quantile Regression and the QUANTREG Procedure* ; Paper n°213-30 ; SAS Institute Inc., Cary, NC ; 24 pp.

¹⁰⁴ Koenker, R. ; 2005 ; *Quantile Regression* ; Cambridge University Press ; 349 pp.

Mais, en plus de mieux saisir les changements dans les paramètres estimés par régression tout au long de la distribution analysée, la régression par quantiles permet de faire face à des phénomènes comme l'hétéroscédasticité (i.e. la non normalité des résidus) présents dans des nombreux types de distributions (Buchinsky, 1998), mais aussi face aux valeurs atypiques (Koenker, 2005). Pour ce faire, ce type de régression s'est basé, à l'origine, sur la méthode de la moindre déviation absolue (MAD ou *minimum absolute deviation*) afin de minimiser, à partir d'un vecteur de coefficients, la somme des carrés des erreurs¹⁰⁵. Pourtant, il faut se garder de penser que la régression par quantiles consiste à faire une sous-division de la variable dépendante (Schlinder, 2003) et procéder ainsi à estimer des coefficients pour chaque sous-échantillon. En effet, Koenker et Hallok (2001) signalent que « *all the sample observations are actively in play in the process of quantile regression* » (p. 145). Elle nécessite donc une simulation spécifique pour y parvenir.

En effet, pour estimer l'impact conditionnel des variables indépendantes pour chaque quantile, il est nécessaire de faire des ré-échantillonnages à partir des données existantes. Pour chacun d'entre eux, une estimation des paramètres est faite (Koenker et Hallok, 2001). Ces ré-échantillonnages avec remise (i.e. pour revenir toujours à la taille originale de l'échantillon mère moins une observation) peuvent être faits à l'aide d'une procédure connue comme *bootstrap* qui crée des nouveaux échantillons en les simulant aléatoirement à partir de l'échantillon original, puis, en répétant l'expérience K fois¹⁰⁶. Schlinder (2003) indique que « *Different quantiles are estimated by weighting the residuals differently. For the median regression, all residuals receive equal weight. However, when estimating the 75th percentile, negative residuals are weighted by 0.25 and positive residuals by 0.75. The criterion is minimized, when 75 percent of the residuals are negative. This is set up as a linear programming problem and solved* » (p. 13). Enfin, un autre point important, c'est que la méthode de régression par quantiles offre des intervalles de limites de confiance pour chaque paramètre estimé à chaque niveau de la distribution de la variable dépendante.

¹⁰⁵ De manière semblable aux modèles MCO qui cherchent le vecteur des coefficients que minimise la somme des carrés des erreurs ou distances à la moyenne, la méthode MAD cherche aussi, pour chaque quantile q et eu égard l'ensemble de variables indépendantes, le vecteur de coefficients qui minimise la somme des carrés des valeurs absolues des erreurs. Ainsi, la régression par quantiles peut être formulée comme la solution pour le q th (*conditional quantile function*) correspondant à $Q_y(q|X) = \beta_0(q) + \beta_1(q)X$ où l'objectif est de trouver les paramètres $\beta_0(q)$ et $\beta_1(q)$ qui résoudront le problème d'optimisation suivant (Panik, 2009 ; pp. 385-397):

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n \rho_q(Y_i - \beta_0 - \beta_1 X_i).$$

Où, pour le quantile(q), β_0 et β_1 son des vecteurs des coefficients, X_i constitue l'ensemble des variables explicatives associées à l'individu i , Y_i correspond aux valeurs de la variable dépendante pour ce même individu.

¹⁰⁶La méthode du *bootstrap* a été proposée par Efron (1979, 1982 cité par Buchinsky, 1998).

Du côté de l'apport de cette technique pour les études reliées aux effets de pairs, Ding et Lehrer (2006) notent que la régression par quantile « (...) provides a more flexible approach to characterizing the effects of observed covariates such as peer groups on different percentiles of the conditional achievement distribution providing a richer pattern on how peer groups affect student achievement. (...) If ability and peer groups are substitutes we would expect the marginal returns on peers to decrease with ability. If ability and peer groups are complements then marginal returns to peers would be higher for the more able.» (p. 26-27). De manière coïncidente, Schindler (2003) indique que « estimating traditional "average effect" models do not give the full picture of the peer group effect, because they estimate the effect only for the average student. This may produce misleading results if peer group effects differ for high and low ability students.» (p. 6). « With quantile regression, we are not only able to address the question whether peer groups matter, but also for whom does the peer group matter? » (p. 11). Schneeweis et Winter-Ebmer (2007) rejoignent ces mêmes arguments.

Enfin, une explication fournie par Schlinder (2003) nous semble assez bien adapté pour mieux comprendre l'ensemble de la démarche : « To recall, quantile regression techniques estimate peer effects at different values of the conditional test score distribution. This means that when we talk about "weak readers", we do not necessarily refer to students with low (absolute) test scores, but to students reading less well than we would expect them to according to their background characteristics. Thus, "weak readers" are, e.g., students from well-educated homes, who read less well than the average student with well-educated parents.» (p. 32).¹⁰⁷

1.7.1 Nos choix d'analyse sur cette méthode

En toute logique, notre intérêt est davantage de nous servir des méthodes disponibles plutôt que de rentrer dans un débat détaillé de leurs caractéristiques, d'autant plus que ces méthodes ont été acceptées par la communauté scientifique et incorporées dans les packages statistiques les plus courants. Ainsi, pour ce travail, nous avons décidé de prendre comme référence l'article de Chen (2005) et les indications fournies par le SAS Institute (2005) vis-à-vis de la méthode de régression par quantiles.¹⁰⁸ Toutefois, nous avons eu en vue des estimations sur nos données faites sur le logiciel STATA®. Parmi les principaux choix opérés en matière

¹⁰⁷ Puis, Schlinder continue sur les répercussions politiques d'une telle démarche en indiquant que « (...) for policy purposes, we are usually interested in students being poor readers unconditional on their background. It is thus useful to examine, how unconditionally poor readers are related to conditionally poor readers. If we, e.g., can show that the conditionally poor readers are typically identical with the unconditionally poor readers, then, our quantile regression results can be transferred directly to the unconditional distribution of test scores. (...). Unconditional and conditional test scores are closely related, which illustrates the tendency of the conditionally poor readers to be identical with those who get low raw test scores. This means that all results extracted above for the conditional distributions are generally valid for the unconditional distribution, too. This relationship makes our results much more easy to interpret for policy advice.» (p. 32-33).

¹⁰⁸ SAS Institute ; 2005 ; *The QUANTREG Procedure (Experimental)* ; SAS Institute Inc., Cary, NC ; 70 pp.

d'analyse, nous avons pris comme méthode d'estimation des erreurs standards des coefficients (et donc, de leurs limites de confiance) l'option *resampling* (Chen, 2005 ; p. 22) qui exécute par défaut une amélioration de l'algorithme *bootstrap* nommée MCBM (*Markov Chain Marginal Bootstrap*, p. 22-23) et qui arbitre entre la dépense du temps et de ressources informatiques et une bonne adéquation pour les grandes bases de données, le tout avec contrôle de l'hétéroscédasticité et des données aberrantes. Le nombre d'itérations utilisées pour calculer les intervalles de confiance des coefficients a été fixé à 200.

Concernant le nombre de quantiles à considérer et/ou les intervalles les plus pertinentes (sur une plage possible de 100, allant de 0 à 1), Birch et Miller (2006)¹⁰⁹ notent que la littérature ne montre pas d'approche théorique pour trancher sur ce sujet. Ainsi, nous avons été plutôt enclins à choisir une plage assez large pour représenter la population entière par intervalles de 5%, partant du 5% le plus faible jusqu'au 5% le plus fort. Ceci pour le cas de chacun des tests ici considérés.

¹⁰⁹ Birch, E.R. ; Miller, P. ; 2006 ; *Student outcomes at University in Australia: a Quantile Regression approach* ; Australian Economic Papers ; Blackwell Publishing Ltd/University of Adelaide and Flinders University ; 17 pp.

ANNEXE MÉTHODOLOGIQUE N°2

2. La correction par biais d'auto-sélection et les raisons invoquées pour le choix d'école

La catégorisation entre établissements que nous avons établi pour nos analyses est censée refléter une situation existante de forte segmentation du système éducatif chilien, autant d'un point de vue social qu'académique. Bien que nos variables cibles –les interactions déclarées et le bien-être à l'école et la confiance en soi scolaire – semblent moins affectées par cette segmentation ou plutôt, moins dépendantes de celles-ci, il nous est apparu prudent de tenir compte que les familles, de par leurs caractéristiques et leur valorisation des différents facteurs, font des choix et/ou ont des probabilités différenciées face aux choix disponibles.

La segmentation la plus importante semble reliée à la sélectivité académique des établissements, puisqu'elle permet à certains d'entre eux de ne travailler qu'avec une population d'élèves triée d'un point de vue de leurs capitaux culturels et de leur préparation académique préalable. Pour leur part, certaines familles font des efforts, déployant des stratégies pour y placer leurs enfants. Le fait que ce type de différenciation, par une sélectivité axée sur le plan académique, soit présente dans toutes les sous-catégories traditionnelles d'établissements renforce à notre avis le choix d'un tel indicateur. Certes, les stratégies des familles sont difficiles à connaître d'une manière objective, mais il est assez probable qu'elles soient liées à certaines caractéristiques observables et aux informations disponibles, ainsi qu'aux situations dans lesquelles ces familles évoluent.

2.1 Les bonnes raisons des familles chiliennes

L'équipe d'examineurs de l'OCDE (2004) indiquait que lors des visites aux écoles, tant les chefs d'établissement du privé que ceux des établissements publics soulignaient que « *s'ils cherchaient d'une certaine manière à faire concurrence aux autres écoles en s'efforçant d'améliorer les résultats aux tests, ils recouraient en fait principalement à d'autres moyens pour attirer les familles.*» (OCDE, 2004, p. 191). Ces mêmes chefs d'établissements ajoutaient aussi que les facteurs décisifs semblant faire l'attraction du privé étaient :

- le niveau socioéconomique moyen du groupe de pairs
- l'absence/présence d'élèves présentant des problèmes de comportement graves
- la possibilité d'accéder à l'enseignement secondaire au sein du même établissement (dans la perspective de l'examen d'entrée à l'université)

A ces constats plutôt qualitatifs et venant du côté de l'offre, s'ajoute, d'une part, le fait que les familles ne semblent pas se procurer d'information suffisante pour faire des choix sur la base des résultats académiques, et d'un autre côté, que les principales raisons de choix ne seraient pas académiques stricto sensu (e.g. Elacqua et al., 2006 ; Raczynski et al. 2009 ; cf. chapitre n°6). De fait, la proximité territoriale arrive en tête des facteurs de choix déclarés par les familles, et ceci en regardant autant le comportement des parents que leurs réponses aux questionnaires. Poussés par ces résultats, nous avons voulu avoir une vision générale sur le sujet à partir des données disponibles. Le tableau C32 suivant, illustre une comparaison de raisons de choix invoquées par les parents d'élèves de 8^{ème} grade, et ceci pour les questionnaires de l'an 2000 et de l'an 2004.

Les opinions des parents répondent ainsi à la question : « *Quels ont été les trois raisons principales pour lesquelles vous avez décidé d'inscrire l'élève dans cet établissement* ». Comme il est possible d'apprécier, il existe une différence entre 2000 et 2004 par rapport à la question formulée. Cela rend difficile une comparaison rigoureuse, mais retenons que dans les deux cas, arrivent en tête des raisons qui ne tiennent pas explicitement de la qualité des établissements. Parmi les cinq raisons les plus considérées en 2004, nous trouvons que plus de 63% des parents la déclarent comme « raison de choix » la proximité établissement-maison. Ceci est conforté par 49% des parents déclarant que leurs enfants vont à pied à l'école. Ensuite, arrive le « coût » en termes de frais de scolarisation, avec près de 51% d'adhésion. Puis, nous trouvons le fait que « *l'établissement donne des valeurs* » (45%) à côté du « *prestige* » (44%), pour avoir enfin, le fait d'avoir (ou avoir eu) un ou plusieurs membres de la famille déjà inscrits à cette école, avec 42% de préférences. Notons que le « *prestige* » reste une dimension plutôt floue, bien qu'il soit possible de penser qu'elle regroupe un ensemble des situations reliées aux résultats académiques des établissements. Les propositions-réponses portant sur les résultats annuels du SIMCE, c'est-à-dire à la qualité mesurée des résultats académiques de l'établissement, ne représentent en 2004 que 7,2% de ceux qui n'ont coché que trois options (et 9,5% si l'on considère le total de répondants).¹¹⁰ De plus, notons qu'en 2000, les raisons invoquées pour choisir l'école étaient fort semblables, quoi que l'option « *l'établissement a des bons enseignants* » arrive à avoir 45% de préférences, ce qui peut être vu indirectement comme une mesure de qualité. Enfin, en ce qui concerne l'appréciation directe et explicite sur les pairs que l'élève rencontrera à l'école, l'affirmation « *bon niveau socioéconomique et culturel des élèves* » réunit en 2004 7% de préférences contre 9% en 2000. Toujours est-il que le « *prestige* » invoqué, avec respectivement 44% et 37%, peut bel et bien considérer aussi cette dimension de choix.

¹¹⁰En ce qui concerne les résultats au test d'entrée à l'université ou PSU — *Prueba de Selección Universitaria* [Examen de selección universitaria] — il n'y a que 4,4% des préférences. Néanmoins, il faut ajouter que tous les établissements n'offrent pas le niveau secondaire.

Tableau C32

Principales raisons invoquées par les parents en 2000 et en 2004 comme étant importantes pour inscrire à l'établissement actuel, leurs enfants.
En pourcentage. (8^{ème} grade 2000 et 2004)

Proposition	2000	2004
<i>n° de répondants</i>	143696	260427
Proximité à la maison	54	63
Coût adéquat	46	51
Prestige	37	44
Autres familiers y sont / ont été scolarisés	36	42
L'établissement donne des Valeurs	34	45
Journée Scolaire Entière (JEC)	10	21
Bon niveau socioéconomique et culturel des élèves	9	7
Bons résultats SIMCE	6	7
Bons résultats PSU	4	4
Seul établissement dans la Commune	4	2
L'établissement a des bons Profs	45	-
Project de l'établissement	14	-
Autre	-	12
Seul établissement à accepter l'élève	-	2

Source : élaboration de l'auteur à partir des bases des données SIMCE 2004 et SIMCE 2000. Pour l'an 2004 le total des réponses obtenues considère 99,2% des parents répondants leur questionnaire. Les pourcentages montrés ne correspondent toutefois qu'à 68%, c'est-à-dire, à ceux qu'ont correctement cochés les trois options demandées, ni plus ni moins. Toutefois, dans le total ces proportions ne varient pas significativement. Pour l'an 2000, le pourcentage de répondants au total est de 74,4% dont 77% a bien coché les trois options demandées et, également, le total des répondants ne fait pas varier les résultats ici illustrés.

Remarquons finalement que nous n'allons utiliser ces réponses que comme des variables auxiliaires pour mieux contrôler l'impact des facteurs socioéconomiques sur le choix de l'école.

2.2 L'objectivation d'une mesure possible de correction de choix d'établissement

Nous avons décidé d'appliquer un facteur de correction par choix d'établissement car il est fort probable que les variables familiales et individuelles censées expliquer les scores des élèves soient aussi à la base des choix référés au type d'établissement. Ne pas prendre en compte ces phénomènes de sélection et d'auto-sélection pourrait nous amener à des estimations biaisées, notamment en ce qui tient à l'efficacité et la performance des différents types d'établissements, et enfin, en ce qui tient aux cadres d'existence des phénomènes qui nous intéressent. Bref, bien qu'ici l'impact net associé aux types d'établissement (i.e. s'ils sont sélectifs ou pas) ne soit pas la cible principale de nos analyses, un tel contrôle nous semble pertinent et nécessaire. Le cas

échéant, nous risquons de réduire encore plus le regard sur les facteurs pouvant altérer, de même qu'expliquer, les différences observées entre élèves, entre groupes d'élèves et entre établissements.

Une procédure assez reconnue et utilisée dans la littérature et celle du système dit de « double équation » proposé par James Heckman (1979)¹¹¹ comme une possible solution face au biais provoqué dans les estimations par l'existence d'échantillons non-aléatoires. Ceci amènerait à la présence de variables omises du fait que l'absence de certains individus dans les échantillons serait le résultat des caractéristiques a priori non-observables mais qui auraient une influence sur les résultats et/ou les comportements analysés. Dans le cas chilien, cette procédure est souvent utilisée dans la littérature économique en éducation – cf. Elacqua et al. (2007), Contreras et al. (2007), mais aussi Aedo et Sapelli (2001), Sapelli et Vial (2005) ou Tokman (2001) – visant à analyser, à partir des modèles du type MCO, les différences d'efficacité entre secteur public et secteur privé (cf. chapitre n°6). L'idée de base de cette utilisation est de revenir à une situation aléatoire dans laquelle les élèves iraient dans les établissements publics ou privés indistinctement de leurs caractéristiques socioéconomiques. Ceci en estimant un facteur ou coefficient qui capture la propension à réaliser un choix donné.

Comme nous l'avons souligné, ladite correction est faite à partir de la procédure proposée par Heckman (1979). Celle-ci considère, d'abord, un modèle Probit pour estimer, à partir des variables explicatives disponibles, la probabilité d'atteindre un type d'établissement donné (ici, académiquement sélectif). Dans notre cas, parmi les variables disponibles nous avons utilisé une bonne partie des mêmes que nous avons inclus après dans les différents modèles explicatifs des scores individuels des élèves. Mais, nous avons ajouté d'autres car ceci est une exigence de la procédure. Les variables choisies, mises dans le modèle rapporté au tableau C33, ont donc été :

- Le nombre d'établissements sélectifs par Km2 dans la commune où habite la famille de l'élève
- Les raisons déclarées par les parents pour avoir choisi l'établissement ;¹¹²
- Les caractéristiques socioéconomiques et scolaires des familles et de l'élève ;
- Des possessions spécifiques de capital culturel (cf. ordinateur, imprimante, internet) en plus de l'indice classique ;
- Les attitudes des parents, observées ici une fois le choix opéré. Toujours est-il que celles-ci reflètent, très probablement, des approches à la scolarité de leurs enfants ;

¹¹¹Heckman, J.J.; 1979 ; *Sample selection bias as a specification error* ; *Econometrica*, Vol.47, n°1; pp. 153-162

¹¹² Tout en y incluant, par exemple, le fait de déclarer que l'élève se mobilise à pied à l'établissement fréquenté (i.e. un choix plutôt « de facto » et pas seulement déclaré).

- Certaines caractéristiques observables de l'établissement (e.g. si l'école existait déjà en 2000, le nombre d'enseignants, le caractère mixte ou pas, l'équipement de base, la moyenne d'expérience des enseignants ou la préoccupation de ceux-ci vis-à-vis des sujets spécifiques, etc.) ;

Avec ces groupes de variables, nous avons construit la première équation suivant les pas indiqués par Robin (2003 ; p. 270-271)¹¹³ pour obtenir après une variable censée exprimer la propension de choix d'une école sélective. Au total, la formule du modèle *probit* (ou *normit*)¹¹⁴ qui résume cette première équation est :

$$P(\text{Etab. Sél.}) = \delta_1 F_i + \delta_2 X_i + \delta_3 \left(n^\circ \frac{ES}{\text{km}^2} \right)_{\text{commune } i} + \delta_4 RC_i + \delta_5 AP_i + \delta_5 CE_{ei} + v_i$$

où P est la probabilité prédite d'atteindre un établissement académiquement sélectif ; F et X sont des vecteurs des caractéristiques familiales et individuelles de l'élève (i) ; et où « n°ES/Km² » correspond au nombre d'établissements sélectifs par Km² dans la commune où habite l'élève i . RC correspond à l'ensemble des raisons de choix données par la famille de l'élève i ; AP prétend capturer les attitudes des parents envers la scolarité de l'enfant et CE correspond aux caractéristiques de l'établissement que de manière présumable les parents peuvent avoir observé ; et enfin v signale un terme d'erreur supposé aléatoire. Les différents δ indiquent des vecteurs des coefficients associés aux groupes de variables.

Dans le cas de la première variable considérée, disons qu'elle s'inscrit dans un raisonnement du type « variable instrumentale » (i.e. corrélée avec le choix mais *pas beaucoup* avec les acquis) et reste tout à fait fondamentale dans l'analyse. Admettons qu'elle est toutefois biaisée car, comme l'indique Belleï (2009), le choix d'emplacement des écoles n'est pas aléatoire non plus. En effet, la corrélation entre cette mesure et le niveau de revenu moyen par commune reste positive ($r=+0,48$). Toutefois, elle n'a qu'une relation réduite avec les scores aux tests, et qui tend à s'effacer lorsqu'on contrôle les caractéristiques des élèves. Mais force est de reconnaître que nos deux groupes de variables *extra*, censées expliquer le choix d'établissement (i.e. les raisons de choix mises en valeur et l'offre disponible) sont aussi, en quelque sorte, reliés aux acquis des élèves, même si elle l'est de manière indirecte. Il faudrait tenir compte de ceci car il est possible que la fonction de contrôle ne soit pas finalement bien remplie.

¹¹³ Robin, J.-M.; 2003 ; *Éndogénéité et variables instrumentales dans les sciences sociales* ; Actes des journées de méthodologie statistique, Document de Travail INSEE-Méthodes n°101 ; pp. 217-276

¹¹⁴ La fonction *probit* est définie comme : $g(p)=\Phi^{-1}(p)$ correspondant à l'inverse de la fonction cumulative standard normale.

Tableau C33

Coefficients du modèle cherchant à expliquer la probabilité de choisir/aller dans un établissement académiquement sélectif. (n=275 936 élèves) (2004 – 8^{ème} grade)

<i>Réf</i>	<i>Modalité active</i>	<i>Coefficient Estimé (NORMIT)</i>	<i>Prob. ChiSq</i>
-	Constante	-0.65	0.0000
-	N° Étabs. Sélectifs par KM ² (Commune)	0.41	0.0000
<i>Les parents valorisent</i>			
non	Proximité à la maison	-0.07	0.0000
non	Autres familiers y sont / ont été scolarisés	-0.02	0.0008
non	Prestige	0.15	0.0000
non	Bons résultats SIMCE	0.01	0.2799
non	Bons résultats PSU	0.09	0.0000
non	Bon NSE et culturel des élèves	0.05	0.0000
non	L'établissement donne des Valeurs	0.20	0.0000
non	Journée Scolaire Entière (JEC)	-0.22	0.0000
non	Coût adéquat	-0.07	0.0000
non	Seul établissement à accepter l'élève	-0.09	0.0000
non	Seul établissement dans la Commune	-0.37	0.0000
non	Autre	-0.08	0.0000
<i>Caractéristiques de l'établissement</i>			
non	L'élève va à pied à l'établissement	-0.07	0.0000
non	Établissement MIXTE	-0.45	0.0000
non	L'établissement existait déjà en 2000	-0.02	0.1485
-	N°Total d'enseignants	0.003	0.0000
-	N°Total du personnel d'appui (UTP)	0.001	0.7304
non	Etab. préoccupé des Résultats (avis des enseignants)	0.16	0.0000
non	Etab. préoccupé des Valeurs (avis des enseignants)	-0.07	0.0000
non	Etab. préoccupé de la Discipline (avis des enseignants)	-0.09	0.0000
non	Moyenne Années d'expérience des enseignants	-0.03	0.0000
-	% d'enseignants ayant une spécialisation	-0.08	0.0000
Université	Moy. Expectatives des Enseignants (PRIMAIRE)	-0.49	0.0000
	Moy. Expectatives des Enseignants (SEC. TECHN.)	-0.92	0.0000
	Moy. Expectatives des Enseignants (SEC. S&H)	-0.76	0.0000
	Moy. Expectatives des Enseignants (Sup. TECHN.)	-0.35	0.0000
non	L'établissement a une Bibliothèque	-0.04	0.0001
non	L'établissement a Internet	0.00	0.8879
non	L'établissement a un Laboratoire de sciences	0.31	0.0000
-	Taille moyenne des classes	0.02	0.0000
-	Taux d'encadrement (n°élèves/prof.)	0.01	0.0000
Rural	Condition urbaine	0.34	0.0000
<i>Caractéristiques et comportements de l'élève et sa famille</i>			
Garçon	Fille	-0.02	0.0001
non	Redoublant UNE fois avant 2004	-0.08	0.0000
non	Redoublant DEUX fois avant 2005	-0.13	0.0000
non	Redoublant ≥ TROIS fois avant 2004	-0.17	0.0009
non	L'élève est passé par Pré-Kinder	0.09	0.0000
non	L'élève est passé par Kinder	0.23	0.0000
Foyer en Couple	Foyer MERE seule	-0.01	0.0995
	Foyer PERE seul	-0.01	0.5206
	Foyer sans Père ni Mère (Autre)	-0.09	0.0000
non	L'élève a au moins un frère	0.04	0.0000

**Modèle expliquant la probabilité d'aller dans un établissement académiquement sélectif
(continuation)**

Réf	Modalité active	Coefficient Estimé (NORMIT)	Prob. ChiSq
	Indice de NSE de l'élève	0.23	0.0000
Entre 1 et 10 Livres à la maison	Zéro Livres à la maison	-0.05	0.0635
	Entre 11 et 50 Livres à la maison	0.08	0.0000
	Entre 51 et 100 Livres à la maison	0.11	0.0000
	Entre 101 et 200 Livres à la maison	0.13	0.0000
	Plus de 200 Livres à la maison	0.14	0.0000
non	Foyer équipé avec Imprimante	0.02	0.1883
non	Foyer équipé avec Ordinateur	0.04	0.0001
non	Foyer équipé avec Internet	0.06	0.0000
Université	Attentes des parents (PRIMAIRE)	-0.06	0.0460
	Attentes des parents (SECONDAIRE)	-0.10	0.0000
	Attentes des parents (SUP. TECHN.)	-0.05	0.0000
> 4 fois par an	S'entretient avec l'enseignant en charge du cours (Jamais)	0.19	0.0000
	S'entretient avec l'enseignant en charge du cours (1 ou 2 fois par an)	0.27	0.0000
	S'entretient avec l'enseignant en charge du cours (3 ou 4 fois par an)	0.10	0.0000
Jamais	S'entretient avec un autre enseignant (1 ou 2 fois par an)	0.00	0.5417
	S'entretient avec un autre enseignant (3 ou 4 fois par an)	-0.05	0.0000
	S'entretient avec un autre enseignant (>4 fois par an)	-0.08	0.0000
Est allé à très peu de réunions	N'est allé à aucune réunion de parents l'année dernière	-0.20	0.0000
	Est allé à quelques réunions l'année dernière	-0.10	0.0000
	Est allé à la plupart des réunions l'année dernière	-0.14	0.0000
	Est allé à toutes les réunions l'année dernière	-0.09	0.0000
> 4 fois par an	Révisé le cahier des notes (JAMAIS)	-0.04	0.0909
	Révisé le cahier des notes (1 ou 2 fois par an)	-0.21	0.0000
	Révisé le cahier des notes (3 ou 4 fois par an)	-0.12	0.0000
-	Dimension d'Appui	-0.04	0.0000
-	Dimension d'Exigence	0.02	0.0000

Source : élaboration de l'auteur. Lecture : le fait d'avoir plus de 200 livres à la maison augmente positivement et significativement la probabilité d'aller dans un établissement sélectif par rapport à la modalité de référence (Avoir entre 1 et 10 livres à la maison).¹¹⁵

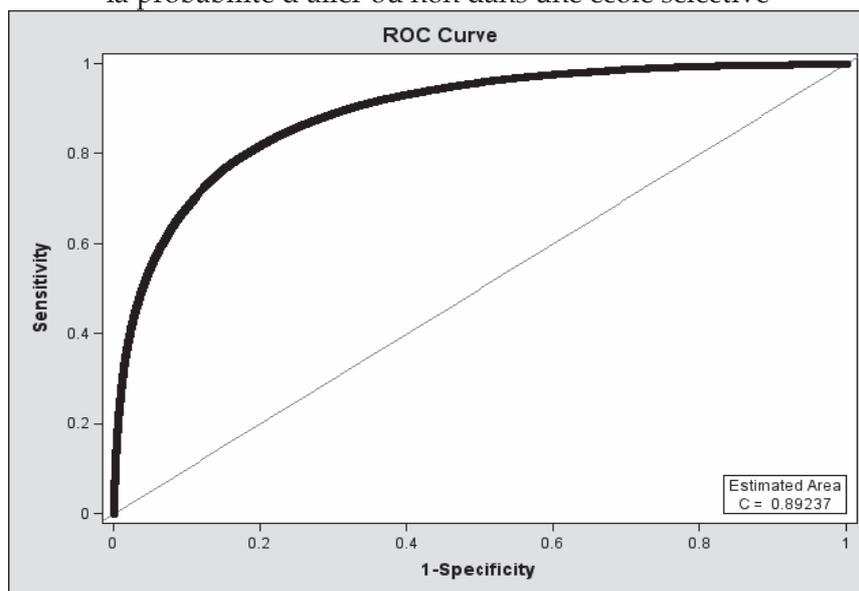
Ainsi, à partir de l'estimateur linéaire de cette probabilité, issu du modèle montré au tableau C33, nous pouvons construire une nouvelle variable qui s'intègre aux équations expliquant les scores aux tests d'apprentissage. Cette mesure, nommée Inverse du Ratio de Mills (i.e. le ratio entre la fonction normale de densité de l'estimateur de choix et la fonction cumulée de celui-ci), illustre la prédisposition, de par les caractéristiques, préférences et situations

¹¹⁵ Des indices d'ajustement du Modèle sont, parmi d'autres : a) Max-Rescaled $R^2=0,56$ et D de Sommers $=0,78$, des valeurs élevées illustrant de fortes liaisons entre les variables indépendantes et la variable à expliquer. b) Test global du Logarithme de Vraisemblance (-2 Log L) passant de 358089,1 à 215454,2, significatif à $p<0,0000$. c) Test de Hosmer-Lemeshow donnant un Khi^2 de 124,2 significatif à $p<0,0000$. Ce qui n'est pas, *a priori*, un bon signe car cela revient à dire qu'il y a des différences significatives à l'heure de reproduire, par le modèle, la distribution originale (en dix sous-groupes égales). Pourtant, bien que très significative, cette valeur du test Khi^2 n'est pas réellement très élevée compte tenu de la taille de l'échantillon.

environnementales déjà nommées, de choisir une école académiquement sélective de la part d'une famille donnée par rapport à un élève donné. A cet égard Wooldridge (2006)¹¹⁶ pointe que « *essentially, the inverse Mills ratio is an omitted variable, and it is generally correlated with the elements of "x"* » (p. 598). Bref, cet Inverse du Ratio de Mills viendrait, théoriquement, à « remettre les comptes à zéro » d'un point de vue des caractéristiques qui expliquent, à la fois, autant le choix d'école que les résultats aux tests. Rappelons que nous nous servons ici d'une version simple de cette procédure (i.e. à choix binaire), forcément imparfaite, d'autant plus que les établissements sélectifs d'un point de vue académique ne sont qu'une partie de ceux existants et analysés ici. Nous entreprenons donc cette démarche à titre exploratoire et d'apprentissage, sachant que cet indice serait davantage important si notre sujet portait directement sur l'efficacité différentielle entre types d'établissement. Les résultats de cette modélisation subsidiaire sont communiqués ici en bas, où nous détaillons les coefficients du modèle construit pour cette première équation.

Graphique C46

Courbe de ROC du modèle détaillé au Tableau C32, cherchant à expliquer la probabilité d'aller ou non dans une école sélective



Source : élaboration de l'auteur.

Enfin, une vision de synthèse de ce modèle est offerte par la courbe de ROC (*receiver operating characteristic*) sur le graphique n°46 qui est définie notamment comme la surface entre l'axe des abscisses et la courbe dessinée par le modèle exprimant le compromis entre sensibilité et spécificité (Panik, 2009 ; p. 230). Cette surface parle alors de l'exactitude ou précision prédictive

¹¹⁶Wooldridge, J.M.; 2006; *Introductory Econometrics: a modern approach* (3th Ed.); Mason ; 890 pp.

du modèle sachant que des hauts niveaux de sensibilité et de spécificité équivalent à dire que le modèle est en mesure de bien prédire, respectivement, les deux événements (1 ou 0, ici, aller ou non à une école académiquement sélective). La surface de ROC ici trouvée est qualifiée comme très satisfaisante (0,89). Notons qu'en enlevant les caractéristiques des établissements, cette valeur serait descendue à 0,86.

Pourtant, et malgré ces bons résultats, nous ne pouvons pas négliger le fait que l'information contenue ici a, en partie, quelque chose de redondant, bien qu'elle soit capable de nous renvoyer la réalité avec une bonne précision. Le fait que les familles choisissent et/ou qu'elles puissent envoyer leurs enfants dans des établissements sélectifs reste un sujet qui nécessite davantage d'exploration et plus de précisions et de travail dans le choix des variables. Nous retiendrons toutefois le caractère illustratif de la démarche ici entreprise en attendant de voir si, une fois dans les équations visant à expliquer les scores, elle est capable de moduler les résultats des estimations des acquis scolaires.

ANNEXE MÉTHODOLOGIQUE N°3

3. La construction de nouvelles variables sous la forme de Dimensions/Indices composites, et/ou, par des regroupements par Clusters

La grande quantité d'informations dont nous disposons, les liens inévitables que les variables entretiennent entre elles, ainsi que le caractère et la forme de certaines sources (questions posées aux acteurs), nous obligent à trouver des procédures adaptées à l'amalgame des informations convergentes et semblables sous la forme de dimensions composites. Pour ce faire, nous allons avoir recours à des procédures largement utilisées dans la littérature, tout en essayant de le faire de manière consistante et avec conscience de leurs conditions d'usage et de leurs limitations et possibilités.

3.1 L'analyse factorielle

Au-delà de l'usage des indices de consistance interne entre variables, tels que l'Alpha de Cronbach¹¹⁷ (Demeuse et Henry, 2004),¹¹⁸ pour parvenir à nos objectifs nous allons travailler d'une façon prioritaire avec ce qu'on connaît comme des analyses factorielles (AF). Falissard (1998) indique que l'AF « fait partie de la famille des modèles linéaires avec variables latentes » (p. 263) qui tente d'expliquer les corrélations entre variables à l'aide de « facteurs communs sous-jacents qui ne peuvent pas être observés directement » (p. 263). L'analyse factorielle (AF) aurait été largement confondue avec l'analyse en composantes principales (ACP) en circonstances qu'il s'agit des procédures différentes, quoi que proches (Falissard, 1998). L'ACP a pour but d'expliquer le plus de variance possible à partir d'un nombre restreint d'axes ou de composantes orthogonales (i.e. indépendantes). Pour ce faire, elle décompose la matrice de corrélations en tenant compte de l'ensemble de la variance des items. En revanche l'AF, tout en décomposant la matrice des corrélations, ne tient compte que de la variance commune à l'ensemble des items. Son but est de reproduire le plus fidèlement possible la matrice originale

¹¹⁷ Cet indicateur nous permet de tester en première l'existence d'une dimension commune de par la fidélité ou la consistance interne de l'instrument utilisé. L'hypothèse de base dans ce cas est que chacune des propositions est une mesure différente d'un même concept ce qui devrait nous amener à des valeurs élevées de cet indice. Toujours est-il que cela ne démontre pas que l'échelle testée mesure plus ou moins parfaitement ce qu'elle prétend mesurer, on s'approche plutôt de savoir si celle-ci est relativement cohérente avec elle-même. De cette façon la fidélité viendrait nous informer sur combien chaque item ou affirmation reflète de manière plutôt équivalente une dimension donnée, à l'occurrence le bien-être scolaire. Le coefficient Alpha de Cronbach est alors considéré comme la borne inférieure de la fidélité d'un test ou d'un ensemble d'items voués à mesurer un même concept ou aptitude (Durand, 2003). Si l'on recherche des valeurs du Alpha autour du 0,7, ils seraient à considérer, avec une plus ou moins grande prudence, à partir de 0,5 (Demeuse et Henry, 2004 ; Durand 2003).

¹¹⁸ Demeuse, M.; Henry, G.; 2004 ; *La fidélité des mesures, Chap. III* ; In Introduction aux Théories et aux Méthodes de la mesure en Sciences Psychologiques et en Sciences de l'Éducation, Demeuse, M.; pp. 149-162

des corrélations (Durand, 2003).¹¹⁹ Elle admet tant l'indépendance que la corrélation entre les axes ou dimensions factorielles extraites. Ces deux types d'analyse (l'ACP et l'AF) peuvent être soumis, après l'extraction initiale, à des procédures dites de rotation, cherchant le meilleur ajustement possible des données aux dimensions proposées (i.e. à des sous-ensembles plus fortement corrélés à l'intérieur de ceux-ci), et donc, visant une meilleure interprétation des résultats (Demeuse, 2003).¹²⁰ Néanmoins, seulement l'AF permet de penser les dimensions comme étant non-indépendantes, hypothèse qui est souvent plus adaptée aux analyses en sciences sociales car il est difficile de poser, *a priori*, que les dimensions à analyser ne seront pas corrélées entre elles. Enfin, Falissard (1998) note que vis-à-vis des *variables latentes*, l'ACP donne des résultats approximatifs à ceux de l'AF, mais qu'elle est en vérité plus utile pour regrouper des variables en clusters homogènes (p. 280).

Or, la notation générale pour un facteur commun extrait à partir de différentes variables est :

$$\gamma_{ij} = x_{i1}b_{j1} + x_{i2}b_{j2} + \dots + x_{iq}b_{jq} + e_{ij}$$

Où γ_{ij} c'est la valeur de la i th observation pour la j th variable ; x_{i1} est la valeur de la i th observation pour le k th facteur commun, tandis que b_{j1} correspond au coefficient de régression du k th facteur commun prédisant la j th variable. Enfin, e_{ij} est assumé comme la valeur de la i th observation pour le j th facteur unique. Pour sa part q correspond au nombre de facteurs communs. Ce facteur commun (souvent nommé facteur tout court) est donc non observé et l'on suppose qu'il contribue à la variance des variables observées (au moins deux). En revanche le facteur unique contribue à la variance d'une seule variable observée, ce qui implique qu'il existe un facteur unique pour chacune de ces variables.

Compte tenu que notre but est autant d'extraire des variables latentes que de comprendre les proximités entre items ou variables, nous utiliserons davantage l'AF pour l'extraction des axes et des coordonnées factorielles liées à chaque individu.

3.1.1 La démarche suivie

¹¹⁹Durand, C. ; 2003 ; *L'analyse factorielle et l'analyse de fidélité notes de cours et exemples* ; Université de Montréal, Département de Sociologie ; 30 pp.

¹²⁰Demeuse, M.; 2003 ; *Les analyses factorielles. Quelques applications d'analyses factorielles dans le domaine de l'éducation et de la formation* ; Séminaire IREDU (16/12/2003).

Une fois considérées ces premières mesures, nous pouvons continuer la réalisation d'une ou de plusieurs analyses factorielles.¹²¹ Un premier indicateur, mis à part l'Alpha de Cronbach, sur la possibilité de réaliser une AF est l'indice MSA (*Kaiser's Measure of Sampling Adequacy*). Cet indice est à la fois estimé pour l'ensemble mais aussi pour chaque variable ou item.¹²² Il nous informe sur le degré de pertinence des variables considérées pour réaliser l'analyse, indiquant jusqu'à quel point s'agit-il d'un ensemble cohérent des variables qui permet de constituer de mesures composites. Bref, cet indicateur nous informe, pour chaque variable, du compromis entre l'apport spécifique de chacune et la corrélation entre celles-ci. Malgré cette préférence manifeste pour l'AF, nous avons réalisé aussi des ACP pour nous donner une référence. Plus encore, il existe plusieurs alternatives parmi les analyses factorielles, ce qui nous conduira à comparer certaines options en quête des solutions factorielles stables et cohérentes entre elles (Durand, 2003), de manière à rendre plus robustes nos estimations.

Nous avons ainsi testé quatre types d'extraction factorielle à l'aide de la Procédure FACTOR de SAS®, y compris l'ACP avec rotation. A savoir, nous avons utilisé dans tous les cas : (a) la méthode de Moindres Carrés Non-Pondérés (*Unweighted Least Squares, ULS*), assez adapté aux échelles d'attitude, (Durand, 2003) ; (b) la méthode nommée Alpha-factorisation, cherchant à dégager un maximum de cohérence intra-dimensions pour ainsi construire des échelles ; et enfin, (c) la méthode de Maximum de Vraisemblance (*Maximum Likelihood, ML*) qui pose en général certaines restrictions comme la normalité des variables, mais qui présente d'autres avantages. Finalement, nous avons considéré l'extraction par composantes principales équivalente à l'ACP.

¹²¹ Disons, au passage, que d'autres procédures quelque peu plus grossières pour construire des échelles à partir d'items Likert et souvent citées dans la littérature, auraient pu être mises en œuvre avec des résultats plutôt semblables à ceux qui nous avons obtenu par l'analyse factorielle. Par exemple, en estimant la moyenne arithmétique ou la somme de tous les items-propositions, pour les garder comme échelles continues. Pourtant, une telle solution reste autant limitée que limitante, puisqu'elle est peu discriminatoire. D'un côté, de par le caractère essentiellement ordinal des degrés d'accord, et d'un autre côté, de par le poids équivalent assigné à chaque item et réponse. En effet, ce type de démarche ne rend pas compte des lignes de force à l'intérieur des données, autant dans le cas des individus que dans celui des variables. En revanche, l'AF permet de surmonter ces limitations car elle regarde l'ensemble de valeurs réponses et, en même temps, les variances par item et par individu. Ainsi, cette méthode définit un nuage de points multidimensionnel plus explicatif et proche du réel. En outre, une caractérisation par clusters aurait laissé plusieurs zones d'ombre, par exemple, vis-à-vis des « mal classés ». Nous retiendrons donc le recours aux solutions factorielles en vue de la construction des larges échelles continues capables d'aider à mieux classer les élèves.

¹²² Le logiciel SPSS® note le MSA global comme Indice KMO (Mesure de Kaiser-Meyer-Olkin). Il propose aussi d'autres indicateurs d'ajustement tels que le Déterminant de la Matrice (Durand, 2003) pour lequel on cherche à avoir une valeur petite mais différente de zéro, ainsi que le test de sphéricité de Bartlett offert aussi par SAS® pour la méthode d'extraction par Maximum de Vraisemblance (ou ML). Pour cette dernière méthode SAS® propose aussi des tests additionnels (e.g. *Tucker and Lewis's reliability coefficient, Akaike's Information Criterion*, ou encore, le *Schwarz's Bayesian Criterion*).

Pour qualifier la qualité de l'extraction des coordonnées, il existe l'analyse des résidus entre la matrice originale et la matrice reproduite par la solution factorielle. L'important dans ce cas c'est d'avoir un pourcentage faible des différences (i.e. peu de valeurs $\geq |0,05|$ (Durand, 2003)), cela indiquant que le but de l'analyse est bien satisfait.

Après l'extraction, nous avons demandé des rotations sous l'option HK (Harris-Kaiser) avec la sub-option HK Power égal à 0 (cf. SAS® 9.1.3 Help and Documentation). Cette dernière équivaut à dire que nous cherchons des clusters des variables les plus indépendants possibles. Des clusters où chaque item est davantage lié à un seul facteur (i.e. une solution dite *simple* et non *complexe* (Durand, 2003)), mais tout en admettant à la fois la possibilité d'une corrélation entre les dimensions ou variables latentes ainsi trouvées – tout cela, bien entendu, si les données l'admettent et le montrent –.

En outre, pour chacune des quatre analyses faites nous n'avons demandé que deux dimensions car demander davantage nous semblait hors de nos propos. D'ailleurs, notons qu'il existe plusieurs critères de choix pour le nombre des facteurs à retenir. La littérature converge sur deux critères base pour qualifier l'importance des axes ou dimensions à choisir (Falissard, 1998) ; d'un côté, les facteurs qui dépassent le nombre « 1 » comme référence, ce qui équivaut à dire qu'il recouvre plus d'information qu'une seule variable ou item (Vivanco, 1996).¹²³ D'un autre côté, l'observation des angulations de la courbe de valeurs propres associées à chaque facteur, pour ne considérer que ceux qui se placent avant la première inflexion (Falissard, 1998). Nous avons cherché à satisfaire les deux critères à la fois pour chaque facteur, sachant que souvent, nous allons décider de ne travailler qu'avec un. D'ailleurs, les tests associés exclusivement à la méthode d'extraction du maximum de vraisemblance sont fort utiles pour déterminer spécifiquement s'il existe, au moins, un facteur commun mais aussi, pour savoir si deux facteurs (ou plus selon la demande) suffisent à résumer les rapports existantes dans les données.

En outre, rappelons qu'une fois l'extraction faite, on cherche une matrice de saturations ou des poids (*loadings*) avec des valeurs absolues importantes et illustrant l'association entre les items ou variables et un facteur. L'idéal étant ainsi que chaque variable soit prioritairement associée à un seul facteur et n'ayant de corrélation supérieure à $|0,3|$ qu'avec un seul (Durand, 2003). Ces saturations s'interprètent comme des coefficients de corrélation (Vivanco, 1996) et les procédures de rotation prétendent justement rendre plus aisée cette tâche en essayant de maximiser l'association de chaque variable à un seul facteur à la fois (Demeuse, 2003). Mais quand les facteurs ne sont pas orthogonaux ou indépendants, l'interprétation de la matrice des

¹²³ Vivanco, M.; 1996 ; *Analisis multivariable* ; Universidad de Chile, Departamento de Sociología, Serie Cuadernos de Trabajo N°3 ; 192 pp.

saturations ou des charges ne peut être faite comme s'il s'agissait des coefficients de corrélation (Vivanco, 1996). De même, l'interprétation de la variance prise en compte par chaque facteur – une nomenclature issue des ACP – devient complexe (Durand, 2003).

La structure de référence (*reference structure*) constitue un autre outil important pour comprendre le sens de la solution factorielle proposée. Elle sert aussi à nommer les axes extraits, notamment quand on a eu recours à une rotation du type oblique ou non orthogonale. Cette structure, souvent mise en graphique, correspond aux corrélations semi-partielles entre les variables et le facteur commun à chacune, tout en contrôlant l'effet ou la relation de celles-ci avec les autres facteurs. Ainsi, une fois que les dimensions ou facteurs ont été extraits et soumis à rotation il faut les nommer en considérant « *l'intensité des relations qu'ils entretiennent avec les variables originales* » (p. 114 ; Vivanco, 1996). Ceci est un travail fortement dépendant de l'intuition et des informations préalables dont dispose le chercheur (Durand, 2003).

Finalement, les communautés (*communality*) finales nous informent de la part de variation commune entre un item donné et le reste (Vivanco, 1996), c'est-à-dire, la part de variance de chacune des variables provenant de facteurs communs (Falissard, 1998). Cette notion est consubstantielle à l'AF car elle s'occupe justement d'interpréter la part de variation qui est commune à toutes les variables et non pas la variance totale du nuage de points, ce qui correspond plutôt à l'ACP. Au total, les valeurs que nous prendrions en compte dans ce cas correspondent à l'après rotation, c'est-à-dire, aux valeurs maximales possibles d'obtenir par la solution factorielle choisie, considérant que plus elles sont élevées, meilleure est la solution. Notons néanmoins que ceci reste un critère parmi d'autres.

En définitive, avec toutes ces informations nous avons fait des arbitrages pour choisir laquelle des solutions nous prenions, sachant que la comparaison entre les hypothèses de départ et les résultats obtenus, ainsi que leur interprétation, ne repose que « *dans la capacité du chercheur à les décoder* » (p. 8 ; Durand, 2003). Nous avons considéré les solutions factorielles ayant moins de résidus entre la matrice de corrélations originale et celle reconstruite et ayant, à la fois, des corrélations élevées entre les facteurs proposés et les variables à l'état original. De même, nous avons tenu en compte l'allure des courbes de distribution des valeurs de chaque dimension extraite, cherchant à conjuguer des hypothèses plausibles avec une tendance, même si faible, à la normalité des nouvelles valeurs produites. Enfin, nous avons considéré les *communalities* que chaque solution factorielle avait produit en cherchant la maximisation de ces valeurs.

Ainsi, nous essayons de prendre de manière judicieuse les choix concernant chaque mesure, dimension ou échelle extraite par AF. Consignons que nous n'irons pas dans ce travail vers les analyses factorielles de type confirmatoire (i.e. se servant des *structural equation modeling using*

covariance analysis) ni des analyses du type *path*. Toutefois, nous reconnaissons dans ces deux cas des chemins tout à fait souhaitables notamment lorsque nous avons pu constater que plusieurs des dimensions subjectives analysées par nous, restent reliées entre elles.

3.2 L'analyse de correspondances multiples et la construction des clusters

Nous avons aussi utilisé trois autres méthodes visant, d'une part, à mieux comprendre les liens entre variables, et d'autre part, à regrouper les élèves, les parents ou les enseignants dans des catégories rendant compte de la réalité de manière plus synthétique. Ceci notamment dans le cas des variables catégorielles, ordinales ou non.

Dans le cas des pratiques d'étude et d'interaction entre élèves, compte tenu de la nature dichotomique de ces données (cf. encadré n°3, Tome I), nous avons estimé, tout d'abord, autant des coefficients Tau de Kendall que des coefficients Tétrachoriques (Bonett et Price, 2005),¹²⁴ cherchant à chiffrer l'intensité des relations entre variables. Au-delà de l'étroite et presque parfaite correspondance entre ces deux types d'estimation (r de Pearson entre ces mesures de +1,0 et +0,97), nous avons préféré communiquer le dernier type de coefficient. En effet, celle-ci s'adapte mieux aux données d'un point de vue théorique, car la corrélation tétrachorique considère que derrière un comportement sous forme binaire, il existe une variable latente continue. Cela semble bien le cas des pratiques d'étude posant que les déclarations faites par les élèves sous-entendent un continuum (e.g. rapporté au niveau de compétences) et qui expliquerait l'apparition de ce comportement à un moment donné.

Ensuite, nous avons eu recours à l'analyse de correspondances multiples (réalisé à l'aide de la Proc CORRESP de SAS®, option MCA, sans usage des variables supplémentaires). Puis enfin, nous nous sommes servis des méthodes vouées à la construction des clusters. Pour le premier cas, Demeuse (2003) indique que l'analyse des correspondances sert à comprendre les liens entre variables qualitatives à partir d'un tableau de contingence ou d'un tableau disjonctif (ici, un tableau de Burt). Ce type de démarche présente l'avantage de ne pas poser d'hypothèses préalables (Falissard, 1998) tout en décrivant de manière synthétique des grands ensembles des données. Ceci en projetant sur un plan factoriel les différentes inerties ou distances à la situation d'indépendance¹²⁵ associées à chacune des modalités des variables.¹²⁶

¹²⁴Bonett, D.; Price, R.; 2005 ; *Inferential Methods for the Tetrachoric Correlation Coefficient*; Journal of Educational and Behavioral Statistics, Vol.30, n°2 ;pp. 213-225

¹²⁵ Notons que le concept d'inertie serait analogue à celui de variance chez l'ACP. Il correspond à la proportion d'information fournie par la décomposition du Khi2 (SAS Institute, 2003).

¹²⁶ De plus, le logiciel SAS® fourni une table sommaire avec les contributions partiales aux inerties, table qui permet de qualifier facilement quel axe ou dimension est mieux associé à chaque variable.

Pour le cas des regroupements, afin d'obtenir rapidement des catégories clairement identifiables, nous avons eu recours à la méthode dite de « clusters » (Falissard, 1998). Ces méthodes de calculs, rapides dans les logiciels spécialisés, nécessitent néanmoins que le chercheur définisse a priori le nombre de clusters ou regroupements à chercher. Dans notre cas nous avons décidé arbitrairement d'examiner entre trois et quatre sous-groupes, que ce soit pour les élèves, enseignants ou parents. Cela ne signifie pas, bien entendu, qu'il ne puisse pas exister d'autres sous-divisions encore, mais ces quantités de sous-groupes (3 ou 4) nous semblent adéquates, eu égard de nos besoins de parcimonie dans les modèles d'analyse, mais aussi, pour simplifier l'interprétation des résultats. Pour y parvenir nous avons eu recours en parallèle à deux logiciels (SPAD® et SAS®) ¹²⁷ pour choisir enfin les découpages qui nous semblaient les plus discriminants et à la fois les plus clairs vis-à-vis des questions posées (différencier les acteurs selon leurs pratiques). Notons que si la procédure FASTCLUS de SAS® est basée sur l'algorithme de *nuées dynamiques* (Falissard, 1998), la procédure PARTI-DECLA de SPAD® l'est sur l'algorithme de *centres mobiles* (CISIA-CERESTA, 2001),¹²⁸ plus relié à l'ACP. Selon Falissard (1998) ces techniques ne diffèrent « *que par quelques détails* » (p. 205). Dans les deux cas, le principe est d'itérer de façon à trouver un maximum de variance interclasses et un minimum intra-classes, et de s'arrêter une fois que des améliorations de ceci ne semblent plus possibles.

¹²⁷ A chaque fois, nous avons comparé les deux résultats (SAS® et SPAD®) et nous avons choisi, dans certains cas, les résultats fournis par la procédure PARTI-DECLA de SPAD®, tandis que pour d'autres on a pris ceux fournis par SAS® (Proc FASTCLUS). En effet, les résultats n'étant pas véritablement éloignés les uns des autres, nous avons privilégié la facilité à l'interprétation des regroupements (i.e. la clarté de ceux-ci par rapport aux variables originales et leur cohérence avec la théorie).

¹²⁸ CISIA-CERESTA ; 2001 ; SPAD-Base, Aide à l'interprétation ; 126 pp.

ANNEXES « D »

MODELISATION DES ACQUIS SCOLAIRES

Tableau D34

MATHS	Principales Statistiques (n=233262)							Variance par Niveau		
	VARIABLE	MEAN	STD	MAX	Q3	Median	Q1	MIN	ETAB.	CLASSE
Score en Mathématiques	0.00	1.00	3.0	0.7	0.0	-0.7	-2.8	29.5	5.0	65.5
AIDANT	0.28	0.45	1	1	0	0	0	0.8	0.5	98.7
AIDÉ	0.41	0.49	1	1	0	0	0	1.3	1.2	97.4
Et SEUL	0.35	0.48	1	1	0	0	0	2.0	1.3	96.7
Et LIVRE	0.12	0.32	1	0	0	0	0	5.5	1.8	92.7
Bas BEE	0.32	0.47	1	1	0	0	0	1.8	1.5	96.7
BEE Moyen	0.33	0.47	1	1	0	0	0	0.2	0.6	99.2
Basse CSS	0.33	0.47	1	1	0	0	0	1.9	1.2	97.0
CSS Moyenne	0.33	0.47	1	1	0	0	0	0.3	0.6	99.1
Étab a Labo Sciences	0.50	0.50	1	1	1	0	0	100	-	-
N°élèves/Prof	21.87	6.16	94.1	25.6	21.6	18.0	0.4	100	-	-
MUNIS	0.08	0.27	1	0	0	0	0	100	-	-
PSUBV NS NFP	0.06	0.23	1	0	0	0	0	100	-	-
PSUBV NS FP	0.11	0.31	1	0	0	0	0	100	-	-
PSUBV S NFP	0.03	0.18	1	0	0	0	0	100	-	-
PSUBV S FP	0.22	0.41	1	0	0	0	0	100	-	-
Privé PAYANT	0.07	0.26	1	0	0	0	0	100	-	-
Années d'Expérience (PM)	20.62	11.50	50.0	30.0	21.0	12.0	0.0	65.9	34.1	-
Postitre (PM)	0.28	0.45	1	1	0	0	0	64.5	35.5	-
Post-grade (PM)	0.03	0.18	1	0	0	0	0	48.6	51.4	-
Autre Travail (PM)	0.16	0.37	1	0	0	0	0	53.5	46.5	-
Niv Prepa Moyen	0.22	0.41	1	0	0	0	0	61.6	38.4	-
Niv Prepa Haut	0.56	0.50	1	1	1	0	0	67.5	32.5	-
Attentes SupTECH (PM)	0.21	0.41	1	0	0	0	0	50.0	50.0	-
Attentes UNIV (PM)	0.19	0.39	1	0	0	0	0	72.1	27.9	-
Préoccupé DISCIPLINE (PM)	0.51	0.50	1	1	1	0	0	59.3	40.7	-
Préoccupé RÉSULTATS (PM)	0.56	0.50	1	1	1	0	0	59.7	40.3	-
Couvert Prog MOYENNE (PM)	0.36	0.48	1	1	0	0	0	49.3	50.7	-
Couvert Prog HAUTE (PM)	0.35	0.48	1	1	0	0	0	62.8	37.2	-
% Heures Prepa Cours (PM)	10.25	11.60	90.0	13.3	6.7	2.6	0.0	43.9	56.1	-
Trav Coopérat Moyen (PM)	0.20	0.40	1	0	0	0	0	78.8	21.2	-
Trav Coopérat Haut (PM)	0.62	0.49	1	1	1	0	0	82.9	17.1	-
Plusieurs PRATs (PM)	0.26	0.44	1	1	0	0	0	60.6	39.4	-
PRATs Indiv (PM)	0.38	0.49	1	1	0	0	0	55.7	44.3	-
Taille de la Classe	35.26	6.45	49.0	41.0	36.0	31.0	20.0	80.1	19.9	-
MOY Années Scol Mères (-i)	10.38	2.46	17.8	11.7	10.0	8.5	4.0	94.8	5.1	0.1
STD Années Scol Mères (-i)	2.55	0.52	5.1	2.9	2.5	2.2	0.0	25.6	72.4	2.0
% Classe (-i) Bas BEE	31.89	11.54	87.1	39.5	31.3	23.8	0.0	29.3	70.0	0.7
% Classe (-i) Haute CSS	33.69	11.41	79.5	41.0	33.3	25.7	0.0	33.7	65.6	0.7
Fille	0.50	0.50	1	1	1	0	0	10.5	1.9	87.6
Redoublant	0.15	0.36	1	0	0	0	0	5.9	3.6	90.5
Pré-scolaire	0.45	0.50	1	1	0	0	0	17.0	0.6	82.3
Foyer NON Couple	0.29	0.45	1	1	0	0	0	2.1	0.5	97.4
Au moins a 1 Frère	0.79	0.41	1	1	1	1	0	1.0	1.0	98.0
Années Scol Mère	10.39	3.56	23.0	13.0	10.0	8.0	1.0	45.1	0.9	54.0
Bas Capital Cult	0.28	0.45	1	1	0	0	0	16.3	1.3	82.4
Moyen Capital Cult	0.40	0.49	1	1	0	0	0	2.7	0.8	96.4
Parents Aidants	0.35	0.48	1	1	0	0	0	0.7	0.5	98.8
Parents Exigeants	0.21	0.40	1	0	0	0	0	1.5	0.6	97.9
Parents Souples	0.10	0.31	1	0	0	0	0	0.4	0.7	98.8
Aime Maths Niv2	0.39	0.49	1	1	0	0	0	0.6	0.8	98.6
Aime Maths Niv3	0.46	0.50	1	1	0	0	0	2.6	1.9	95.5
Rapport Prof Maths	0.00	1.00	1.2	0.8	0.2	-0.6	-3.3	12.1	11.3	76.6
Étude M >= 'des fois' Hebdo	0.44	0.50	1	1	0	0	0	2.4	2.3	95.3
Croit ira SupTECH	0.18	0.38	1	0	0	0	0	4.7	0.8	94.5
Croit ira UNIV	0.60	0.49	1	1	1	0	0	19.2	1.6	79.2
Inv Ratio MILLS (Km²)	0.00	1.00	3.76	0.78	-0.10	-0.96	-1.24	90.0	0.4	9.6

Tableau D35

LANGAGE	Principales Statistiques (n=235094)							Variance par Niveau		
	VARIABLE	MEAN	STD	MAX	Q3	Median	Q1	MIN	ETAB.	CLASSE
SCORE en LANGAGE	0 00	1 00	2 6	0 7	0 0	-0 7	-3 1	23 0	4 9	72 1
AIDANT (L)	0 27	0 44	1	1	0	0	0	2 2	1 1	96 7
AIDÉ (L)	0 33	0 47	1	1	0	0	0	1 3	1 3	97 4
Et SEUL (L)	0 40	0 49	1	1	0	0	0	4 0	1 4	94 6
Et LIVRE (L)	0 16	0 36	1	0	0	0	0	6 0	1 9	92 0
Bas BEE	0 32	0 47	1	1	0	0	0	1 8	1 5	96 7
BEE Moyen	0 33	0 47	1	1	0	0	0	0 2	0 6	99 2
Basse CSS	0 33	0 47	1	1	0	0	0	1 8	1 2	97 0
CSS Moyenne	0 33	0 47	1	1	0	0	0	0 3	0 6	99 1
Étab a Labo de Sciences	0 50	0 50	1	1	1	0	0	100	-	-
N°élèves/Prof	21 87	6 16	94 1	25 6	21 6	18 0	0 4	100	-	-
MUNI S	0 08	0 26	1	0	0	0	0	100	-	-
PSUBV NS NFP	0 06	0 23	1	0	0	0	0	100	-	-
PSUBV NS FP	0 11	0 31	1	0	0	0	0	100	-	-
PSUBV S NFP	0 03	0 18	1	0	0	0	0	100	-	-
PSUBV S FP	0 22	0 41	1	0	0	0	0	100	-	-
Privé PAYANT	0 07	0 26	1	0	0	0	0	100	-	-
Années d'Expérience (PL)	20 68	11 98	50 0	31 0	22 0	10 0	0 0	59 5	40 5	-
Postitre (PL)	0 29	0 45	1	1	0	0	0	59 4	40 6	-
Post-grade (PL)	0 05	0 22	1	0	0	0	0	45 6	54 4	-
Autre Travail (PL)	0 13	0 34	1	0	0	0	0	57 2	42 8	-
Niv Prepa Moyen	0 36	0 48	1	1	0	0	0	51 1	48 9	-
Niv Prepa Haut	0 32	0 47	1	1	0	0	0	57 2	42 8	-
Attentes SupTECH (PL)	0 24	0 43	1	0	0	0	0	43 7	56 3	-
Attentes UNIV (PL)	0 21	0 41	1	0	0	0	0	71 1	28 9	-
Préoccupé DISCIPLINE (PL)	0 50	0 50	1	1	0	0	0	53 0	47 0	-
Préoccupé RÉSULTATS (PL)	0 53	0 50	1	1	1	0	0	55 8	44 2	-
Couvert Prog MOYENNE (PL)	0 38	0 48	1	1	0	0	0	42 0	58 0	-
Couvert Prog HAUTE (PL)	0 30	0 46	1	1	0	0	0	53 2	46 8	-
% Heures Prepa Cours (PL)	10 40	11 62	93 8	13 6	6 7	2 6	0 0	42 8	57 2	-
Trav Coopérat Moyen (PL)	0 16	0 36	1	0	0	0	0	73 3	26 7	-
Trav Coopérat Haut (PL)	0 69	0 46	1	1	1	0	0	79 5	20 5	-
Plusieurs PRATs (PL)	0 19	0 40	1	0	0	0	0	52 7	47 3	-
PRATs Indiv (PL)	0 51	0 50	1	1	1	0	0	49 0	51 0	-
PRATs Group (PL)	0 18	0 39	1	0	0	0	0	49 9	50 1	-
Taille de la Classe	35 25	6 44	49	41	36	31	20	80 1	19 9	-
MOY Années Scol Mères (-i)	10 38	2 46	17 8	11 7	10 0	8 5	4 0	94 8	5 1	0 1
STD Années Scol Mères (-i)	2 55	0 52	5 1	2 9	2 5	2 2	0 0	25 6	72 4	2 0
% Classe (-i) Bas BEE	31 91	11 55	87 1	39 5	31 3	23 8	0 0	29 4	69 9	0 7
% Classe (-i) Haute CSS	33 69	11 41	79 5	41 0	33 3	25 7	0 0	33 8	65 5	0 7
Fille	0 50	0 50	1	1	1	0	0	10 5	1 9	87 6
Redoublant	0 15	0 36	1	0	0	0	0	5 9	3 5	90 5
Préscolaire	0 45	0 50	1	1	0	0	0	17 3	0 6	82 1
Foyer NON Couple	0 29	0 45	1	1	0	0	0	2 1	0 5	97 4
Au moins a 1 Frère	0 79	0 41	1	1	1	1	0	0 9	1 0	98 0
Années Scol Mère	10 39	3 55	23 0	13 0	10 0	8 0	1 0	45 3	0 9	53 8
Bas Capital Cult	0 28	0 45	1	1	0	0	0	16 0	1 2	82 7
Moyen Capital Cult	0 41	0 49	1	1	0	0	0	2 8	0 8	96 3
Parents Aidants	0 35	0 48	1	1	0	0	0	0 7	0 4	98 8
Parents Exigeants	0 20	0 40	1	0	0	0	0	1 5	0 6	97 9
Parents Souples	0 10	0 31	1	0	0	0	0	0 5	0 7	98 8
Aime Lang Niv2	0 46	0 50	1	1	0	0	0	1 4	1 4	97 2
Aime Lang Niv3	0 40	0 49	1	1	0	0	0	5 8	3 2	91 0
Rapport Prof Lang	0 00	1 00	1 4	0 8	0 2	-0 6	-3 3	11 9	10 3	77 8
Étudie L >= 'des fois' Hebdo	0 40	0 49	1	1	0	0	0	3 5	2 1	94 4
Croit ira SupTECH	0 18	0 38	1	0	0	0	0	4 7	0 8	94 5
Croit ira UNIV	0 60	0 49	1 0	1 0	1 0	0 0	0 0	19 2	1 6	79 2
Inv Ratio MILLS (Km²)	0 00	1 00	3 76	0 78	-0 10	-0 96	-1 24	90 1	0 4	9 5

Tableau D36

Scs. de la NATURE	Principales Statistiques (n=235744)							Variance par Niveau		
	VARIABLE	MEAN	STD	MAX	Q3	Median	Q1	MIN	ETAB.	CLASSE
SCORE en Scs de la NATURE	0.00	1.00	3.0	0.7	0.0	-0.7	-2.7	26.9	5.3	67.8
AIDANT (ScN)	0.21	0.41	1	0	0	0	0	1.2	0.8	98.0
AIDÉ (ScN)	0.28	0.45	1	1	0	0	0	1.1	1.2	97.7
Et SEUL (ScN)	0.43	0.49	1	1	0	0	0	2.9	1.3	95.8
Et LIVRE (ScN)	0.23	0.42	1	0	0	0	0	5.5	2.2	92.4
Bas BEE	0.32	0.47	1	1	0	0	0	1.8	1.5	96.7
BEE Moyen	0.33	0.47	1	1	0	0	0	0.2	0.6	99.2
Basse CSS	0.33	0.47	1	1	0	0	0	1.9	1.2	97.0
CSS Moyenne	0.33	0.47	1	1	0	0	0	0.3	0.6	99.1
Étab a Labo de Sciences	0.50	0.50	1	1	1	0	0	100	-	-
N°élèves/Prof	21.87	6.16	94.1	25.6	21.6	18.0	0.4	100	-	-
MUNI S	0.08	0.26	1	0	0	0	0	100	-	-
PSUBV NS NFP	0.06	0.23	1	0	0	0	0	100	-	-
PSUBV NS FP	0.11	0.31	1	0	0	0	0	100	-	-
PSUBV S NFP	0.03	0.18	1	0	0	0	0	100	-	-
PSUBV S FP	0.22	0.41	1	0	0	0	0	100	-	-
Privé PAYANT	0.07	0.26	1	0	0	0	0	100	-	-
Années d'Expérience (PScN)	19.39	11.86	50.0	30.0	20.0	9.0	0.0	69.6	30.4	-
Postitre (PScN)	0.26	0.44	1	1	0	0	0	68.5	31.5	-
Post-grade (PScN)	0.04	0.19	1	0	0	0	0	61.9	38.1	-
Autre Travail (PScN)	0.16	0.37	1	0	0	0	0	59.5	40.5	-
Niv Prepa Moyen	0.29	0.46	1	1	0	0	0	62.7	37.3	-
Niv Prepa Haut	0.27	0.45	1	1	0	0	0	67.4	32.6	-
Attentes: SupTECH (PScN)	0.23	0.42	1	0	0	0	0	55.6	44.4	-
Attentes: UNIV (PScN)	0.20	0.40	1	0	0	0	0	71.4	28.6	-
Préoccupé DISCIPLINE (PScN)	0.52	0.50	1	1	1	0	0	63.1	36.9	-
Préoccupé RÉSULTATS (PScN)	0.54	0.50	1	1	1	0	0	63.8	36.2	-
Couvert Prog MOYENNE (PScN)	0.31	0.46	1	1	0	0	0	47.6	52.4	-
Couvert Prog HAUTE (PScN)	0.33	0.47	1	1	0	0	0	63.1	36.9	-
% Heures Prepa Cours (PScN)	10.36	12.35	93.8	13.3	6.7	2.3	0.0	52.4	47.6	-
Trav Coopérat Moyen (PScN)	0.24	0.42	1	0	0	0	0	84.2	15.8	-
Trav Coopérat Haut (PScN)	0.53	0.50	1	1	1	0	0	83.6	16.4	-
Tradit PRATs (PScN)	0.47	0.50	1	1	0	0	0	56.8	43.2	-
Plusieurs PRATs (PScN)	0.27	0.44	1	1	0	0	0	63.5	36.5	-
Taille de la Classe	35.25	6.44	49	41	36	30	20	80.1	19.9	-
MOY Années Scol Mères (-i)	10.38	2.46	17.8	11.7	10.0	8.5	4.0	94.8	5.1	0.1
STD Années Scol Mères (-i)	2.55	0.52	5.1	2.9	2.5	2.2	0.0	25.6	72.5	2.0
% Classe (-i) Bas BEE	31.92	11.55	87.1	39.5	31.3	23.8	0.0	29.4	69.9	0.7
% Classe (-i) Haute CSS	33.69	11.41	79.5	41.0	33.3	25.7	0.0	33.8	65.5	0.7
Fille	0.50	0.50	1.0	1.0	1.0	0.0	0.0	10.5	1.9	87.6
Redoublant	0.15	0.36	1	0	0	0	0	6.0	3.5	90.5
Précolaire	0.45	0.50	1	1	0	0	0	17.3	0.6	82.1
Foyer NON Couple	0.29	0.45	1	1	0	0	0	2.1	0.5	97.4
Au moins a 1 Frère	0.79	0.41	1	1	1	1	0	0.9	1.0	98.0
Années Scol Mère	10.38	3.55	23	13	10	8	1	45.3	0.9	53.8
Bas Capital Cult	0.28	0.45	1.0	1.0	0.0	0.0	0.0	16.0	1.2	82.8
Moyen Capital Cult	0.41	0.49	1	1	0	0	0	2.8	0.8	96.4
Parents Aidants	0.35	0.48	1	1	0	0	0	0.7	0.4	98.8
Parents Exigeants	0.20	0.40	1	0	0	0	0	1.5	0.6	97.9
Parents Souples	0.10	0.31	1	0	0	0	0	0.5	0.7	98.8
Aime ScScNat Niv2	0.45	0.50	1	1	0	0	0	0.7	0.9	98.4
Aime ScScNat Niv3	0.41	0.49	1	1	0	0	0	3.2	2.0	94.9
Étudie ScN >= 'des fois' Hebdo	0.38	0.49	1	1	0	0	0	2.8	1.9	95.4
Croit ira SupTECH	0.18	0.38	1	0	0	0	0	4.7	0.8	94.5
Croit ira UNIV	0.60	0.49	1	1	1	0	0	19.2	1.6	79.2
Inv Ratio MILLS (Km²)	0.00	1.00	3.76	0.78	-0.10	-0.96	-1.25	90.1	0.4	9.5

Tableau D37

Scs. SOCIALES	Principales Statistiques (n=232771)							Variance par Niveau		
	VARIABLE	MEAN	STD	MAX	Q3	Median	Q1	MIN	ETAB.	CLASSE
SCORE en Scs SOCIALES	0.00	1.00	2.7	0.7	0.0	-0.7	-2.8	24.2	4.2	71.6
AIDANT (ScS)	0.22	0.41	1	0	0	0	0	1.2	0.8	98.0
AIDÉ (ScS)	0.27	0.44	1	1	0	0	0	1.0	1.0	98.0
Et SEUL (ScS)	0.43	0.49	1	1	0	0	0	3.1	1.3	95.6
Et LIVRE (ScS)	0.26	0.44	1	1	0	0	0	5.8	2.4	91.8
Bas BEE	0.32	0.47	1	1	0	0	0	1.8	1.5	96.7
BEE Moyen	0.33	0.47	1	1	0	0	0	0.2	0.6	99.2
Basse CSS	0.32	0.47	1	1	0	0	0	1.9	1.2	97.0
CSS Moyenne	0.33	0.47	1	1	0	0	0	0.3	0.6	99.1
Étab a Labo de Sciences	0.50	0.50	1	1	1	0	0	100	-	-
N°élèves/Prof	21.87	6.16	94.1	25.6	21.6	18.0	0.4	100	-	-
MUNI S	0.08	0.27	1	0	0	0	0	100	-	-
PSUBV NS NFP	0.06	0.23	1	0	0	0	0	100	-	-
PSUBV NS FP	0.11	0.31	1	0	0	0	0	100	-	-
PSUBV S NFP	0.03	0.18	1	0	0	0	0	100	-	-
PSUBV S FP	0.22	0.41	1	0	0	0	0	100	-	-
Privé PAYANT	0.07	0.26	1	0	0	0	0	100	-	-
Années d'Expérience (PSS)	19.05	12.14	50.0	30.0	20.0	8.0	0.0	65.3	34.7	-
Postitre (PSS)	0.27	0.44	1	1	0	0	0	64.7	35.3	-
Post-grade (PSS)	0.06	0.24	1	0	0	0	0	66.1	33.9	-
Autre Travail (PSS)	0.16	0.36	1	0	0	0	0	62.6	37.4	-
Niv Prepa Moyen	0.14	0.35	1	0	0	0	0	63.2	36.8	-
Niv Prepa Haut	0.43	0.50	1	1	0	0	0	65.8	34.2	-
Attentes: SupTECH (PSS)	0.22	0.41	1	0	0	0	0	51.5	48.5	-
Attentes: UNIV (PSS)	0.20	0.40	1	0	0	0	0	73.8	26.2	-
Préoccupé DISCIPLINE (PSS)	0.52	0.50	1	1	1	0	0	60.4	39.6	-
Préoccupé RÉSULTATS (PSS)	0.53	0.50	1	1	1	0	0	58.7	41.3	-
Couvert Prog MOYENNE (PSS)	0.46	0.50	1	1	0	0	0	52.3	47.7	-
Couvert Prog HAUTE (PSS)	0.28	0.45	1	1	0	0	0	64.1	35.9	-
% Heures Prepa Cours (PSS)	10.47	11.98	93.8	13.6	6.7	2.5	0.0	46.7	53.3	-
Trav Coopérat Moyen (PSS)	0.23	0.42	1	0	0	0	0	78.1	21.9	-
Trav Coopérat Haut (PSS)	0.54	0.50	1	1	1	0	0	80.4	19.6	-
PRATs Indiv (PSS)	0.24	0.43	1	0	0	0	0	53.5	46.5	-
PRATs Group (PSS)	0.11	0.31	1	0	0	0	0	50.8	49.2	-
Plusieurs PRATs (PSS)	0.48	0.50	1	1	0	0	0	55.6	44.4	-
Taille de la Classe	35.26	6.45	49.0	41.0	36.0	31.0	20.0	80.1	19.9	-
MOY Années Scol Mères (-i)	10.39	2.46	17.8	11.7	10.0	8.5	4.0	94.8	5.1	0.1
STD Années Scol Mères (-i)	2.55	0.52	5.1	2.9	2.5	2.2	0.0	25.6	72.4	2.0
% Classe (-i) Bas BEE	31.89	11.54	87.1	39.5	31.3	23.8	0.0	29.3	70.0	0.7
% Classe (-i) Haute CSS	33.69	11.41	79.5	41.0	33.3	25.7	0.0	33.8	65.5	0.7
Fille	0.50	0.50	1	1	1	0	0	10.5	1.9	87.6
Redoublant	0.15	0.36	1	0	0	0	0	5.9	3.6	90.6
Pré-scolaire	0.45	0.50	1	1	0	0	0	17.0	0.6	82.3
Foyer NON Couple	0.29	0.45	1	1	0	0	0	2.1	0.5	97.4
Au moins a 1 Frère	0.79	0.41	1	1	1	1	0	1.0	1.0	98.0
Années Scol Mère	10.39	3.56	23.0	13.0	10.0	8.0	1.0	45.1	0.9	54.0
Bas Capital Cult	0.28	0.45	1	1	0	0	0	16.3	1.3	82.4
Moyen Capital Cult	0.40	0.49	1	1	0	0	0	2.7	0.9	96.4
Parents Aidants	0.35	0.48	1	1	0	0	0	0.7	0.5	98.8
Parents Exigeants	0.20	0.40	1	0	0	0	0	1.5	0.6	97.9
Parents Souples	0.10	0.31	1	0	0	0	0	0.5	0.7	98.8
Aime ScSSat Niv2	0.40	0.49	1	1	0	0	0	0.7	0.8	98.5
Aime ScSSat Niv3	0.47	0.50	1	1	0	0	0	2.5	2.2	95.3
Étudie SS >= 'des fois' Hebdo	0.41	0.49	1	1	0	0	0	2.7	2.1	95.2
Croit ira SupTECH	0.18	0.38	1	0	0	0	0	4.7	0.8	94.5
Croit ira UNIV	0.60	0.49	1	1	1	0	0	19.2	1.6	79.1
Inv Ratio MILLS (Km²)	0.00	1.00	3.76	0.78	-0.10	-0.96	-1.24	90.0	0.4	9.6

Les quatre tableaux précédents (D34 à D37) correspondent aux caractéristiques générales de la distribution de chaque variable introduite dans les modèles analysant les scores aux tests par discipline. Chacune d'elles considère une population spécifique mais qui est, à la rigueur, la même, car les différences sont dues aux élèves n'ayant pas passé quelques-uns des tests respectifs.¹²⁹ En effet, le nombre de classes et d'établissements est le même pour les quatre cas (7061 et 3565 respectivement), correspondant aux classes urbaines ayant plus de 19 élèves. La variance par niveau communiquée correspond à une analyse de variance faite à l'aide de la Proc NESTED de SAS®. Ce type d'analyse est, à un décimal près, tout à fait équivalent à un modèle vide estimé avec la Proc MIXED pour chaque variable. Notons que pour ce faire, les variables muettes ont été prises comme étant continues, ces résultats sont donc approximatifs et deviennent seulement illustratifs.

Dans la spécification de chaque modèle nous avons eu le soin de regarder à tous moments les degrés de colinéarité entre variables à l'aide des indices mentionnés dans la partie Annexes Méthodologiques. Pour chaque modèle complet, nous communiquons en bas l'indice VIF associé à chacune des variables introduites. Bien que nous n'allons pas détailler ici d'autres analyses concomitantes (Index d'état), ni non plus détailler l'évolution de ces indices au fur et à mesure que l'on estimait les différents modèles (de 1 à 15), disons que les variables les plus colinéaires ont trait au choix d'école et à la moyenne (-i) d'années de scolarité des mères par classe. Il n'y a pas des valeurs de l'Index d'état dépassant le chiffre 20 avant le modèle N°10 et ceci pour toutes les disciplines. De toute évidence, il faut faire attention à la relation entre composition sociale du public scolaire, choix d'école et caractéristiques de l'enseignement délivré. En effet, et même si cela tient en partie à l'ordre d'introduction des variables, c'est entre le modèle N°13 et le modèle N°14, que les indices de colinéarité s'accroissent de manière importante (i.e. frôlant les limites critiques, passant de $\approx 3,75$ à $\approx 5,25$ pour le VIF et de ≈ 33 à ≈ 85 pour l'Index d'état). Autrement dit, quand nous ajoutons les caractéristiques, attentes et pratiques des enseignants ainsi que les variables définissant le public par classe, ces variables montrent les relations fortes qu'elles entretiennent avec le type d'école choisi, ce qui reste tout à fait logique. Plus encore, notons que dans le modèle N°10 (intégrant en plus de variables cibles, les trois aspects concernant l'enseignant plus la composition de la classe) on apprécie déjà un Indice d'état maximal autour de 54, ce qui confirme l'idée que ces deux groupes de variables comportent de fortes relations entre elles.

¹²⁹ Afin de garder le plus d'informations originales possible nous n'avons pas voulu, pour ces analyses, fixer cette population à tous ceux ayant répondu aux quatre tests à la fois. De ce fait, l'échantillon de travail est composé, globalement, de 236113 élèves. Pourtant, cette restriction s'est imposée d'elle-même du moment où nous avons modélisé les dimensions subjectives, ramenant l'échantillon à 231 200 individus.

Cela dit, nous avons des raisons de penser que la qualité des estimations principales n'est pas affectée. D'abord, parce que la valeur du VIF n'est pas proche de 10 dans aucun cas, et ceci en tenant compte du nombre important des paramètres estimés (≈ 60). Puis, parce que les variables cibles montrent toutes de fortes singularités (i.e. $VIF < 1,7$ avec un VIF moyen pour les pratiques d'étude de 1,22, et de 1,39 pour les sentiments de confiance et de bien-être. Cela renforce à notre avis l'importance explicative de ces facteurs. D'ailleurs, leurs variations restent fortement attachées au niveau individu. Enfin, si les Indices d'état s'élèvent au delà de 25, dans aucun cas ils sont corrélés avec deux variables à la fois, même dépassant la valeur de 0,35 pour la composante respective (voir Annexes Méthodologiques). Ainsi, cela ne conseille pas, de manière spécifique, de faire sortir aucune variable des analyses. Finalement, à l'instar du logiciel STATA nous communiquons l'indice VIF moyen pour chaque modèle, ce qui montre de manière synthétique que la grande majorité de variables semble rester éloignée des possibilités de colinéarité affectant les estimations.

Enfin, notons que tous ces indices ont été obtenus en doublant nos estimations multiniveaux par des analyses du type MCO. Consignons aussi que l'ajout de termes d'interaction entre variables fait, en toute logique, accroître le VIF des variables, notamment pour ces termes, et parfois, pour les variables originales de base. Cela est le cas avec des scolarités des mères (sous forme individuelle ou en moyenne par groupe dans classe) avec un VIF très élevé (≈ 39). Pourtant cela n'arrive pas dans les deux catégories de capital culturel ($VIF \leq 3,7$).

Tableau D38

Indices VIF par Modèle

<i>Variables (Modèle N°14)</i>	<i>MATHS</i>	<i>LANGAGE</i>	<i>Scs. de la NATURE</i>	<i>Scs. SOCIALES</i>
AIDANT	1.25	1.28	1.21	1.23
AIDÉ	1.28	1.29	1.22	1.23
Et. SEUL	1.25	1.38	1.29	1.31
Et. LIVRE	1.05	1.06	1.08	1.09
Bas BEE	1.48	1.49	1.47	1.47
BEE Moyen	1.37	1.37	1.37	1.37
Basse CSS	1.67	1.65	1.61	1.61
CSS Moyenne	1.41	1.41	1.40	1.40
Étab a Labo Scs.	1.48	1.47	1.48	1.48
N°élèves/Prof	1.61	1.61	1.62	1.62
MUNI S	1.29	1.27	1.27	1.27
PSUBV NS NFP	1.14	1.14	1.14	1.13
PSUBV NS FP	1.53	1.53	1.54	1.54
PSUBV S NFP	1.17	1.17	1.17	1.17
PSUBV S FP	2.45	2.45	2.48	2.45
Privé PAYANT	2.95	2.98	2.97	2.98
Années d'Expérience	1.28	1.32	1.35	1.33
Postitre	1.06	1.05	1.06	1.06
Post-grade	1.04	1.10	1.04	1.09
Autre Travail	1.05	1.03	1.07	1.05
Niv Prepa Moyen	1.61	1.50	1.30	1.19
Niv Prepa Haut	1.85	1.79	1.46	1.33
Attentes: SupTECH	1.30	1.30	1.25	1.32
Attentes: UNIV	2.18	2.21	2.10	2.20
Préoccupé DISCIPLINE	1.02	1.04	1.04	1.04
Préoccupé RÉSULTATS	1.02	1.04	1.03	1.04
Couvert. Prog. MOYENNE	1.53	1.50	1.34	1.59
Couvert. Prog. HAUTE	1.77	1.88	1.52	1.75
% Heures Prepa Cours	1.07	1.07	1.06	1.08
Trav Coopérat. Moyen	1.73	1.72	1.55	1.57
Trav Coopérat. Haut	1.87	1.83	1.68	1.72
Taille de la Classe	1.59	1.59	1.60	1.60
MOY Années Scol. Mères (-i)	5.36	5.34	5.26	5.37
STD Années Scol. Mères (-i)	1.18	1.17	1.17	1.17
% Classe (-i) Bas BEE	1.30	1.30	1.30	1.30
% Classe (-i) Haute CSS	1.19	1.20	1.19	1.19

Indices VIF par Modèle (continuation)

<i>Variables (Modèle N°14)</i>	<i>MATHS</i>	<i>LANGAGE</i>	<i>Scs. de la NATURE</i>	<i>Scs. SOCIALES</i>
Fille	1.06	1.04	1.03	1.04
Redoublant	1.12	1.12	1.12	1.12
Préscolaire	1.19	1.20	1.19	1.19
Foyer NON Couple	1.08	1.08	1.08	1.08
Au moins a 1 Frère	1.06	1.06	1.06	1.06
Années Scol. Mère	2.02	2.03	2.03	2.02
Capital Culturel Bas	1.90	1.88	1.88	1.90
Capital Culturel Moyen	1.52	1.52	1.52	1.52
Parents Aidants	1.40	1.41	1.40	1.40
Parents Exigeants	1.39	1.41	1.39	1.39
Parents Souples	1.29	1.31	1.28	1.28
Aime la discipline Niv2 (moy)	2.30	2.57	2.40	2.55
Aime la discipline Niv3 (haut	2.66	2.93	2.59	2.77
Étudie >= 'des fois' Hebdo	1.10	1.13	1.10	1.10
Croit ira SupTECH	1.57	1.57	1.57	1.57
Croit ira UNIV	2.03	2.02	2.02	2.03
Inv.Ratio MILLS (Km²)	5.24	5.24	5.18	5.30
Rapport aux Enseignants	1.25	1.31	-	-
Plusieurs PRATs (PM)	1.41	-	-	-
PRATs Indiv (PM)	1.33	-	-	-
Plusieurs PRATs (PL / PSS)	-	2.59	-	2.11
PRATs Indiv. (PL / PSS)	-	2.80	-	1.89
PRATs Group. (PL / PSS)	-	2.27	-	1.52
Tradit. PRATs (PScN)	-	-	1.55	-
Plusieurs PRATs (PScN)	-	-	1.66	-
<i>Moyenne Indice VIF</i>	1.61	1.68	1.60	1.63
<i>Valeur maximale Index d'État</i>	85.9	86.4	84.5	85.7

Résidus des Modèles n°14 par discipline

Au niveau élève, sont communiqués les résidus standardisés (option résidus de Pearson conditionnels), qui suivent une tendance globale à la normalité (cf. graphique D47). Les résidus des niveaux supérieurs sont illustrés en bas (cf. graphique D48). Ces derniers correspondent à l'estimation faite, pour chaque classe et établissement, de leur écart respectif à la moyenne.

Enfin, la corrélation entre les scores réels et les scores prédits par les modèles multiniveaux choisis, sont sur le tableau C39 suivant :

Tableau C39

Relation entre scores réels et prédits
(Model N°14). Par discipline.

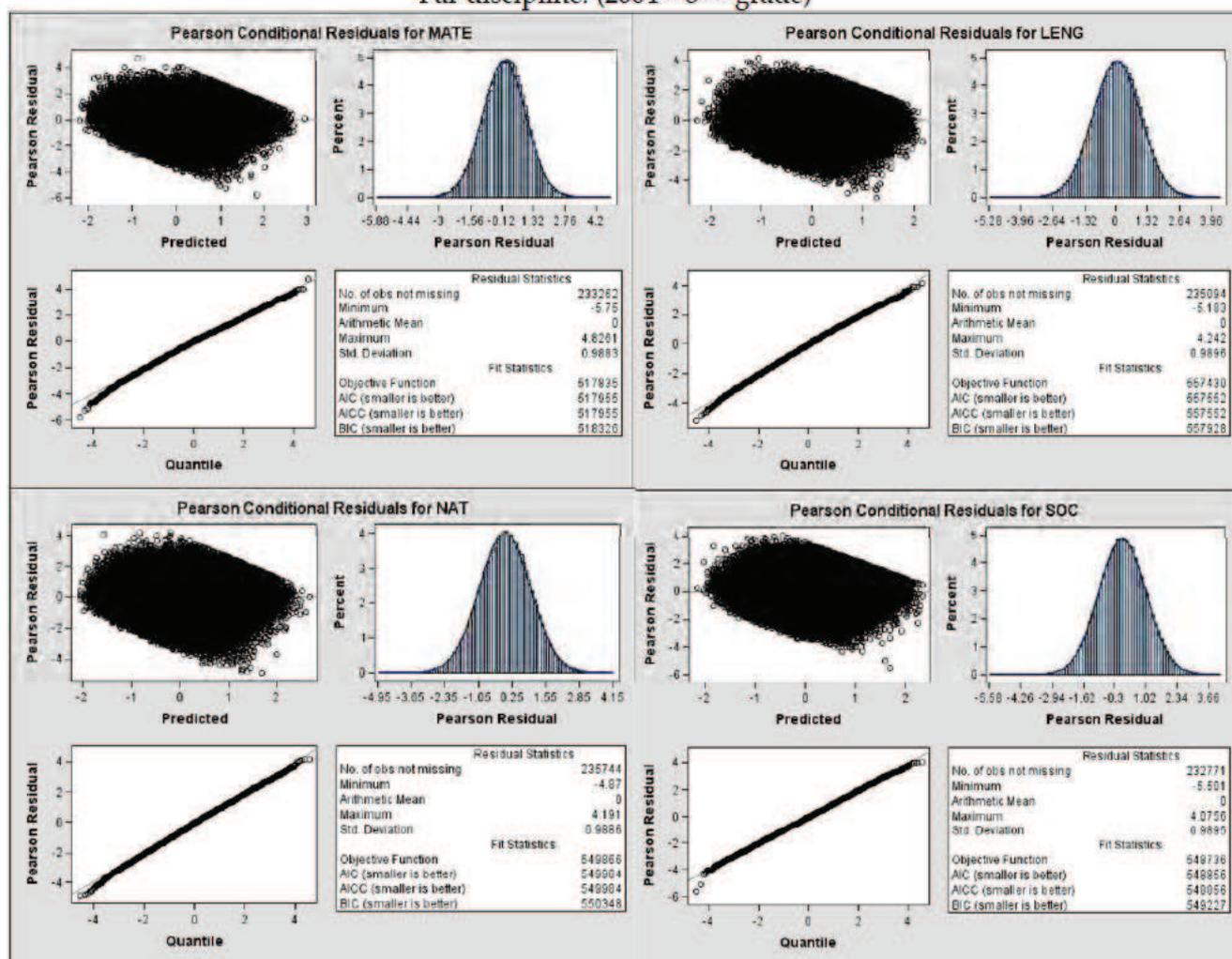
<i>Discipline</i>	<i>r</i>	<i>R²</i>
Maths	0.706	0.499
Langage	0.640	0.409
Scs. de la Nature	0.662	0.438
Scs. Sociales	0.647	0.419

Rappelons que Singer et Willet (2003, cités par Bressoux, 2008) indiquent ce dernier type d'estimation comme un possible indice de variance expliquée par les modèles.

Puis, enfin, les tableaux D40 et D41 considèrent les principaux résultats, en termes de coefficients fixes et aléatoires, de même que l'ajustement des modèles, en comparant les modèles n°14 originaux, mais appliqués autant sur l'échantillon de travail (i.e. avec remplissage des données concernant les dimensions BEE et CSS) que sur un échantillon qui élimine toutes les individus dont les données pour les principales dimensions subjectives n'étaient pas disponibles en entier à chaque fois. Cette dernière comparaison permet donc d'apprécier le biais d'estimation qui résulte du fait d'avoir rempli les données manquantes à cet égard.

Graphique D47

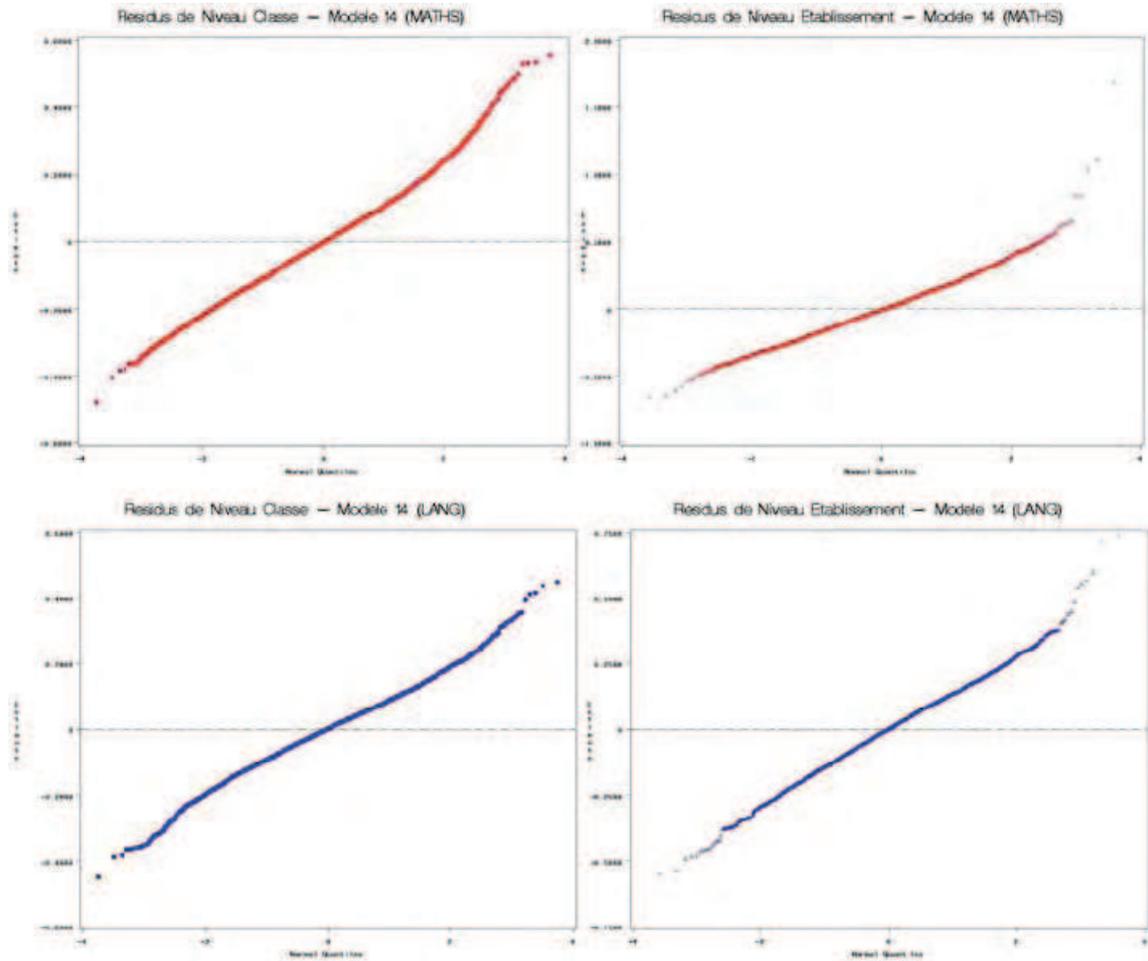
Résidus de Niveau Élève pour chaque modèle principal estimé (Modèles N°14).
Par discipline. (2004 - 8^{ème} grade)



Source : sorties SAS® pour les modèles multiniveaux respectifs estimés. Lecture : chaque cadran illustre différentes manières de regarder la distribution des résidus individuels pour chacun des modèles estimés visant à expliquer les résultats aux tests (quadrants en Z : maths, langage, sciences de la nature et sciences sociales).

Graphique D48

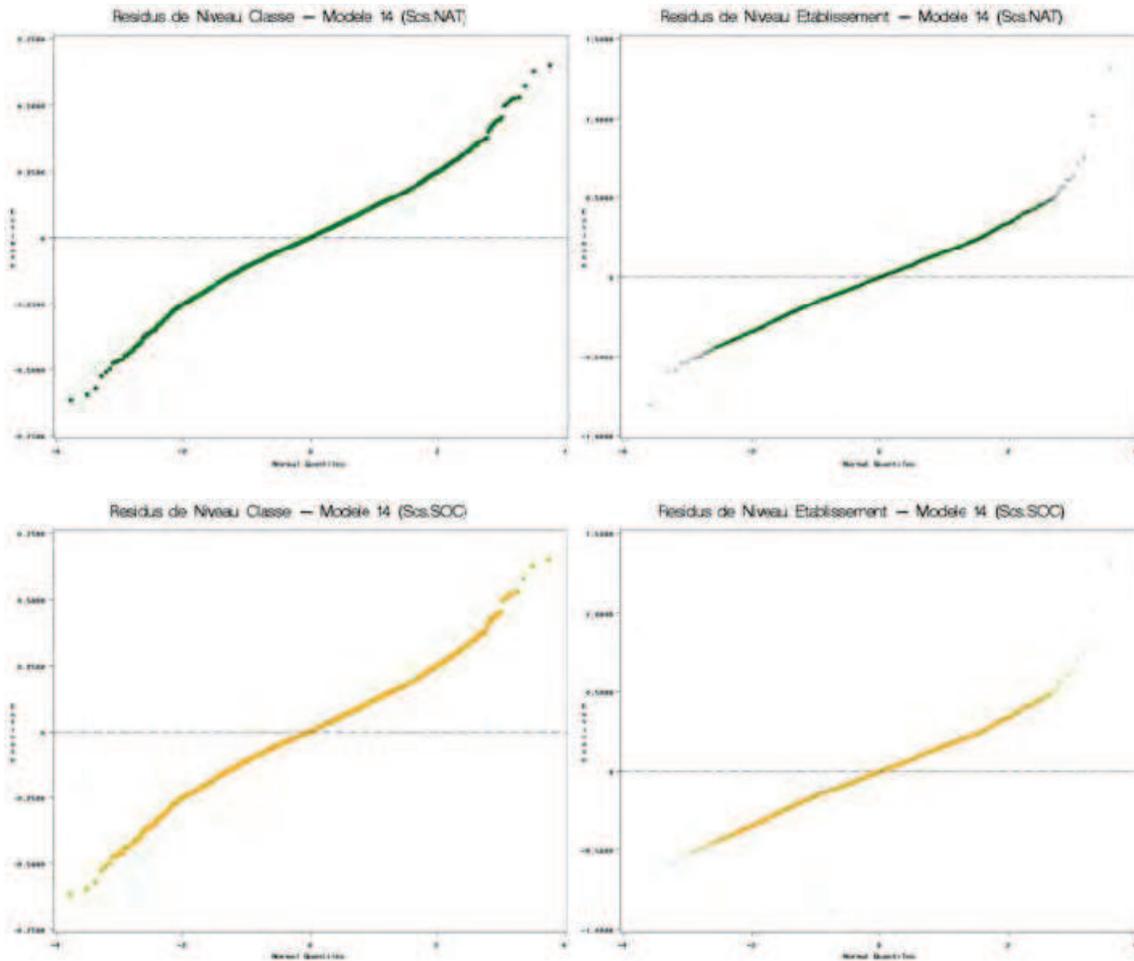
Distribution des résidus estimés par niveau classe et établissement
(Modèle N°14). Mathématiques et Langage



Source : élaboration de l'auteur à partir des résultats de la Proc MIXED pour le modèle N°14. A gauche, se trouvent les variations de niveau classe et à droite celles de niveau établissement.

Graphique E49

Distribution des résidus estimés par niveau classe et établissement
(Modèle N°14). Sciences de la Nature et Sciences Sociales



Source : élaboration de l'auteur à partir des résultats de la Proc MIXED pour le modèle N°14. A gauche, se trouvent les variations de niveau classe et à droite celles de niveau établissement.

Tableau D40

Comparaison entre les modèles (n°14) estimés avec l'ensemble des données (i.e. après remplissage) et les modèles estimés sur des données restreints (i.e. en excluant ceux dont manquait de l'information de base pour les dimensions de BEE et de CSS). (2004 – 8^{ème} grade ; échantillons d'élèves urbains dans des classes à taille ≥20) (PARTIE I)

	MATHS			LANGAGE			Sciences de la NATURE			Sciences SOCIALES		
	Données avec remplissage pour les ITEMS manquantes sur la Confiance en soi scolaire et le Bien-être à l'école									Échantillon Utilisé		
	Coeff	StdE	Sign	Coeff	StdE	Sign	Coeff	StdE	Sign	Coeff	StdE	Sign
PARTIE FIXE												
AIDANT	0.39	0.004	0.000	0.19	0.005	0.000	0.24	0.005	0.000	0.25	0.005	0.000
AIDÉ	-0.05	0.004	0.000	0.004	0.004	0.302	0.02	0.004	0.000	-0.02	0.004	0.000
Non	0.14	0.004	0.000	0.17	0.004	0.000	0.14	0.004	0.000	0.15	0.004	0.000
Et. SEUL	0.15	0.005	0.000	0.23	0.005	0.000	0.25	0.004	0.000	0.25	0.004	0.000
Et. LIVRE												
Haut BEE	-0.08	0.004	0.000	-0.16	0.005	0.000	-0.07	0.004	0.000	-0.08	0.004	0.000
BEE Moyen	-0.02	0.004	0.000	-0.06	0.004	0.000	-0.02	0.004	0.000	-0.03	0.004	0.000
Basse CSS	-0.16	0.004	0.000	-0.23	0.005	0.000	-0.22	0.004	0.000	-0.22	0.004	0.000
Haute CSS	-0.08	0.004	0.000	-0.11	0.004	0.000	-0.11	0.004	0.000	-0.11	0.004	0.000
<i>n</i>	233262			235094			235744			232771		
PARTIE ALEATOIRE												
	Mod Vide	Model T4		Mod Vide	Model T4		Mod Vide	Model T4		Mod Vide	Model T4	
Coeffs.												
Niv ETABLISS	0.29	0.05		0.23	0.04		0.27	0.05		0.24	0.05	
Niv CLASSE	0.05	0.03		0.05	0.02		0.05	0.03		0.04	0.02	
Niv ELEVE	0.66	0.51		0.72	0.60		0.68	0.58		0.72	0.59	
Variance	Var Totale	Var Mod Niv		Var Totale	Var Mod Niv		Var Totale	Var Mod Niv		Var Totale	Var Mod Niv	
entre Établissements	29.1%	81.9%		23.1%	84.9%		26.6%	82.4%		23.9%	81.1%	
entre Classes	5.0%	41.4%		4.9%	53.1%		5.3%	37.0%		4.2%	49.7%	
entre Élèves	65.8%	21.7%		72.0%	16.4%		68.1%	15.2%		71.9%	17.1%	

Lecture : en gris sont marqués les coefficients dont il existe au moins une incohérence avec les modèles estimés sur les données non remplies (échantillon restreint, tableau suivant).

Modèles intermédiaires (Entre les modèles vides et les modèles n°1)

Dans le chapitre n°11 nous avons délibérément omis les modèles faisant appel de manière séparée aux variables cibles (i.e. les pratiques d'étude et interactions d'une part, et les deux principales dimensions subjectives, d'autre part). Ainsi, dans ce chapitre, le modèle n°1 pour chaque discipline considère l'ensemble de ces variables prises en même temps. Or, nous notons ici-bas (tableaux D42 et D43) les résultats des modèles dits intermédiaires entre le modèle vide et ces modèles n°1 de manière à ne pas perdre de vue cette partie de l'évolution des coefficients reliant ces facteurs aux acquis scolaires. En effet, si notre intérêt principal est de regarder finalement l'évolution de ces coefficients vis-à-vis de l'ensemble d'autres facteurs, nous avons noté, à plusieurs reprises (cf. chapitre n°10) les relations entretenues entre celles-ci. Les modèles par discipline ici présentés montrent donc ces mêmes relations mais sous la forme des coefficients fixes dans des modèles de régression.

Dans certains cas, notamment pour celui de la Confiance en soi scolaire (CSS), des variances résiduelles négatives de niveau établissement apparaissent dans toutes les disciplines. De même, en maths les pratiques d'étude et d'interaction entre élèves montrent, quand on les prend de manière isolée, une légère variance négative de niveau classe. Certes, ces augmentations de la variance résiduelle ne sont finalement que très réduites sur le total, mais ce phénomène semble indiquer des différences existantes dans la distribution, par niveau, des variables. Autrement dit, suivant Bressoux (2008 ; p. 314), les moyennes des variables introduites semblent varier de manière importante d'un groupe à l'autre selon le niveau, ce qui peut se voir accentué parce que nous n'avons pas pris dans ce cas des mesures centrées mais plutôt catégorielles. Nous croyons en fait, que ces augmentations de la variance résiduelle nous indiquent que les rapports entre acquis scolaires et dimension de CSS se vérifient mieux en termes relatifs (i.e. au sein des unités classes/écoles) plus que sur une échelle nationale unique. D'ailleurs, la CSS semble davantage dépendante du niveau élève et reliée ainsi aux caractéristiques individuelles.

Tableau D43

Modèles de régression multiniveaux faisant appel aux variables cibles de manière isolée et combinée.
Par discipline (8^{ème} grade – 2004). (Partie II)

Impacts en SCIENCES de la NATURE

Impacts en SCIENCES SOCIALES

Coefficients Fixes					
Variable	Interactions	BEE	CSS	BEE+CSS	MODEL N°1
AIDANT	0.32				0.27
AIDÉ	0.00				-0.01
Et SEUL	0.19				0.18
Et LIVRE	0.30				0.28
Bas BEE		-0.15		-0.09	-0.08
BEE Moyen		-0.06		-0.02	-0.02
Basse CSS			-0.30	-0.28	-0.23
CSS Moyenne			-0.15	-0.14	-0.11

Coefficients Aléatoires						
NIVEAU	VIDE	Interactions	BEE	CSS	BEE+CSS	MODEL N°1
École	0.265	0.242	0.258	0.272	0.267	0.244
Classe	0.053	0.050	0.051	0.052	0.051	0.049
Élève	0.679	0.650	0.675	0.664	0.663	0.640

Coefficients Fixes					
Variable	Interactions	BEE	CSS	BEE+CSS	MODEL N°1
AIDANT	0.35				0.30
AIDÉ	-0.05				-0.05
Et SEUL	0.22				0.21
Et LIVRE	0.30				0.29
Bas BEE		-0.17		-0.11	-0.10
BEE Moyen		-0.08		-0.04	-0.04
Basse CSS			-0.32	-0.30	-0.24
CSS Moyenne			-0.16	-0.15	-0.12

Coefficients Aléatoires						
NIVEAU	VIDE	Interactions	BEE	CSS	BEE+CSS	MODEL N°1
École	0.238	0.213	0.230	0.245	0.239	0.214
Classe	0.042	0.039	0.040	0.041	0.040	0.038
Élève	0.716	0.679	0.712	0.699	0.698	0.667

Variances Modélisées par Niveau et au Total					
	entre Établissements	entre Classes...	entre Élèves...	Total Modèle MN	R ² (Réf. MCO)
	10.5%	6.1%	5.2%	6.5%	7.6%
	3.3%	4.2%	0.6%	1.4%	1.7%
	-2.6%	2.9%	2.4%	1.2%	1.2%
	-0.3%	5.2%	2.5%	2.0%	2.4%
	10.3%	10.6%	6.8%	7.8%	9.2%

Indicateur BIC						
	593715	583498	588876	592476	588416	579805
	597678	585133	592271	596217	591730	581034

Source : élaboration de l'auteur. Lecture : Idem page antérieure. Les coefficients en italique (cf. « Aidé » en sciences de la nature) ne sont pas significatifs (à p<0,1). Tous les autres coefficients, fixes ou aléatoires, le sont (à p<0,0000).

Variations à partir des Modèles n°14

Les Modèles N°14 ont été la base des différentes autres analyses par discipline illustrées au chapitre n°11, dont:

a) Les Modèles n°14a qui correspondent aux mêmes Modèles n°14 mais sous la technique MCO avec des écarts-types robustes (cf. annexes méthodologiques). Leurs résultats peuvent être appréciés sur le chapitre n°11 du Tome I.

b) Les Modèles n°14b correspondant à l'application du Modèle n°14, mais par type d'établissement. Cela équivaut alors à sept modèles différents pour chaque discipline et dont les détails sont communiqués ici en bas sur le tableau D44 (i.e. statistiques d'ajustement, variance totale et par niveaux, taille des échantillons) et sur le graphique D50 (variances totales et variances modélisés).¹³⁰

c) Pour ces mêmes Modèles n°14b, le tableau D45 s'attarde sur les résultats obtenus pour les variables de composition utilisées dans les analyses compte tenu l'importance que la littérature assigne à ce type de mesure et de leur probable concurrence ou complémentarité avec nos indicateurs d'effets de pairs. Le graphique D51 est issu, pour sa part, des modèles n°14 originaux. Il représente les coefficients des deux variables de composition assez importantes : moyenne et écart type de la scolarité des mères des élèves entourant un élève donné, mais obtenus sous la technique de régression par quantiles, donc, estimés pour différentes parties de la distribution des résultats académiques des élèves. Enfin, le tableau D46 illustre les résultats correspondant aux principales dimensions subjectives analysés (coefficients fixes).

d) Quant à eux, les modèles n°15f correspondent à l'incorporation des termes d'interaction du capital culturel (livres à la maison) et humain (scolarité des mères) des familles, avec le fait d'être un élève « aidé ». Le tableau D47 illustre les résultats, pour les quatre disciplines concernées, en termes de coefficients fixes, tout en ayant à la fois, une vision de comparaison avec les Modèles n°14 originales. Le tableau D48 montre les changements (presqu'inexistants) des coefficients aléatoires dus à l'inclusion de ces termes d'interaction dans la partie fixe.

e) Pour leur part, les Modèles n°15-fa se distinguent par considérer des termes d'interaction entre le fait d'aider ou d'être aidé, avec le genre et la condition de redoublement des élèves,

¹³⁰Notons que pour la variable « Attentes des élèves » la modalité « supérieur technique ou court » a été éliminée pour ne garder que son homologue « université ». Ceci compte tenu de la presque nulle prévalence de la première modalité chez certains types d'école.

mais aussi, parce qu'ils admettent que ces mêmes probables effets pourraient être autant fixes qu'aléatoires. Ainsi, le tableau D49 ne communique que les coefficients aléatoires issus de ces analyses, la partie fixe ayant été présentée au chapitre n°11 du Tome I.

f) Finalement, le tableau D50 et le graphique D52 nous permettent d'apprécier l'évolution de l'indicateur BIC pour les différents modèles communiqués au chapitre n°11, par discipline (i.e. entre le modèle Vide et les Modèles N°15-fa construits), de même que le nombre de paramètres (coefficients fixes et aléatoires) à chaque fois considérés.

Tableau D44

Indicateurs généraux sur les Modèles n°14b, appliqués par type d'établissement
(2004 - 8^{ème} grade ; échantillons d'élèves urbains dans des classes à taille ≥ 20)

Nombre d'unités par Niveau d'analyse (Modèle 14b)

Type d'établissement	Maths	Langage	Scs NAT	Scs SOC	N° Classes	N° Étabs
(1) M_NS	102664	103594	103921	102445	3205	1514
(2) MS	17753	17851	17901	17726	482	172
(3) PSUBV_NSFP	12922	13022	13069	12869	397	270
(4) PSUBV_NSFP	24547	24756	24832	24493	738	445
(5) PSUBV_SFP	7966	7997	8021	7954	219	116
(6) PSUBV_SFP	50620	50909	51022	50501	1416	739
(7) Privé PAYANT	16790	16965	16978	16783	604	309
Ensemble	233262	235094	235744	232771	7061	3565

Statistiques d'ajustement BIC par Modèle

Type d'Étab	MATHS		LANGAGE		Scs de la NATURE		Scs SOCIALES	
	Mod VIDE	Mod 14b	Mod VIDE	Mod 14b	Mod VIDE	Mod 14b	Mod VIDE	Mod 14b
1	278084	253467	282901	262005	281329	263418	278922	258547
2	42093	38024	44633	41784	43102	40358	43374	40359
3	34633	31549	35155	32654	34676	32474	34761	32161
4	65068	59131	66676	62131	66002	61988	65993	61326
5	20878	18756	21446	20049	21316	19952	21360	19698
6	133345	119932	137645	129016	135889	127134	135853	126572
7	44332	39551	46144	43469	45684	42752	45700	42723
Ensemble	579771	518326	605560	557928	593715	550348	597678	549227

Estimations générales de Variance prise en compte par les Modèles

Type d'Étab	MATHS		LANGAGE		Scs de la NATURE		Scs SOCIALES	
	Mod 14b	R ² (OLS)	Mod 14b	R ² (OLS)	Mod 14b	R ² (OLS)	Mod 14b	R ² (OLS)
1	0 25	0 26	0 25	0 23	0 20	0 21	0 22	0 23
2	0 36	0 45	0 39	0 36	0 30	0 39	0 32	0 39
3	0 31	0 31	0 31	0 27	0 25	0 26	0 28	0 28
4	0 30	0 30	0 30	0 25	0 24	0 24	0 26	0 26
5	0 35	0 35	0 36	0 27	0 27	0 27	0 30	0 29
6	0 30	0 31	0 31	0 23	0 24	0 24	0 24	0 24
7	0 35	0 35	0 35	0 23	0 25	0 24	0 23	0 23
Ensemble	0 40	0 41	0 34	0 34	0 34	0 34	0 34	0 34

Source : élaboration de l'auteur. Lecture : les modèles n°14b correspondent aux mêmes modèles principaux (n°14) appliqués dans l'ensemble, mais avec un paramètre en moins (i.e. élève qui croit arrivera dans le supérieur technique) et appliqués séparément pour chacune des catégories d'établissement. Les coefficients aléatoires ne sont pas détaillés. Le nombre d'unités correspond aux élèves, classes et écoles. En bas, on détaille la variance modélisée et la variance totale trouvée par type d'établissement. Notons que le contrôle par niveaux (i.e. estimations MN v/s estimations MCO ou OLS ; cf. annexes méthodologiques) s'avère assez importante, notamment pour la catégorie public sélectif.

Tableau D45

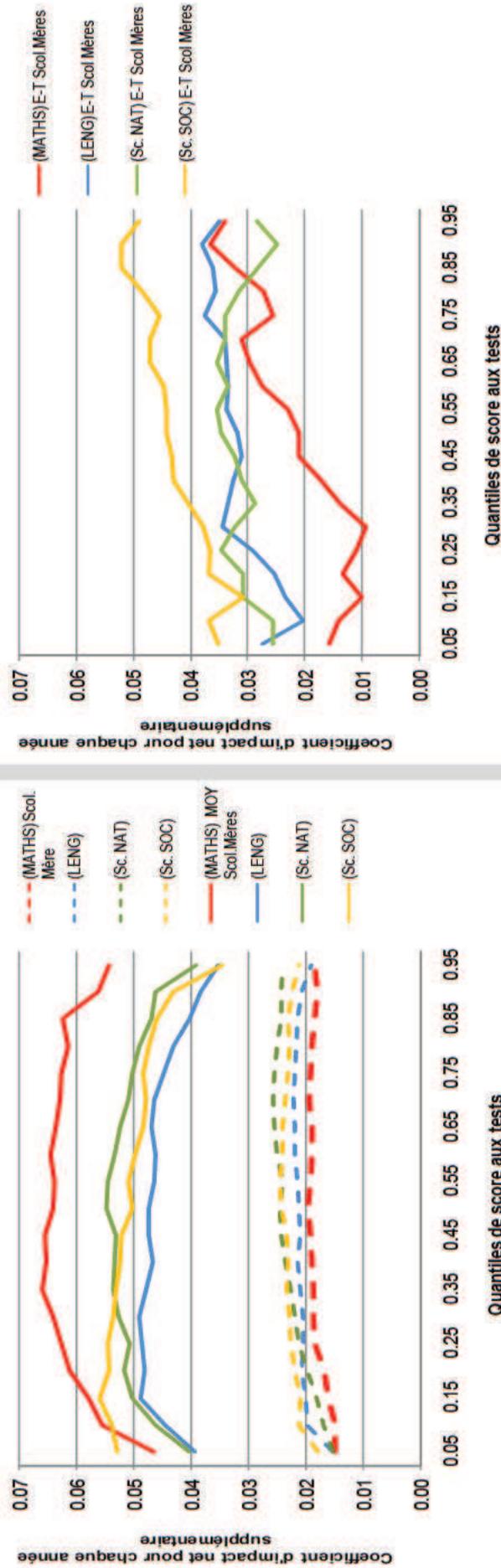
Coefficients d'impact moyen des variables de composition, issus des régressions multiniveaux par type d'établissement (modèles n°14b). Par discipline.

	Type d'école	Taille de la Classe		Moyenne d'années de scolarité des mères dans la classe (-i)		Écart-type scolarité des mères dans la classe (-i)		% d'élèves en Bas BEE dans la classe (-i)		% d'élèves en Haute CSS dans la classe (-i)	
		Code	Coeff.	Sign	Coeff.	Sign	Coeff.	Sign	Coeff.	Sign	Coeff.
Mathématiques	1	0.004	***	0.05	***	0.04	***	-0.005	***	-0.0001	ns
	2	-0.001	ns	0.08	***	-0.01	ns	-0.003	**	-0.0011	ns
	3	0.001	ns	0.05	***	0.01	ns	-0.005	***	0.0000	ns
	4	0.005	**	0.04	***	0.02	ns	-0.006	***	-0.0001	ns
	5	-0.016	***	0.06	***	0.06	ns	-0.004	**	-0.0036	*
	6	0.008	***	0.05	***	-0.02	ns	-0.002	***	0.0004	ns
	7	0.006	**	0.11	***	0.03	ns	-0.005	***	-0.0026	***
	<i>Ensemble</i>	0.005	***	0.06	***	0.01	**	-0.004	***	-0.001	*
Langage	1	0.003	**	0.03	***	0.04	***	-0.005	***	-0.0003	ns
	2	0.004	ns	0.06	***	-0.02	ns	-0.003	**	-0.0016	*
	3	0.001	ns	0.04	***	0.03	ns	-0.003	**	-0.0007	ns
	4	0.004	**	0.03	***	0.01	ns	-0.007	***	-0.0010	ns
	5	-0.012	***	0.04	**	0.04	ns	-0.005	***	-0.0020	ns
	6	0.008	***	0.03	***	-0.01	ns	-0.002	***	0.0005	ns
	7	-0.001	ns	0.06	***	0.02	ns	-0.004	***	-0.0018	ns
	<i>Ensemble</i>	0.004	***	0.039	***	0.017	***	-0.004	***	-0.001	**
Sciences de la Nature	1	0.002	*	0.04	***	0.04	***	-0.005	***	-0.0001	ns
	2	0.001	ns	0.06	***	-0.01	ns	-0.004	***	-0.0012	ns
	3	-0.001	ns	0.05	***	0.00	ns	-0.004	***	-0.0011	ns
	4	0.004	*	0.04	***	0.01	ns	-0.008	***	-0.0011	ns
	5	-0.015	***	0.02	ns	0.03	ns	-0.006	***	-0.0018	ns
	6	0.006	***	0.04	***	-0.03	*	-0.003	***	-0.0002	ns
	7	0.002	ns	0.08	***	0.08	***	-0.005	***	-0.0006	ns
	<i>Ensemble</i>	0.003	***	0.044	***	0.015	**	-0.005	***	-0.001	**
Sciences Sociales	1	0.003	**	0.04	***	0.04	***	-0.004	***	-0.0002	ns
	2	0.001	ns	0.08	***	0.00	ns	-0.003	**	-0.0016	*
	3	0.001	ns	0.05	***	0.09	***	-0.003	***	-0.0001	ns
	4	0.005	**	0.03	***	0.01	ns	-0.006	***	-0.0004	ns
	5	-0.007	ns	0.04	*	0.04	ns	-0.002	ns	-0.0007	ns
	6	0.005	***	0.04	***	-0.01	ns	-0.002	***	0.0000	ns
	7	0.003	ns	0.04	**	0.03	ns	-0.005	***	-0.0019	*
	<i>Ensemble</i>	0.003	***	0.044	***	0.022	***	-0.004	***	-0.001	*

Source : élaboration de l'auteur. Lecture : on illustre les impacts sur les acquis scolaires des différentes variables liées à la composition du public par classe. Ceci selon type d'école et dans l'ensemble. Signification statistique : *** = p<0.01 ; ** = p<0,05 à p≥0,01 ; * = p<0,1 à p≥0,05 ; ns= non significatif (p>0,1).

Graphique D51

Coefficients d'impact par quantiles de score des élèves (i.e. par niveau académique) des variables associées à la scolarité des mères des élèves. (8^{ème} grade – 2004)



Source : élaboration de l'auteur. Lecture : en abscisses se trouvent les différents quantiles analysés (entre 0,05 et 0,95) et en ordonnées, la taille des coefficients associés à la scolarité de la mère de l'élève (lignes à traits à gauche) ; puis, à la moyenne d'années de scolarité des mères par classe (-i) (lignes solides à gauche) ; et enfin, à l'écart-type de la scolarité des mères dans la classe (-i) (à droite). Les coefficients sont tous significatifs à $p < 0,05$, sauf en maths chez le quantile 0,15 pour la variable hétérogénéité ($p < 0,1$). Les spécifications de ces régressions par quantiles sont notées aux annexes méthodologiques. Les échantillons sont urbains et n'intègrent que des classes à taille ≥ 20 . Le nombre d'élèves est de 233262 (maths), 235094 (langage), 235744 (sciences de la nature), et de 232771 (sciences sociales).

Tableau D46

Coefficients fixes des conditions Basse et Moyenne pour le sentiment de Bien-être à l'école et de Confiance en soi scolaire Par discipline et selon type d'établissement. (Modèle n°14b). (2004 - 8^{ème} grade).

Pratique	Discipline		Type d'établissement							TOTAL
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	
BEE Bas	MATHS	C	-0.14	-0.05	-0.13	-0.09	-0.05	-0.03	0.06	-0.08
		SE	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.00
	LANG	C	-0.22	-0.13	-0.25	-0.18	-0.16	-0.12	-0.003 ns	-0.16
		SE	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.00
	Scs.NAT	C	-0.14	-0.05	-0.15	-0.08	-0.07	-0.03	0.08	-0.07
		SE	0.01	0.02	0.02	0.01	0.03	0.01	0.02	0.00
	Scs.SOC	C	-0.14	-0.04	-0.14	-0.11	-0.07	-0.01 ns	0.07	-0.08
		SE	0.01	0.02	0.02	0.01	0.03	0.01	0.02	0.00
BEE Moyen	MATHS	C	-0.06	-0.01 ns	-0.04 **	-0.02 ns	0.02 ns	0.01 ns	0.05	-0.02
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.00
	LANG	C	-0.10	-0.05	-0.12	-0.08	-0.06 **	-0.03	0.02 ns	-0.06
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.00
	Scs.NAT	C	-0.05	0.00 ns	-0.05	-0.03 **	0.00 ns	0.01 *	0.05	-0.02
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.00
	Scs.SOC	C	-0.07	-0.02 ns	-0.06	-0.06	0.02 ns	0.01 ns	0.04	-0.03
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.00
CSS Basse	MATHS	C	-0.19	-0.16	-0.18	-0.17	-0.16	-0.18	-0.25	-0.16
		SE	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.00
	LANG	C	-0.26	-0.21	-0.24	-0.23	-0.23	-0.26	-0.31	-0.23
		SE	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.00
	Scs.NAT	C	-0.24	-0.20	-0.23	-0.23	-0.24	-0.24	-0.34	-0.22
		SE	0.01	0.02	0.02	0.01	0.03	0.01	0.02	0.00
	Scs.SOC	C	-0.24	-0.21	-0.26	-0.21	-0.25	-0.23	-0.31	-0.22
		SE	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.00
CSS Moyenne	MATHS	C	-0.09	-0.08	-0.09	-0.08	-0.07	-0.09	-0.12	-0.08
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.00
	LANG	C	-0.13	-0.10	-0.11	-0.12	-0.10	-0.12	-0.13	-0.11
		SE	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.00
	Scs.NAT	C	-0.12	-0.10	-0.10	-0.12	-0.10	-0.12	-0.17	-0.11
		SE	0.01	0.02	0.02	0.01	0.03	0.01	0.02	0.00
	Scs.SOC	C	-0.12	-0.09	-0.13	-0.11	-0.11	-0.11	-0.16	-0.11
		SE	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.00

Source : élaboration de l'auteur. Lecture : Idem tableau antérieur. En sciences sociales et dans les établissements privés subventionnés non-sélectifs et sans FP, les élèves classés en condition de CSS Moyenne obtiendraient un score moyen de $-0,13\sigma$ par rapport à ceux qui se classent en condition de Haute confiance en soi scolaire (colonne n°3 en bas). Cette différence serait fortement significative du point de vue statistique. Pour chaque discipline, on communique le coefficient (C) et l'erreur standard qui va avec (SE). Les résultats pour l'ensemble ou total (i.e. Modèles n°14 des tableaux n°11 à n°18) sont reproduits comme référence.

Tableau D47

Coefficients fixes des termes d'interaction entre le fait d'être aidé et quatre variables de choix. Par discipline. (Modèles n°15f)(2004 - 8^{ème} grade)

Variable	MATHÉMATIQUES				LANGAGE			
	Modèle N°14		Modèle N°15		Modèle N°14		Modèle N°15	
	Coeff	StdE	Coeff	StdE	Coeff	StdE	Coeff	StdE
AIDANT	0.39	0 004	0.39	0 004	0.19	0 005	0.19	0 005
AIDÉ	-0.05	0 004	0.09	0 018	0.00	0 004 ns	0.04	0 020 **
Et. SEUL	0.14	0 004	0.14	0 004	0.17	0 004	0.16	0 004
Et. LIVRE	0.15	0 005	0.15	0 005	0.23	0 005	0.23	0 005
MOY Années Scol. Mères (-i)	0.06	0 004	0.06	0 004	0.04	0 003	0.04	0 003
Années Scol. Mère	0.02	0 001	0.02	0 001	0.02	0 001	0.02	0 001
Capital Culturel BAS	-0.20	0 005	-0.22	0 006	-0.20	0 005	-0.23	0 006
Capital Culturel MOYEN	-0.11	0 004	-0.12	0 005	-0.12	0 004	-0.12	0 005
NSE Classe(-i) * Aidé			-0.01	0 002			0.002	0 002 ns
Scolarité Mère * Aidé			-0.01	0 001			-0.01	0 001
CapCult BAS * Aidé			0.052	0 009			0.075	0 010
CapCult MOY * Aidé			0.026	0 007			0.023	0 009

Variable	Sciences de la NATURE				Sciences SOCIALES			
	Modèle N°14		Modèle N°15		Modèle N°14		Modèle N°15	
	Coeff	StdE	Coeff	StdE	Coeff	StdE	Coeff	StdE
AIDANT	0.24	0 005	0.24	0 005	0.25	0 005	0.25	0 005
AIDÉ	0.02	0 004	0.11	0 021	-0.02	0 004	0.05	0 021 **
Et. SEUL	0.14	0 004	0.14	0 004	0.15	0 004	0.15	0 004
Et. LIVRE	0.25	0 004	0.25	0 004	0.25	0 004	0.25	0 004
MOY Années Scol. Mères (-i)	0.04	0 003	0.05	0 004	0.04	0 003	0.05	0 003
Années Scol. Mère	0.02	0 001	0.02	0 001	0.02	0 001	0.02	0 001
Capital Culturel BAS	-0.21	0 005	-0.23	0 006	-0.22	0 005	-0.24	0 006
Capital Culturel MOYEN	-0.13	0 004	-0.13	0 005	-0.13	0 004	-0.13	0 005
NSE Classe(-i) * Aidé			-0.008	0 002			-0.004	0 002 *
Scolarité Mère * Aidé			-0.005	0 001			-0.006	0 001
CapCult BAS * Aidé			0.051	0 010			0.073	0 011
CapCult MOY * Aidé			0.029	0 009			0.029	0 009

Source : élaboration de l'auteur. Lecture : le fait d'être aidé en maths montre une interaction significative et négative avec la scolarité des mères (modèles n°15), équivalente à $-0,01\sigma$ pour chaque année supplémentaire d'études de la mère d'un élève ayant fait cette déclaration. Chez les coefficients des modèles n°14 (cf. chapitre n°11) on peut comparer le changement des variables originales dus à l'inclusion de cette interaction. Tous les coefficients significatifs à $p < 0,001$ sauf ceux marqués autrement : ** = $p < 0,05$ à $p \geq 0,001$; ns= non significatif ($p > 0,1$).

Tableau D48

Détails des parties aléatoires pour les modèles incorporant termes d'interaction du capital culturel et humain (Modèles n°15f) (2004 – 8^{ème} grade ; Échantillons d'élèves urbains dans des classes à taille ≥20)

	MATHS			LANGAGE			Scs. de la NATURE			Scs. SOCIALES		
	Mod 14	Mod 15	Mod 14	Mod 15	Mod 14	Mod 15	Mod 14	Mod 15	Mod 14	Mod 15	Mod 14	Mod 15
Coefficients Aléatoires												
NIVEAU 1	0.51325	0.51256	0.60331	0.60285	0.57571	0.57533	0.59399	0.59353	0.59399	0.59353	0.59399	0.59353
NIVEAU 2	0.02942	0.02934	0.02303	0.02297	0.03334	0.03335	0.02112	0.02109	0.02112	0.02109	0.02112	0.02109
NIVEAU 3	0.05245	0.05253	0.03514	0.03507	0.04665	0.04665	0.04518	0.04516	0.04518	0.04516	0.04518	0.04516
Entre Élèves	0.14278	0.14348	0.11793	0.11839	0.10319	0.10358	0.12277	0.12323	0.12277	0.12323	0.12277	0.12323
Entre Classes	0.02088	0.02096	0.02596	0.02602	0.01965	0.01965	0.02094	0.02097	0.02094	0.02097	0.02094	0.02097
Entre Étabs.	0.23846	0.23838	0.19642	0.19649	0.21945	0.21945	0.19387	0.19388	0.19387	0.19388	0.19387	0.19388
% variance modélisée par Niveau												
Entre Élèves	0.21685	0.21791	0.16388	0.16452	0.15161	0.15217	0.17082	0.17146	0.17082	0.17146	0.17082	0.17146
Entre Classes	0.41406	0.41562	0.53054	0.53182	0.37015	0.37008	0.49698	0.49783	0.49698	0.49783	0.49698	0.49783
Entre Étabs.	0.81902	0.81874	0.84858	0.84889	0.82424	0.82425	0.81051	0.81056	0.81051	0.81056	0.81051	0.81056
Variance TOT Modélisée (MMN)	0.4021	0.4028	0.3403	0.3409	0.3423	0.3427	0.3376	0.3381	0.3376	0.3381	0.3376	0.3381
R ² réf (OLS)	0.4054	0.4060	0.3392	0.3397	0.3442	0.3446	0.3401	0.3406	0.3401	0.3406	0.3401	0.3406

Source : élaboration de l'auteur. Lecture : les modèles N°15f correspondent aux mêmes modèles principaux (n°14) mais avec des termes d'interaction pour la variable « *Je suis aidé par mes camarades* » avec le capital culturel des élèves (modalités bas et moyen) et avec la scolarité des mères de ceux-ci. À l'aide de ces coefficients et de leurs points de comparaison on peut apprécier que l'incorporation de ces termes d'interaction (pris tous au niveau individu) n'améliore que très faiblement la qualité explicative des modèles. Consignons que les modèles N°15f ici commentés ne seront pas intégrés après dans la suite comparative. D'ailleurs, les variables « capital culturel » (modalité bas) et le nombre d'années d'étude de la mère, ne voient pas leur impact modifié significativement par classe. Des analyses approfondies sur ce sujet ne seront pas commentées ici.

Tableau D49
Coefficients aléatoires des modèles N°15-fa.
Par discipline. (2004 – 8^{ème} grade)

Coefficient Aléatoire	Niveau / Variable	MATHS			LANGAGE		
		Estimate	StdErr	ProbZ	Estimate	StdErr	ProbZ
Constante	Établissement	0.052	0.002	***	0.035	0.002	***
Constante	Classe	0.027	0.001	***	0.021	0.001	***
Pente	Aidant	0.019	0.002	***	0.011	0.002	***
Pente	Aidé	0	0	ns	0	0	ns
Pente	Fille	0.004	0.001	***	0.003	0.001	**
Pente	Redoublant	0.010	0.002	***	0.010	0.002	***
Pente	<i>Fille*Aidant</i>	0.001	0.003	ns	0	0	ns
Pente	<i>Redoub*Aidant</i>	0.009	0.007	†	0.001	0.006	ns
Pente	<i>Fille*Aidé</i>	0	0	ns	0	0	ns
Pente	<i>Redoub*Aidé</i>	0	0	ns	0	0	ns
	Résiduel	0.506	0.002	***	0.597	0.002	***

Coefficient Aléatoire	Niveau / Variable	Scs. de la NATURE			Scs. SOCIALES		
		Estimate	StdErr	ProbZ	Estimate	StdErr	ProbZ
Constante	Établissement	0.046	0.002	***	0.045	0.002	***
Constante	Classe	0.032	0.001	***	0.020	0.001	***
Pente	Aidant	0.017	0.002	***	0.019	0.002	***
Pente	Aidé	0	0	ns	0	0	ns
Pente	Fille	0.002	0.001	*	0.001	0.001	ns
Pente	Redoublant	0.007	0.002	**	0.003	0.002	†
Pente	<i>Fille*Aidant</i>	0	0	ns	0	0	ns
Pente	<i>Redoub*Aidant</i>	0	0	ns	0	0	ns
Pente	<i>Fille*Aidé</i>	0	0	ns	0	0	ns
Pente	<i>Redoub*Aidé</i>	0	0	ns	0	0	ns
	Résiduel	0.571	0.002	***	0.590	0.002	***

Source : élaboration de l'auteur. Lecture : l'impact du fait d'être un élève déclarant aider les autres, pour toutes les disciplines, varie significativement entre classes (du moins, pour certaines d'entre elles). Signification statistique (*ProbZ*) : *** $p < 0,0000$; ** = entre $p \geq 0,001$ et $p < 0,05$; * = entre $p \geq 0,05$ et $p < 0,1$; † = entre $p \geq 0,1$ et $p < 0,15$; ns = non significatif (à $p < 0,15$). NOTE : Les Modèles N°15-fa sont identiques aux Modèles N°14 mais ils ajoutent des termes d'interaction sur des variables spécifiques (i.e. genre*(aidé/aidant), puis, redoublant*(aidé/aidant)) tout en admettant que les pentes de ces interactions de niveau élève puissent être aléatoires au niveau classe. De plus, ces modèles admettent que les variables originales (genre, redoublement, aidé, aidant) peuvent aussi varier leurs impacts selon les classes. Statistiques d'ajustement et comparaison avec d'autres modèles sont disponibles en annexes. Les modèles estimés contraignent les covariances à zéro. Toutes les pentes sont proposées comme étant de niveau classe.

Tableau D50

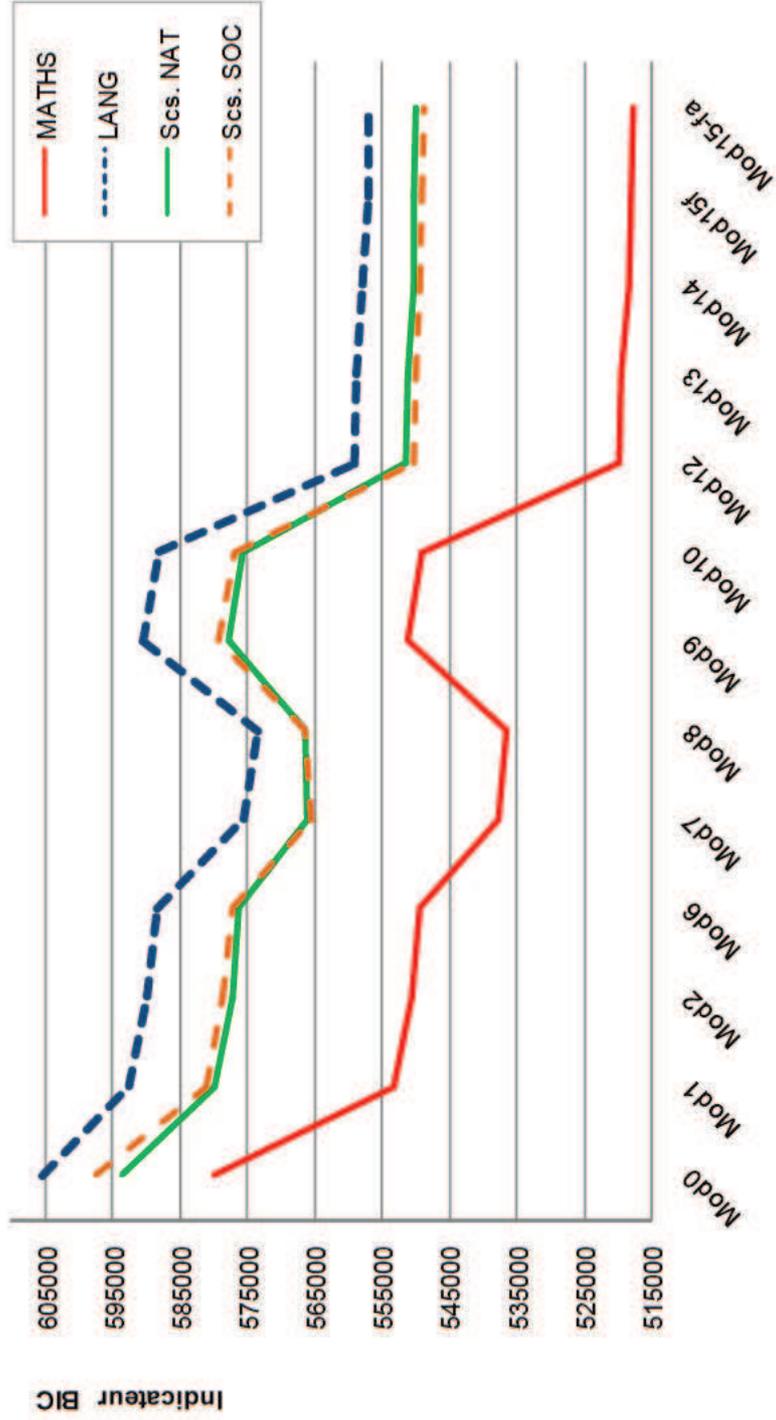
Différences entre les Indicateurs BIC selon variables incluses dans les Modèles communiqués.
Par Discipline. (8^{ème} grade – 2004 échantillons d'élèves urbains dans des classes à taille ≥ 20)

MODEL	Variables incluses	Deltas Indicateur BIC (réf= Model antérieur)					N° Param Fixes (*)
		MATHS	LANG	N° Param Fixes (*)	Scs. NAT	Scs. SOC	
Model Vide	Aucune			0			0
Model 1	Cibles	26553	12990	8	13910	16644	8
Model 2	Caractéristiques de l'Établissement	2705	2615	16	2617	2464	16
Model 6	Composition Classe	1130	1553	13	1006	1336	13
Model 7	Caractéristiques Élève (Dims. Subjs.)	11834	12748	16	10178	11630	15
Model 8	Caractéristiques Élève (Socio-Démog.)	1012	2227	17	-485	-794	17
Model 9	Enseignant (Caract./Croy./Prats.)	-14699	-17134	25/26	-11450	-12809	25/26
Model 10	Enseign. (Caract./Croy./Prats.) + Comp. Classe	2311	2469	30/31	2116	2220	30/31
Model 12	Model 7+8+2	29131	28880	33	24381	26664	33
Model 13	Model 7+8+2+IRM	367	453	34	268	342	34
Model 14	Todas las Variables	1102	831	56	826	753	56
Model 15f	Model 14 + Interactions Fixes	461	815	60	198	251	60
Model 15fa	Model 14 + Interactions Fixes et Aléatoires (**)	222	56	60/61	88	94	59/60
		BIC			BIC		
Model Vide		579771	605560	0	593715	597678	0
Model 15fa	Model 14 + Interactions Fixes et Aléatoires	517643	557057	60/61	550062	548882	59/60

Source : élaboration de l'auteur. Pour chaque modèle, la référence est la valeur du BIC du modèle antérieur. Mais stricto sensu, seulement les modèles n°14 à n°15fa sont comparables car ils considèrent une même base commune des variables incluses sur laquelle on teste des effets d'interaction, soient-ils, pensés exclusivement comme étant des effets fixes (n°15f), soient-ils, en les acceptant comme des possibles effets aléatoires (pentes) (n°15fa). Plusieurs modèles ne considèrent pas des variables communes entre eux (cf. variables de niveau individuel, en gris). Un delta positif indique un gain explicatif. En bas sont notés les deux valeurs totales extrêmes (cf. Modèle Vide et Modèle n°15fa). L'addition de chaque colonne des valeurs BIC conduit à la valeur du Modèle N°15fa. (*) Ils sont notés seulement les paramètres fixes inclus dans chaque modèle (hors constante). Les différences entre disciplines proviennent de fait qu'en sciences sociales et en langage il y a quatre catégories des pratiques de maîtres tandis que dans les autres disciplines, il n'y en a que trois. (**) Les effets aléatoires considèrent, en plus des constantes et des interactions (cf. genre*aidé/aidant et redoublant*aidé/aidant), les variables genre, redoublement, aidant et aidé, déclarées comme des possibles effets aléatoires de niveau classe. Remarque : ces interactions sont différentes de celles considérées dans les Modèles N°15i, commentées plus haut.

Graphique D52

Evolution générale de l'Indicateur BIC selon Modèles d'analyse communiqués.
 Par discipline. (8^{ème} grade - 2004 - échantillons d'élèves urbains dans des classes à taille ≥ 20)



Source : élaboration de l'auteur. Lecture : pour les variables incluses dans chaque modèle voir tableau antérieur.

ANNEXE « E »

MODELISATION DES PRINCIPALES DIMENSIONS SUBJECTIVES

MODELISATION DES DIMENSIONS SUBJECTIVES

Dans le cas des dimensions de Bien-être à l'école et de Confiance en soi scolaire nous avons suivi une démarche semblable à celle menée pour les modèles portant sur les acquis scolaires. Il faudrait noter que nous avons remplacé certaines variables catégorielles par des indices continus préalablement construits et justifiés dans d'autres parties de ce travail. De même, nous avons testé la probable existence de certaines relations quadratiques. Dès lors, l'usage de la standardisation, tant sur la variable dépendante que sur plusieurs facteurs explicatifs, ne nous permettra pas seulement d'assumer la plupart des impacts en termes de variations d'écart-type, mais cet usage va contribuer à diminuer les phénomènes de colinéarité au sein des modèles. En effet, l'Indice d'état ne dépasse pas dans ce cas la valeur 55 (Modèles N°7) et il n'existe aucun duo de variables qui soient en même temps associées à la respective composante à plus de 50%. Néanmoins, comme nous le communiquons dans le tableau D51 suivant, l'indicateur VIF s'accroît pour certaines variables, il le fait sans dépasser toutefois la valeur critique de 10. Cette hausse est due sans doute à l'inclusion des mesures inévitablement reliées entre elles. C'est le cas de notre indice de compétences (addition des scores aux tests) exprimé par classes (-i) avec l'Indice de NSE par classe aussi ; ou encore, celui des attentes des parents que l'on sait fortement corrélées avec celles de leurs enfants.

Pour sa part, le tableau E52 note les variances par niveau pour les différentes variables utilisées dans les analyses. Ces estimations sont issues de la Proc NESTED de SAS®. Finalement, le tableau E53 communique les résultats des modèles n°7, autant pour la dimension de CSS que pour celle de BEE. Le graphique E53 illustre, pour les trois niveaux concernés, les résidus issus de ces mêmes modèles (n°7). Le tableau E56 fait état des corrélations entre les scores prédits par les modèles et les scores observés. Le tableau E57, quant à lui, compare les statistiques d'ajustement et les principaux coefficients fixes et aléatoires entre les modèles n°7, d'un côté, et une variation de ceux-ci (n°7a) qui accepte des termes d'interaction au niveau de paramètres fixes (entre NSE et type d'établissement) ainsi que des probables pentes aléatoires pour ces mêmes variables, d'autre côté. Notons que ces modèles n°7a servent de base au graphique n°37 du chapitre n°13 au Tome I.

Enfin, les tableaux E54 et E55 montrent l'ensemble des résultats, en termes de coefficients fixes, des suites des modèles construites pour expliquer ces deux dimensions (modèles n°1 à n°7). Les tableaux E58 et E59, quant à eux, comparent les modèles n°7 sur deux populations distinctes : l'échantillon utilisée et celui qui élimine les données où les variables à la base des dimensions subjectives présentent des valeurs manquantes.

Tableau E51

VARIABLE	STATISTIQUES GÉNÉRALES					Variance par Niveau (en %)			VIF (Mod.7)
	Moyenne	Écart-Type	Q3	Mediane	Q1	Étab.	Classe	Élève	
AIDANT (général)	0.24	0.43	0	0	0	0.8	0.3	98.9	1.5
AIDANT - AIDÉ (général)	0.15	0.36	0	0	0	1.0	0.9	98.1	1.3
AIDÉ (général)	0.31	0.46	1	0	0	0.5	0.3	99.2	1.4
Position dans la classe (Basse)	0.34	0.47	1	0	0	-	-	100	2.7
Position dans la classe (Haute)	0.32	0.47	1	0	0	-	-	100	2.7
Niveau de Compétences	0.00	1.00	0.7	0.0	-0.8	31.2	5.4	63.4	7.6
Niv. Compét. (carré)	1.00	1.19	1.5	0.5	0.1	10.5	3.2	86.3	1.3
Niveau Moyen de Compétences (Classe-i)	0.00	1.00	0.6	-0.1	-0.7	80.9	18.9	0.2	6.2
Niv. Moyen Compét. (carré)	1.00	1.35	1.3	0.5	0.1	70.8	28.9	0.3	1.9
Hétérogénéité Acad. (STD Niv. Compét. -i)	0.00	1.00	0.6	0.0	-0.7	23.7	74.3	2.0	1.3
Croit ira UNIV (élève)	0.60	0.49	1	1	0	19.2	1.6	79.1	2.7
Croit ira SupTECH (élève)	0.18	0.38	0	0	0	4.7	0.8	94.5	1.9
Bien-être à l'école	0.00	1.00	0.8	0.1	-0.6	2.8	2.0	95.2	1.4
BEE (carré)	1.00	1.50	1.3	0.5	0.1	0.3	0.3	99.4	1.2
Confiance en Soi Scolaire	0.00	1.00	0.8	0.2	-0.6	2.8	1.4	95.9	2.1
CSS (carré)	1.00	1.61	1.3	0.5	0.2	0.7	0.5	98.8	1.6
Rapport Moyen aux Profs	0.00	1.00	0.8	0.2	-0.6	12.2	9.3	78.5	1.7
Rapport Moyen aux Profs (carré)	1.00	1.42	1.3	0.5	0.1	4.1	6.9	89.0	1.4
Appui des Parents	0.00	1.00	0.8	0.1	-0.7	2.3	1.6	96.1	1.6
Appui (carré)	1.00	1.24	1.5	0.6	0.1	0.6	0.6	98.9	1.3
Aime Étudier (Niv2 - général)	0.49	0.50	1	0	0	1.4	0.8	97.8	2.2
Aime Étudier (Niv3 - général)	0.34	0.47	1	0	0	7.0	1.6	91.4	2.7
Étudie 'des fois' (dans la semaine) (général)	0.39	0.49	1	0	0	1.6	1.3	97.1	1.2
Étudie 'tous les jours ou presque' (général)	0.07	0.26	0	0	0	1.3	1.1	97.5	1.2
Fille	0.50	0.50	1	1	0	10.5	1.9	87.6	1.1
Redoublant	0.15	0.36	0	0	0	5.8	3.6	90.6	1.2
Indice de NSE (élève)	0.00	1.00	0.6	-0.15	-0.7	63.4	0.9	35.7	3.6
NSE (carré)	1.00	1.36	1.3	0.5	0.1	49.4	0.6	50.1	1.9
Moins de 11 livres à la maison	0.28	0.45	1	0	0	16.3	1.3	82.4	2.0
Entre 11 et 50 Livres à la maison	0.40	0.49	1	0	0	2.7	0.8	96.4	1.6
Foyer NON Couple	0.29	0.45	1	0	0	2.1	0.5	97.4	1.1
Au moins a 1 Frère	0.79	0.41	1	1	1	0.9	1.0	98.0	1.1
Attentes Parents (UNIV)	0.49	0.50	1	0	0	28.3	1.4	70.3	3.7
Attentes Parents (SupTECH)	0.30	0.46	1	0	0	8.5	0.5	91.0	2.3
Préscolaire	0.45	0.50	1	0	0	17.0	0.6	82.3	1.2

Tableau E52

(continuation)	STATISTIQUES GÉNÉRALES					Variance par Niveau (en %)		VIF (Mod.7)
	Moyenne	Écart-Type	Q3	Médiane	Q1	Étab.	Classe	
VARIABLE								VIF
Moyenne Années d'Expérience (4 Profs)	19.94	8.38	26.5	20.0	13.5	77.4	22.6	1.9
Un Enseignant au moins a un autre travail	0.41	0.49	1	0	0	62.1	37.9	1.1
25% d'Enseignants a une spécialisation	0.35	0.48	1	0	0	39.9	60.1	1.5
>=50% d'Enseignants a une spécialisation	0.38	0.49	1	0	0	58.8	41.2	1.6
Niv. Moyen Couvertures des Programmes (éch)	0.00	1.00	0.7	0.1	-0.6	71.8	28.2	1.3
Trav Coopérat. Moyen (général)	0.41	0.49	1	0	0	70.3	29.7	2.0
Trav Coopérat. Haut (général)	0.41	0.49	1	0	0	79.8	20.2	2.3
Attentes Profs : SupTECH	0.21	0.41	0	0	0	46.7	53.3	1.6
Attentes Profs : UNIV	0.18	0.38	0	0	0	77.1	22.9	3.0
Niveau Moyen de Préparation déclaré (échelle)	0.00	1.00	0.7	0.1	-0.6	77.5	22.5	1.4
Profs Préoccupés DISCIPLINE (Niv. Moyen)	0.32	0.47	1	0	0	55.6	44.4	1.3
Profs Préoccupés DISCIPLINE (Niv. Haut)	0.31	0.46	1	0	0	29.5	70.5	1.3
Profs Préoccupés RÉSULTATS (Niv. Moyen)	0.29	0.46	1	0	0	28.2	71.8	1.4
Profs Préoccupés RÉSULTATS (Niv. Haut)	0.41	0.49	1	0	0	57.0	43.0	1.5
Moyenne de l'Indice NSE (Classe -i)	0.00	1.00	0.5	-0.2	-0.8	97.1	2.9	8.0
STD Indice NSE (-i)	0.00	1.00	0.6	0.0	-0.6	36.4	62.0	1.2
Taille de la Classe	35.27	6.45	41.0	36.0	31.0	80.1	19.9	1.7
% Classe (-i) Bas BEE	31.88	11.54	39.5	31.3	23.8	29.4	70.0	1.4
% Classe (-i) Haute CSS	33.69	11.41	41.0	33.3	25.7	33.8	65.5	1.2
N° élèves/Prof	21.87	6.16	25.6	21.6	18.0	100	-	1.6
Étab a Labo Scs.	0.51	0.50	1	1	0	100	-	1.5
MUNIS	0.08	0.27	0	0	0	100	-	1.3
PSUBV NS NFP	0.06	0.23	0	0	0	100	-	1.2
PSUBV NS FP	0.11	0.31	0	0	0	100	-	1.6
PSUBV S NFP	0.03	0.18	0	0	0	100	-	1.2
PSUBV S FP	0.22	0.41	0	0	0	100	-	2.7
Privé PAYANT	0.07	0.26	0	0	0	100	-	3.9
Inverse du Ratio de MILLS	-0.20	0.84	0.5	-0.3	-1.0	90.0	0.4	6.8

Tableau E53

		Confiance en Soi Scolaire			Bien-être à l'école		
Variable (Modèle N°7)		Coeff.	StdE	Sign.	Coeff.	StdE	Sign.
Profil Scolaire	AIDANT (général)	0.174	0 005	***	0.101	0 005	***
	AIDANT - AIDÉ (général)	0.026	0 006	***	0.182	0 006	***
	AIDÉ (général)	-0.070	0 005	***	0.111	0 005	***
	Position dans la classe (Basse)	-0.005	0 006	ns	-0.022	0 007	***
	Position dans la classe (Haute)	0.030	0 006	***	-0.018	0 007	**
	Niveau de Compétences Estimé	0.171	0 005	***	0.084	0 005	***
	Niv. Compét. (carré)	0.031	0 002	***	-0.056	0 002	***
Classe (PUBLIC)	Niveau Moyen de Compétences (Classe -i)	-0.082	0 005	***	0.008	0 005	*
	Niv. Moyen Compét. (carré)	-0.018	0 002	***	0.023	0 002	***
	Hétérogénéité Acad. (STD Niv. Compét. -i)	-0.010	0 002	***	0.009	0 002	***
	Moyenne de l'Indice NSE (Classe -i)	-0.008	0 006	ns	0.039	0 006	***
	STD Indice NSE (-i)	-0.002	0 002	ns	-0.004	0 002	**
	Taille de la Classe	-0.001	0 000	**	0.000	0 000	ns
	% Classe (-i) Bas BEE	0.001	0 000	***	-0.010	0 000	***
	% Classe (-i) Haute CSS	0.005	0 000	***	-0.001	0 000	***
Dimensions SUBJECTIVES	Croit ira SupTECH (élève)	0.136	0 006	***	0.015	0 007	**
	Croit ira UNIV (élève)	0.218	0 006	***	0.035	0 006	***
	Bien-être à l'école	0.127	0 002	***			
	BEE (carré)	0.014	0 001	***			
	Confiance en Soi Scolaire				0.152	0 003	***
	CSS (carré)				0.006	0 001	***
	Rapport Moyen aux Profs (M&L)	0.210	0 002	***	0.114	0 003	***
	Rapport Moyen aux Profs (carré)	0.027	0 001	***	0.016	0 002	***
	Appui des Parents	0.185	0 002	***	0.061	0 002	***
	Appui (carré)	0.042	0 002	***	-0.023	0 002	***
	Aime Étudier (Niv.Moyen - général)	0.291	0 005	***	0.281	0 006	***
	Aime Étudier (Niv. Haut - général)	0.513	0 006	***	0.440	0 007	***
	Étudie 'des fois' (dans la semaine) (général)	0.150	0 004	***	-0.043	0 004	***
	Étudie 'tous les jours ou presque' (général)	0.322	0 007	***	-0.165	0 008	***

		(continuation)	Confiance en Soi Scolaire			Bien-être à l'école		
		Variable (Modèle N°7)	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.
Socio-Scolaires		Fille	-0.008	0 004	**	-0.004	0 004	ns
		Redoublant	-0.078	0 005	***	-0.054	0 006	***
		Indice de NSE de l'élève	-0.021	0 003	***	0.008	0 004	**
		Moins de 11 livres à la maison	0.014	0 005	**	0.033	0 006	***
		Entre 11 et 50 Livres à la maison	0.001	0 004	ns	0.028	0 005	***
		Foyer NON Couple	-0.011	0 004	**	-0.023	0 004	***
		Au moins a 1 Frère	0.003	0 004	ns	0.020	0 005	***
		Attentes Parents (SupTECH)	0.021	0 006	***	-0.008	0 006	ns
		Attentes Parents (UNIV)	0.080	0 007	***	0.013	0 007	*
Classe (ENSEIGNANT)		Moyenne Années d'Expérience (4 Profs)	0.001	0 000	***	0.003	0 000	***
		Un Enseignant au moins a un autre travail	0.013	0 004	**	-0.001	0 004	ns
		25% d'Enseignants a une spécialisation	-0.016	0 005	**	-0.007	0 005	ns
		>=50% d'Enseignants a une spécialisation	-0.017	0 005	**	-0.004	0 005	ns
		Niv. Moyen Couvertures des Programmes (éch.)	0.004	0 002	*	-0.003	0 002	ns
		Trav Coopérat. Moyen (général)	-0.006	0 006	ns	-0.008	0 006	ns
		Trav Coopérat. Haut (général)	-0.014	0 006	**	-0.010	0 006	*
		Attentes Profs : SupTECH	-0.025	0 006	***	-0.006	0 006	ns
		Attentes Profs : UNIV	-0.035	0 009	***	-0.015	0 009	*
		Niveau Moyen de Préparation déclaré (échelle)	-0.005	0 002	**	0.002	0 002	ns
		Profs Préoccupés DISCIPLINE (Niv. Moyen)	-0.001	0 005	ns	0.011	0 005	**
		Profs Préoccupés DISCIPLINE (Niv. Haut)	0.002	0 005	ns	0.004	0 005	ns
		Profs Préoccupés RÉSULTATS (Niv. Moyen)	0.006	0 005	ns	0.003	0 005	ns
	Profs Préoccupés RÉSULTATS (Niv. Haut)	0.001	0 005	ns	0.006	0 005	ns	
Établissement		N°élèves/Prof	0.001	0 000	ns	-0.001	0 000	**
		Étab a Labo Scs.	-0.032	0 005	***	-0.012	0 005	**
		MUNI S	0.004	0 010	ns	-0.015	0 009	*
		PSUBV NS NFP	-0.042	0 010	***	0.001	0 009	ns
		PSUBV NS FP	-0.002	0 009	ns	-0.010	0 008	ns
		PSUBV S NFP	-0.023	0 013	*	-0.001	0 012	ns
		PSUBV S FP	-0.016	0 009	*	-0.002	0 008	ns
	Privé PAYANT	0.017	0 016	ns	0.063	0 015	***	

Source : élaboration de l'auteur. Lecture : sur l'échelle de confiance en soi scolaire, le fait qu'un élève déclare penser qu'il arrivera à l'université apporte, en moyenne, un gain de +0,22 σ par rapport à ceux qui pensent qu'ils n'arriveront pas au supérieur. Cette même caractéristique n'apporte qu'un gain moyen de +0,035 σ sur l'échelle de bien-être.

Tableau E54

	MODEL 0		MODEL 1		MODEL 2		MODEL 3		MODEL 4		MODEL 5		MODEL 6		MODEL 7																																		
	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE	Coeff.	StdE																																	
	Sign.		Sign.		Sign.		Sign.		Sign.		Sign.		Sign.		Sign.																																		
Constante	-0.01	0.00	0.0125	-0.16	0.01	0.0000	-0.10	0.01	0.0000	0.15	0.02	0.0000	-0.15	0.01	0.0000	0.10	0.00	0.0000	0.00	0.00	0.0000	0.09	0.00	0.0000	0.08	0.01	0.0000	0.08	0.01	0.0000																			
AIDANT (général)	0.26	0.01	0.0000	0.26	0.01	0.0000	0.25	0.01	0.0000	0.25	0.01	0.0000	0.25	0.01	0.0000	0.10	0.01	0.0000	0.10	0.01	0.0000	-0.03	0.02	0.1668	0.10	0.01	0.0000	0.10	0.01	0.0000	0.10	0.01	0.0000																
AIDANT - AIDÉ (général)	0.29	0.01	0.0000	0.29	0.01	0.0000	0.28	0.01	0.0000	0.28	0.01	0.0000	0.29	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000													
AIDÉ (général)	0.13	0.01	0.0000	0.13	0.01	0.0000	0.14	0.01	0.0000	0.14	0.01	0.0000	0.14	0.01	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000	0.11	0.00	0.0000													
Position dans la classe (Basse)																																																	
Position dans la classe (Haute)																																																	
Niveau de Compétences																																																	
Niv Compét (carré)																																																	
Niveau Moyen de Compétences (Classe -i)																																																	
Niv Moyen Compét (carré)																																																	
Hétérogénéité Acad (STD Niv Compét -i)																																																	
Croit ira SupTECH (élève)																																																	
Croit ira UNIV (élève)																																																	
Confiance en Soi Scolaire																																																	
CSS (carré)																																																	
Rapport Moyen aux Profs (M&L)																																																	
Rapport Moyen aux Profs (carré)																																																	
Appui des Parents																																																	
Appui (carré)																																																	
Aime Étudier (Niv Moyen - général)																																																	
Aime Étudier (Niv Haut - général)																																																	
Étude 'des fois' (dans la semaine) (général)																																																	
Étude 'tous les jours ou presque' (général)																																																	
Fille																																																	
Redoublant																																																	
Indice de NSE de l'élève																																																	
NSE (carré)																																																	
Moins de 11 livres à la maison																																																	
Entre 11 et 50 Livres à la maison																																																	
Foyer NON Couple																																																	
Au moins a 1 Frère																																																	
Attentes Parents (SupTECH)																																																	
Attentes Parents (UNIV)																																																	
Préscolaire																																																	

(continuation)

BEE	MODEL 0	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6	MODEL 7				
	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.
Moyenne Années d'Expérience (4 Profes)				0 00	0 00	0 0000	0 00	0 00	0 0000	0 00	0 00	0 0000
Un Enseignant au moins a un autre travail				-0 01	0 00	0 0589	0 00	0 00	0 8231	0 00	0 00	0 8132
25% d'Enseignants a une spécialisation				-0 01	0 01	0 0391	-0 01	0 01	0 1811	-0 01	0 01	0 1825
>=50% d'Enseignants a une spécialisation				0 00	0 01	0 4872	0 00	0 01	0 4537	0 00	0 01	0 4554
Niv Moyen Couvertures des Programmes (éch)				0 01	0 00	0 0085	0 00	0 00	0 1628	0 00	0 00	0 1657
Trav Coopérat Moyen (général)				0 00	0 01	0 7046	-0 01	0 01	0 1724	-0 01	0 01	0 1729
Trav Coopérat Haut (général)				0 00	0 01	0 5031	-0 01	0 01	0 0952	-0 01	0 01	0 0946
Attentes Profes SupTECH				-0 01	0 01	0 2441	-0 01	0 01	0 3272	-0 01	0 01	0 3262
Attentes Profes UNIV				-0 02	0 01	0 0899	-0 02	0 01	0 0878	-0 02	0 01	0 0853
Niveau Moyen de Préparation déclaré (échelle)				0 00	0 00	0 7888	0 00	0 00	0 4332	0 00	0 00	0 4335
Profes Préoccupés DISCIPLINE (Niv Moyen)				0 02	0 00	0 0022	0 01	0 00	0 0229	0 01	0 00	0 0227
Profes Préoccupés DISCIPLINE (Niv Haut)				0 01	0 00	0 2440	0 00	0 00	0 4241	0 00	0 00	0 4262
Profes Préoccupés RÉSULTATS (Niv Moyen)				0 00	0 01	0 7155	0 00	0 01	0 5839	0 00	0 01	0 5818
Profes Préoccupés RÉSULTATS (Niv Haut)				0 01	0 00	0 2049	0 01	0 00	0 2495	0 01	0 00	0 2473
Moyenne de l'Indice NSE (Classe -i)				-0 01	0 01	0 0169	0 04	0 01	0 0000	0 04	0 01	0 0000
STD Indice NSE (-i)				0 00	0 00	0 5712	0 00	0 00	0 0459	0 00	0 00	0 0470
Taille de la Classe				0 00	0 00	0 1870	0 00	0 00	0 4979	0 00	0 00	0 5019
% Classe (-i) Bas BEE				-0 01	0 00	0 0000	-0 01	0 00	0 0000	-0 01	0 00	0 0000
% Classe (-i) Haute CSS				0 00	0 00	0 0000	0 00	0 00	0 0000	0 00	0 00	0 0000
N°élèves/Prof				0 00	0 00	0 0433	0 00	0 00	0 0079	0 00	0 00	0 0076
Étab a Labo Scs				-0 02	0 00	0 0000	-0 01	0 00	0 0133	-0 01	0 00	0 0135
MUNIS				-0 02	0 01	0 0355	-0 01	0 01	0 0974	-0 01	0 01	0 0943
PSUBV NS NFP				0 00	0 01	0 7471	0 00	0 01	0 8919	0 00	0 01	0 8916
PSUBV NS FP				-0 02	0 01	0 0081	-0 01	0 01	0 2070	-0 01	0 01	0 2055
PSUBV S NFP				-0 01	0 01	0 4998	0 00	0 01	0 9285	0 00	0 01	0 9274
PSUBV S FP				-0 02	0 01	0 0281	0 00	0 01	0 8211	0 00	0 01	0 8188
Privé PAYANT				0 03	0 01	0 0671	0 06	0 02	0 0000	0 06	0 02	0 0000
Inverse du Ratio de MILLS				-0 08	0 01	0 0000	-0 05	0 01	0 0000	-0 05	0 01	0 0000

Tableau E55

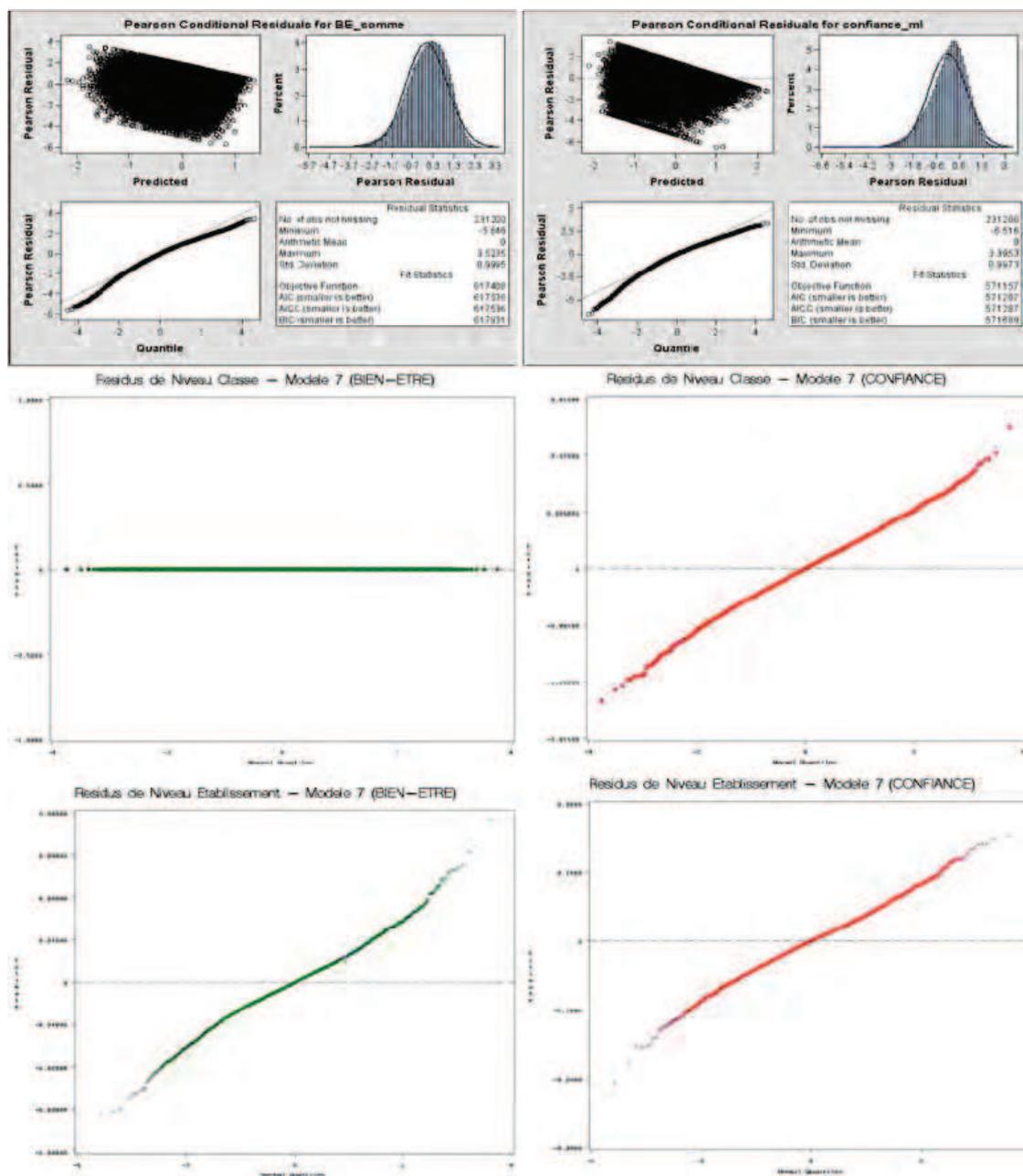
CSS	MODEL 0	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6	MODEL 7	
	Coef.	StdE	Sign.	Coef.	StdE	Sign.	Coef.	StdE	Sign.
Constante	0.00	0.00	0.2936	-0.10	0.01	0.00000	-0.70	0.01	0.00000
AIDANT (général)	0.49	0.01	0.00000	0.43	0.01	0.00000	0.17	0.00	0.00000
AIDANT - AIDÉ (général)	0.21	0.01	0.00000	0.21	0.01	0.00000	0.02	0.01	0.00000
AIDÉ (général)	-0.04	0.01	0.00000	-0.01	0.01	0.05388	-0.07	0.00	0.00000
Position dans la classe (Basse)									
Position dans la classe (Haute)									
Niveau de Compétences									
Niv Compét (carré)	0.18	0.00	0.00000	0.17	0.00	0.00000	0.20	0.00	0.00000
Niveau Moyen de Compétences (Classe -i)	0.01	0.00	0.00000	0.02	0.00	0.00000	0.03	0.00	0.00000
Niv Moyen Compét (carré)	-0.09	0.00	0.00000	-0.08	0.00	0.00000	-0.10	0.00	0.00000
Hétérogénéité Acad (SID Niv Compét -i)	-0.05	0.00	0.00000	-0.02	0.00	0.00000	-0.03	0.00	0.00000
Croit ira SupTECH (élève)	-0.02	0.00	0.00000	-0.01	0.00	0.00006	-0.01	0.00	0.00000
Croit ira UNIV (élève)									
Bien-être à l'école									
BEE (carré)									
Rapport Moyen aux Profs (M&L)									
Rapport Moyen aux Profs (carré)									
Appui des Parents									
Appui (carré)									
Aime Étudier (Niv Moyen - général)									
Aime Étudier (Niv Haut - général)									
Étudie 'des fois' (dans la semaine) (général)									
Étudie 'tous les jours ou presque' (général)									
Fille									
Redoublant									
Indice de NSE de l'élève									
NSE (carré)									
Moins de 11 livres à la maison									
Entre 11 et 50 Livres à la maison									
Foyer NON Couple									
Au moins a 1 Frère									
Attentes Parents (SupTECH)									
Attentes Parents (UNIV)									
Préscolaire									

(continuation)

CSS	MODEL 0	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6	MODEL 7				
	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.	Coeff.	StdE	Sign.
Moyenne Années d'Expérience (4 Profés)				0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Un Enseignant au moins a un autre travail				0.00	0.00	0.6888	0.01	0.00	0.0040	0.01	0.00	0.0039
25% d'Enseignants a une spécialisation				-0.02	0.01	0.0040	-0.02	0.01	0.0023	-0.02	0.01	0.0023
>=50% d'Enseignants a une spécialisation				-0.01	0.01	0.0665	-0.02	0.01	0.0012	-0.02	0.01	0.0012
Niv Moyen Couvertures des Programmes (éch)				0.02	0.00	0.0000	0.00	0.00	0.0707	0.00	0.00	0.0711
Trav Coopérat Moyen (général)				0.00	0.01	0.5992	-0.01	0.01	0.3062	-0.01	0.01	0.3056
Trav Coopérat Haut (général)				0.00	0.01	0.5692	-0.01	0.01	0.0302	-0.01	0.01	0.0304
Attentes Profés SupTECH				-0.02	0.01	0.0004	-0.02	0.01	0.0001	-0.02	0.01	0.0001
Attentes Profés UNIV				-0.03	0.01	0.0002	-0.03	0.01	0.0002	-0.03	0.01	0.0002
Niveau Moyen de Préparation déclaré (échelle)				-0.01	0.00	0.0062	-0.01	0.00	0.0417	-0.01	0.00	0.0417
Profés Préoccupés DISCIPLINE (Niv Moyen)				0.01	0.01	0.2590	0.00	0.01	0.8542	0.00	0.01	0.8523
Profés Préoccupés DISCIPLINE (Niv Haut)				0.00	0.00	0.3553	0.00	0.00	0.7601	0.00	0.00	0.7591
Profés Préoccupés RÉSULTATS (Niv Moyen)				0.00	0.01	0.9231	0.01	0.01	0.2497	0.01	0.01	0.2500
Profés Préoccupés RÉSULTATS (Niv Haut)				0.00	0.01	0.7851	0.00	0.01	0.8729	0.00	0.01	0.8753
Moyenne de l'Indice NSE (Classe -i)				-0.07	0.01	0.0000	-0.01	0.01	0.1541	-0.01	0.01	0.1536
STD Indice NSE (-i)				0.00	0.00	0.3195	0.00	0.00	0.2692	0.00	0.00	0.2633
Taille de la Classe				0.00	0.00	0.0000	0.00	0.00	0.0012	0.00	0.00	0.0011
% Classe (-i) Bas BEE				0.00	0.00	0.0000	0.00	0.00	0.0001	0.00	0.00	0.0001
% Classe (-i) Haute CSS				0.01	0.00	0.0000	0.01	0.00	0.0000	0.01	0.00	0.0000
N°élèves/Prof				0.00	0.00	0.1881	0.00	0.00	0.1543	0.00	0.00	0.1520
Étab a Labo Scs				-0.05	0.01	0.0000	-0.03	0.01	0.0000	-0.03	0.01	0.0000
MUNI S				0.00	0.01	0.8721	0.00	0.01	0.6868	0.00	0.01	0.6812
PSUBV NS NFP				-0.04	0.01	0.0001	-0.04	0.01	0.0000	-0.04	0.01	0.0000
PSUBV NS FP				-0.02	0.01	0.0082	0.00	0.01	0.8516	0.00	0.01	0.8540
PSUBV S NFP				-0.04	0.01	0.0019	-0.02	0.01	0.0784	-0.02	0.01	0.0784
PSUBV S FP				-0.04	0.01	0.0000	-0.02	0.01	0.0651	-0.02	0.01	0.0653
Privé PAYANT				-0.06	0.02	0.0000	0.02	0.02	0.3080	0.02	0.02	0.3091
Inverse du Ratio de MILLS				-0.11	0.01	0.0000	-0.05	0.01	0.0000	-0.05	0.01	0.0000

Graphique E53

Résidus par Niveau (élève, classe et établissement) des Modèles MN portant sur les dimensions de Bien-être (à gauche) et de Confiance en soi scolaire (à droite) (Modèles N°7) (8^{ème} grade - 2004 ; 23 1200 élèves dans des classes urbaines ≥ 20)



Source : sorties SAS® pour des modèles estimés visant à expliquer les dimensions de CSS et de BEE. Note : les résidus de niveau établissement pour la dimension de CSS occupent une fourchette très réduite et ils ne sont pas significatifs. Il est de même avec les résidus de niveau établissement pour la dimension de BEE.

Tableau E56

Relation entre scores réels et prédits
(Model N°7)

<i>Dimension</i>	<i>r</i>	<i>R²</i>
Bien-être à l'école	0.393	0.15
Confiance en soi scolaire	0.562	0.32

Tableau E57

Indicateurs d'ajustement des modèles faisant appel à des interactions entre l'indice de NSE par élève et le type d'établissement. Par dimension subjective.
(2004 - 8^{ème} grade ; 231200 élèves urbains issus de classes ≥20)

		<i>Modèles Tableaux n°33 et n°34</i>	<i>Modèles avec Termes d'interaction (NSE). Graph. n°37</i>	
		Model VIDE	Model N°7	Model N°7a
Bien-être à l'école	-2 Log L	651092.5	617407.6	617435.4
	BIC	651125.3	617931.0	618089.7
	DELTA (BIC)	-	33194.2	-158.7
	Delta DDL = 12			
Confiance en soi scolaire	-2 Log L	651865.8	571157.0	571191.4
	BIC	651898.5	571688.6	571870.2
	DELTA (BIC)	-	80209.9	-181.6
	Delta DDL = 12			

Lecture : Ce tableau présente deux statistiques d'ajustement pour différents modèles (Vide, modèle n°7, puis, modèles n°7 avec interactions et pentes aléatoires). Toutes les interactions ont été déclarées comme des possibles effets aléatoires (pentes). Cela signifie douze termes additionnels sur le modèle n°7a par rapport au n°7 (i.e. l'Indice de NSE, cet Indice au carré, puis, les six catégories d'école). Une structure de covariance restreinte a été demandée (option « VC » sur SAS®), cherchant à connaître seulement les termes de variance. Donc, on postule comme nulles les covariances entre pentes et entre pentes et constantes. Les Deltas de l'indicateur BIC sont estimés à chaque fois sur la valeur BIC du modèle précédant. Les coefficients aléatoires estimés ne montrent pas de signification statistique (à $p < 0,05$) pour le cas de la CSS, tandis que seulement l'interaction entre la catégorie privé payant et l'indice de NSE se montre significative pour le cas de BEE.

Tableau E58

Comparaison entre les modèles estimés avec l'ensemble des données (i.e. après remplissage) et les modèles estimés sur des données restreints (i.e. en excluant ceux dont manquait de l'information de base pour les dimensions de BEE et de CSS). (2004 – 8^{ème} grade ; échantillons d'élèves urbains dans des classes à taille ≥ 20)

Confiance en soi scolaire	<i>Model N°7</i> <i>Sans Remplissage (n=199467)</i>				<i>Model N°7</i> <i>Avec Remplissage (n=231200)</i>			
	<i>Coeff.</i>	<i>StdErr</i>	<i>DDL</i>	<i>Prob. T</i>	<i>Coeff.</i>	<i>StdErr</i>	<i>DDL</i>	<i>Prob. T</i>
Variables partie FIXE								
AIDANT (général)	0.180	0.005	195855	0.0000	0.174	0.005	227582	0.0000
AIDANT - AIDÉ (général)	0.023	0.006	195855	0.0001	0.026	0.006	227582	0.0000
AIDÉ (général)	-0.071	0.005	195855	0.0000	-0.070	0.005	227582	0.0000
Position dans la classe (Basse)	-0.004	0.007	195855	0.5405	-0.005	0.006	227582	0.3892
Position dans la classe (Haute)	0.029	0.006	195855	0.0000	0.030	0.006	227582	0.0000
Niveau de Compétences	0.172	0.005	195855	0.0000	0.171	0.005	227582	0.0000
Niv. Compét. (carré)	0.032	0.002	195855	0.0000	0.031	0.002	227582	0.0000
Niveau Moyen de Compétences (Classe -i)	-0.085	0.005	195855	0.0000	-0.082	0.005	227582	0.0000
Niv. Moyen Compét. (carré)	-0.018	0.002	195855	0.0000	-0.018	0.002	227582	0.0000
Bien-être à l'école (échelle continue)	0.127	0.002	195855	0.0000	0.127	0.002	227582	0.0000
BEE (carré)	0.015	0.002	195855	0.0000	0.014	0.001	227582	0.0000
Coefficients partie ALÉATOIRE								
	<i>Coeff.</i>	<i>StdErr</i>	<i>Z Value</i>	<i>Prob Z</i>	<i>Coeff.</i>	<i>StdErr</i>	<i>Z Value</i>	<i>Prob Z</i>
Niveau Établissement	0.005	0.000	11.9	0.000	0.006	0.000	11.9	0.000
Niveau Classe	0.000				0.000	0.001	0.8	0.207
Niveau Élève	0.679	0.002	313.0	0.000	0.688	0.002	333.3	0.000
	<i>Model VIDE</i>		<i>Model N°7</i>		<i>Model VIDE</i>		<i>Model N°7</i>	
-2 Log V	561270.3		490066.4		651865.8		571157.0	
BIC	561303.0		490589.8		651898.5		571688.6	

Source : élaboration de l'auteur. Ces modèles ne communiquent que les variables cibles, mais, ils comportent d'autres contrôles communiqués dans les tableaux précédents.

Tableau E59

Comparaison entre les modèles estimés avec l'ensemble des données (i.e. après remplissage) et les modèles estimés sur des données restreints (i.e. en excluant ceux dont manquait de l'information de base pour les dimensions de BEE et de CSS). (2004 – 8^{ème} grade ; échantillons d'élèves urbains dans des classes à taille ≥ 20)

Bien-être à l'école	<i>Model N°7</i> <i>Sans Remplissage (n=199467)</i>				<i>Model N°7</i> <i>Avec Remplissage (n=231200)</i>			
	<i>Coeff.</i>	<i>StdErr</i>	<i>DDL</i>	<i>Prob. T</i>	<i>Coeff.</i>	<i>StdErr</i>	<i>DDL</i>	<i>Prob. T</i>
<i>Variables partie FIXE</i>								
AIDANT (général)	0.108	0.006	195855	0.0000	0.101	0.005	227582	0.0000
AIDANT - AIDÉ (général)	0.184	0.006	195855	0.0000	0.182	0.006	227582	0.0000
AIDÉ (général)	0.112	0.005	195855	0.0000	0.111	0.005	227582	0.0000
Position dans la classe (Basse)	-0.029	0.007	195855	0.0000	-0.022	0.007	227582	0.0007
Position dans la classe (Haute)	-0.013	0.007	195855	0.0813	-0.018	0.007	227582	0.0065
Niveau de Compétences	0.074	0.006	195855	0.0000	0.084	0.005	227582	0.0000
Niv. Compét. (carré)	-0.058	0.002	195855	0.0000	-0.056	0.002	227582	0.0000
Niveau Moyen de Compétences (Classe -i)	0.010	0.005	195855	0.0599	0.008	0.005	227582	0.0981
Niv. Moyen Compét. (carré)	0.023	0.002	195855	0.0000	0.023	0.002	227582	0.0000
Confiance en Soi Scolaire	0.149	0.00	195855	0.0000	0.152	0.00	227582	0.0000
CSS (carré)	0.003	0.00	195855	0.0752	0.006	0.00	227582	0.0001
<i>Coefficients partie ALÉATOIRE</i>								
	<i>Coeff.</i>	<i>StdErr</i>	<i>Z Value</i>	<i>Prob Z</i>	<i>Coeff.</i>	<i>StdErr</i>	<i>Z Value</i>	<i>Prob Z</i>
Niveau Établissement	0.0007	0.0004	1.9	0.0284	0.0009	0.0003	2.7	0.0040
Niveau Classe	0.0000				0.0000			
Niveau Élève	0.8346	0.0027	313.3	0.0000	0.8450	0.0025	337.6	0.0000
		<i>Model VIDE</i>	<i>Model N°7</i>		<i>Model VIDE</i>	<i>Model N°7</i>		
-2 Log V		560445.5	530151.2		651092.5	617407.6		
BIC		560478.2	530674.5		651125.3	617931.0		

Source : élaboration de l'auteur. Lecture : Idem tableau antérieur. En gris sont marquées les différences portant sur au moins un aspect (sens des coefficients, signification, etc.) entre les modèles estimés selon l'échantillon pris.

ANNEXE « F »

MODELISATION DES PRATIQUES D'ENTRAIDE

Vision préliminaire

Pour donner une première vision sur l'ensemble, le graphique F54 illustre les courbes de distribution générale de la proportion par classe déclarant chacune des deux pratiques étudiées. Nous tenons à mettre en exergue le caractère plutôt normal de celles-ci, à l'exception du fait de déclarer aider les autres en langage qui s'étend davantage vers la droite avec un écart-type de 11% contre 9% pour les autres trois disciplines. Nous voyons, en outre, la présence de ces comportements selon la discipline, retrouvant ainsi le fait qu'en maths les deux pratiques d'entraide sont beaucoup plus courantes qu'ailleurs. Ce cas est suivi par langage, puis, par les deux autres matières.

Cependant, rappelons que si les variations dans la présence des pratiques s'expriment en tant que proportions à un niveau donné (cf. graphique E54), cela ne veut pas dire que ce niveau et ses caractéristiques les expliquent entièrement, ou même partiellement. Il se peut, en effet, que ce soit tout simplement l'agrégation des caractéristiques individuelles reliées aux pratiques celle qui explique les différentes proportions apprécées à simple vue. Conséquemment, chaque modèle communiqué dans les tableaux n°23 à n°30 (chapitre n°12, Tome I) considère les coefficients aléatoires respectifs des niveaux classe et école. Ces coefficients permettent, d'ailleurs, d'estimer la proportion de variation par niveaux des probabilités de déclarer chaque pratique.

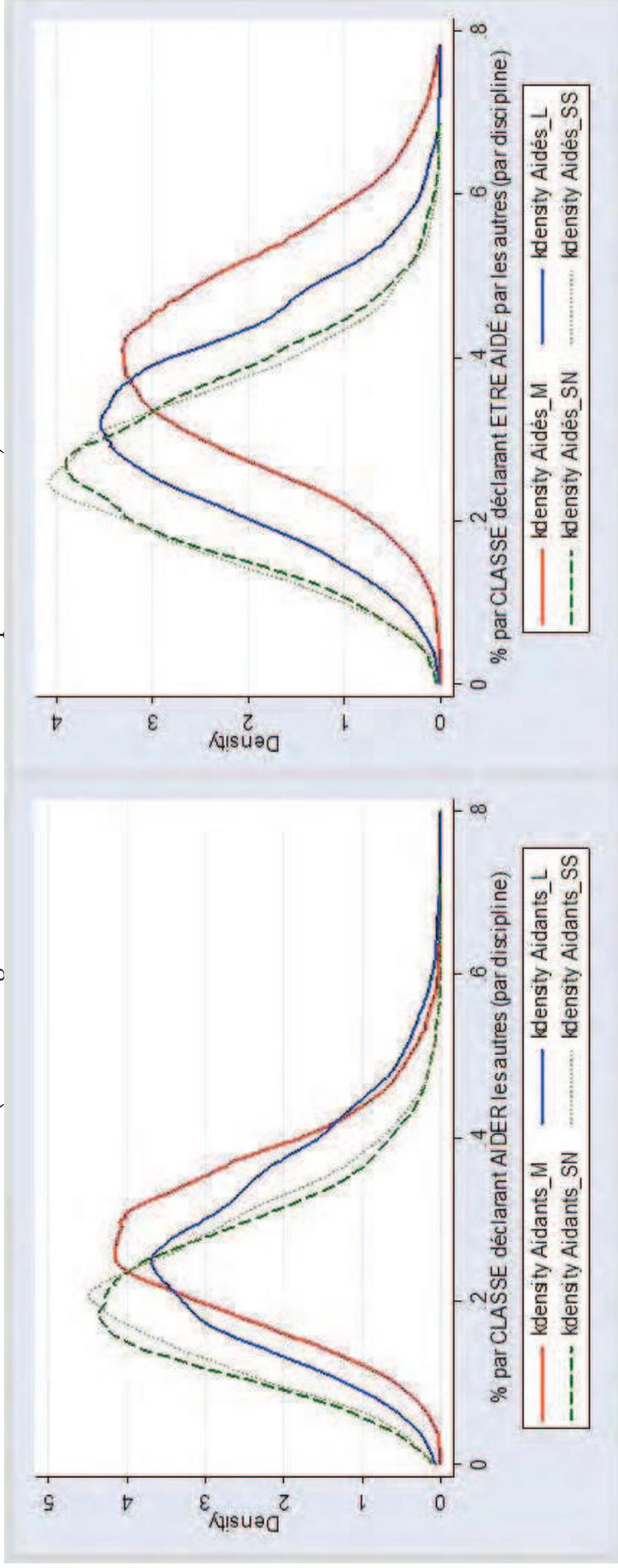
Enfin, notons que, par souci de cohérence, une correction spécifique a été appliquée. Sur l'ensemble de classes on a éliminé, par discipline, celles dont il y avait des élèves aidés mais pas d'élèves aidants et vice-versa. Ceci sachant qu'il est difficile de qualifier ces inconsistances car, comme on l'a dit, certains élèves peuvent être aidés par plusieurs autres et encore, ils peuvent être aidés par d'autres camarades du même établissement mais pas de la même classe. Au total, cela aboutit à l'élimination de 4 classes en maths, 18 en langage, 28 et 37 en sciences sociales et de la nature respectivement, ce qui signifie pour chaque estimation une perte d'effectifs allant de 0,04% à 0,4% (entre 90 et 1120 élèves). Toutefois, et malgré ces corrections, les analyses restent presque identiques quel que soit l'échantillon pris.

Les analyses menées

Les suites de modèles, par discipline, illustrées sur les tableaux F60 à F63 correspondent à la modélisation de la pratique « aider les autres », tandis que les suites montrées par les tableaux F63 à F67 s'attardent sur la pratique d'être aidé. Enfin, le tableau F68 communique une série des corrélations de Pearson, par discipline, entre les proportions par classe de ces deux comportements avec d'autres caractéristiques de contexte.

Graphique F54

Densité des distributions de la proportion par classe déclarant *aider* ou *être aidés*. Selon discipline.
(2004 – 8^{ème} grade – 7061 classes urbaines à plus de 19 élèves)



Source : élaboration de l'auteur. Lecture : les courbes respectives correspondent aux densités de Kernel (méthode Epanechnikov) pour les proportions d'élèves par classe déclarant *être aidés* par les autres (droite) ou *aider* les autres (gauche). M = maths ; L = langage ; SN = sciences de la nature ; et SS = sciences sociales.

Tableau F60 : Aidant - Maths

	MODEL A	MODEL B	MODEL C	MODEL D	MODEL E	MODEL F	MODEL G	MODEL H	MODEL I	MODEL J		
	Logit	Staff	Prob1	Logit	Staff	Prob1	Logit	Staff	Prob1	Logit	Staff	Prob1
Constante	-0.95	0.01	0.0000	-1.22	0.02	0.0000	-1.19	0.04	0.0000	-0.82	0.07	0.0000
Années Scolarité Mère	0.01	0.00	0.0000	0.02	0.00	0.0000	0.03	0.00	0.0000	-0.82	0.07	0.0000
Entre 11 et 50 Livres	0.15	0.01	0.0000	0.16	0.01	0.0000	0.17	0.01	0.0000	0.10	0.01	0.0000
Entre 51 et 100 Livres	0.22	0.02	0.0000	0.25	0.02	0.0000	0.26	0.02	0.0000	0.19	0.02	0.0000
Entre 101 et 200 Livres	0.36	0.02	0.0000	0.39	0.02	0.0000	0.41	0.02	0.0000	0.28	0.02	0.0000
Plus de 200 Livres	0.36	0.02	0.0000	0.40	0.02	0.0000	0.41	0.02	0.0000	0.28	0.02	0.0000
Indice de la classe en M (Moyenne)												
Position dans la classe en M (Haute)												
Score en Maths (SIMCE 2004)												
Score en Maths (carré)												
Ecart-Type Score M (Classe -)												
AIDÉ (M)												
Et SEUL (M)												
Et LIVRE (M)												
Confiance en So4 Scolaire												
CS (carré)												
Appui des Parents												
Bien-être à l'école												
BEE (carré)												
Aime Maths (Niv2)												
Aime Maths (Niv3)												
Aime Maths (Niv4)												
Étude M = 2 = 'des fois'; 'Hebdo'												
Va obligé à l'école (d'accord)												
Va obligé à l'école (très d'accord)												
... avoir un bon travail (d'accord)												
... avoir un bon travail (très d'accord)												
Croît tra SupTECH												
Croît tra UNIV												
Fille												
Redoublant												
Préscolaire												
Foyer NON Couple												
Au moins a 1 Frere												
Taille de la Classe												
% Années Scol. Mères (-)												
% Années Scol. Mères (4)												
% Classe (4) Bas REE												
% Classe (4) Haute CS												
Années d'Expérience (PM)												
Postitre (PM)												
Post-grade (PM)												
Autre Travail (PM)												
Niv Prepa Moyen (PM)												
Niv Prepa Haut (PM)												
Attentes: SupTECH (PM)												
Attentes: UNIV (PM)												
Préoccupé DISCIPLINE (PM)												
Préoccupé RESULTATS (PM)												
Convert. Prog. MOYENNE (PM)												
Convert. Prog. HAUTE (PM)												
% Heures Prepa Cours (PM)												
Trax Cooperat. Moyenne (PM)												
Trax Cooperat. Haute (PM)												
Plusieurs PRATIS (PM)												
PRATIS Indiv. (PM)												
Étab a Labo Scs.												
N'élèves/Prof												
MUNI S												
PSUBV NS NFP												
PSUBV NS FP												
PSUBV S NFP												
PSUBV S FP												
PRIV PAVANT												

Tableau F66 : Aidé - Sciences de la Nature

	MODEL A	MODEL B	MODEL C	MODEL D	MODEL E	MODEL F	MODEL G	MODEL H	MODEL I	MODEL J		
	Logr	Staf	Pwch	Logr	Staf	Pwch	Logr	Staf	Pwch	Logr	Staf	Pwch
Constante	-0.97	0.01	0.0000	-0.97	0.01	0.0000	-0.97	0.01	0.0000	-0.97	0.01	0.0000
Année de Scolarité Mère	-0.01	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000
Entre 11 et 30 Livres	0.00	0.01	0.0000	-0.02	0.01	0.0000	-0.02	0.01	0.0000	-0.01	0.01	0.0000
Entre 31 et 100 Livres	0.06	0.01	0.0000	-0.02	0.01	0.0000	-0.03	0.01	0.0000	-0.06	0.02	0.0000
Entre 101 et 200 Livres	0.06	0.01	0.0000	-0.02	0.01	0.0000	-0.03	0.01	0.0000	-0.06	0.02	0.0000
Entre 201 et 300 Livres	0.15	0.02	0.0000	-0.12	0.02	0.0000	-0.09	0.02	0.0000	-0.15	0.03	0.0000
Entre 301 et 400 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 401 et 500 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 501 et 600 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 601 et 700 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 701 et 800 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 801 et 900 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Entre 901 et 1000 Livres	0.15	0.02	0.0000	-0.13	0.02	0.0000	-0.10	0.02	0.0000	-0.11	0.03	0.0000
Position dans la classe en SN (Moyenne)												
Position dans la classe en SN (Ecart)												
Score en Sc. NAT (SJMCE 2004)												
Score en Sc. NAT (carte)												
Ecart-Type-Scote M (Classe-i)												
MIDANT (SN)												
EL SEUL (SN)												
Confiance en Soi Scolaire												
CSS (carte)	-0.16	0.01	0.0000	-0.15	0.01	0.0000	-0.14	0.01	0.0000	-0.13	0.01	0.0000
Appui des Parents	0.15	0.01	0.0000	-0.07	0.00	0.0000	-0.07	0.00	0.0000	-0.07	0.00	0.0000
Appui (carte)	0.16	0.01	0.0000	-0.07	0.00	0.0000	-0.07	0.00	0.0000	-0.07	0.00	0.0000
Bien-être à l'école	0.02	0.00	0.0000	0.02	0.00	0.0000	0.02	0.00	0.0000	0.02	0.00	0.0000
BEE (carte)	0.16	0.01	0.0000	0.02	0.00	0.0000	0.02	0.00	0.0000	0.02	0.00	0.0000
Aime Sc. Nat (Niv2)	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000
Aime Sc. Nat (Niv3)	-0.22	0.02	0.0000	-0.22	0.02	0.0000	-0.22	0.02	0.0000	-0.22	0.02	0.0000
Importe Sc. Nat (Niv3)	0.12	0.01	0.0000	0.12	0.01	0.0000	0.12	0.01	0.0000	0.12	0.01	0.0000
Étudie SCN > "des fois Hebdo"	0.02	0.01	0.0000	0.02	0.01	0.0000	0.02	0.01	0.0000	0.02	0.01	0.0000
Va obligé à l'école (d'accord)	-0.02	0.01	0.0000	-0.02	0.01	0.0000	-0.02	0.01	0.0000	-0.02	0.01	0.0000
Va obligé à l'école (très d'accord)	0.01	0.01	0.0000	0.01	0.01	0.0000	0.01	0.01	0.0000	0.01	0.01	0.0000
... avoir un bon travail (très d'accord)	0.19	0.02	0.0000	0.19	0.02	0.0000	0.19	0.02	0.0000	0.19	0.02	0.0000
Credit tra SupTECH	0.15	0.02	0.0000	0.15	0.02	0.0000	0.15	0.02	0.0000	0.15	0.02	0.0000
Credit tra UNIV	-0.12	0.02	0.0000	-0.12	0.02	0.0000	-0.12	0.02	0.0000	-0.12	0.02	0.0000
Étude	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000
Reculant	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000	0.18	0.01	0.0000
Préscolaire	0.03	0.01	0.0000	0.03	0.01	0.0000	0.03	0.01	0.0000	0.03	0.01	0.0000
Foyer NON Couple	0.02	0.01	0.0000	0.02	0.01	0.0000	0.02	0.01	0.0000	0.02	0.01	0.0000
Aliments à l'ère	0.04	0.01	0.0000	0.04	0.01	0.0000	0.04	0.01	0.0000	0.04	0.01	0.0000
Taille de la Classe	0.09	0.01	0.0000	0.09	0.01	0.0000	0.09	0.01	0.0000	0.09	0.01	0.0000
Années de Sc. Nat (Mère (i))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (j))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (k))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (l))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (m))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (n))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (o))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (p))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (q))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (r))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (s))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (t))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (u))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (v))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (w))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (x))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (y))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (z))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (aa))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ab))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ac))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ad))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ae))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (af))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ag))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ah))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ai))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (aj))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ak))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (al))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (am))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (an))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ao))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ap))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (aq))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ar))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (as))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (at))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (au))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (av))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (aw))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ax))	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
Années de Sc. Nat (Mère (ay))	0.00	0.00	0.0000	0.00	0.00	0.000						

Tableau F67 : Aidé - Sciences Sociales

	MODEL A		MODEL B		MODEL C		MODEL D		MODEL E		MODEL F		MODEL G		MODEL H		MODEL I		MODEL J					
	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev	Logit	StDev				
Constante	-1.00	0.01	0.0000	-0.79	0.02	0.0000	-0.84	0.04	0.0000	-0.99	0.02	0.0000	-1.06	0.08	0.0000	-0.86	0.09	0.0000	-1.06	0.10	0.0000	-0.11	0.01	0.3169
Années Scolaires Mère	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.01	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000	-0.02	0.00	0.0000
Entre 11 et 500 Livres	-0.03	0.01	0.0031	-0.05	0.01	0.0000	-0.06	0.01	0.0000	-0.03	0.01	0.0062	-0.04	0.01	0.0010	-0.02	0.01	0.0076	-0.03	0.01	0.0014	0.00	0.01	0.8807
Entre 51 et 100 Livres	-0.10	0.02	0.0000	-0.12	0.02	0.0000	-0.13	0.02	0.0000	-0.08	0.02	0.0000	-0.10	0.02	0.0000	-0.07	0.02	0.0000	-0.08	0.02	0.0001	-0.04	0.02	0.0025
Entre 101 et 200 Livres	-0.14	0.02	0.0000	-0.17	0.02	0.0000	-0.18	0.02	0.0000	-0.11	0.02	0.0000	-0.15	0.02	0.0000	-0.10	0.02	0.0000	-0.08	0.02	0.0000	-0.04	0.02	0.5177
Plus de 200 Livres	-0.23	0.02	0.0000	-0.26	0.02	0.0000	-0.27	0.02	0.0000	-0.24	0.02	0.0000	-0.24	0.02	0.0000	-0.18	0.01	0.0000	-0.15	0.02	0.0000	-0.16	0.02	0.0000
Position dans la classe en SS (Moyenne)																								
Position dans la classe en SS (Haute)																								
Score en Ss: Soc (SIMCET 2004)																								
Score en Ss: Soc (carré)																								
Kart-Type Score S-S (Classe -1)																								
AIDANT (SS)																								
Et. SEUL (S-S)																								
Et. LIVRE (S-S)																								
Confiance en Soi Scolaire																								
CSS (carré)																								
Appui des Parents																								
Appui (carré)																								
Bien-être à l'école																								
BIE (carré)																								
Années Soc (Niv2)																								
Années Soc (Niv3)																								
Années Soc (Niv4)																								
Impense Ss: Soc (Niv3)																								
Impense Ss: Soc (Niv4)																								
Étude Ss > 'des fois' Heblo																								
Étude Ss > 'des fois' Heblo																								
Va obligé à l'école (d'accord)																								
Va obligé à l'école (d'accord)																								
Va obligé à l'école (très d'accord)																								
... avoir un bon travail (d'accord)																								
... avoir un bon travail (très d'accord)																								
Crat tra Sup TECH																								
Crat tra UNIV																								
Fille																								
Redoublant																								
Préscolaire																								
Foyer NON Couple																								
Au moins a 1 Frere																								
Taille de la Classe																								
Moy Années Scol. Mère (4)																								
SPD Années Scol. Mère (4)																								
% Classe (-) Bas CSS																								
% Classe (-) Haute CSS																								
Années d'Expérience (PSS)																								
Positif (PSS)																								
Post-gradé (PSS)																								
Autre 1 Travail (PSS)																								
Niv Prepa Moyen (PSS)																								
Niv Prepa Haut (PSS)																								
Attentes: Sup TECH (PSS)																								
Attentes: UNIV (PSS)																								
Preoccup: DISCIPLINE (PSS)																								
Preoccup: RESULTATS (PSS)																								
Convert. Prog: MOYENNE (PSS)																								
Convert. Prog: HAUTE (PSS)																								
% Heures Prepa Cours (PSS)																								
Trav Coopérat: Moyen (PSS)																								
Trav Coopérat: Haut (PSS)																								
PRAT Indiv: (PSS)																								
PRAT Grp: (PSS)																								
PRAT Lib: (PSS)																								
Heabs Libs Ss:																								
N'States / Prof																								
MUNIS																								
PSUBV NS NFP																								
PSUBV NS FP																								
PSUBV 5 NFP																								
PSUBV 5 FP																								
Privé PAYANT																								

Tableau F68
Corrélations de Pearson entre les proportions d'élèves *aidants* et *aidés*
par classe et différentes caractéristiques de niveau classe et établissement.
(7061 classes - 8^{ème} grade - 2004)

	MATHÉMATIQUES				LANGAGE				SCIENCES de la NATURE				SCIENCES SOCIALES			
	% Aidants		% Aidés		% Aidants		% Aidés		% Aidants		% Aidés		% Aidants		% Aidés	
	Coeff	Sign	Coeff	Sign	Coeff	Sign	Coeff	Sign	Coeff	Sign	Coeff	Sign	Coeff	Sign	Coeff	Sign
Taille de la Classe	-0.03	**	0.14	***	-0.12	***	0.13	***	-0.06	***	0.07	***	-0.07	***	0.05	***
MOY Années Scol Mères	0.04	***	0.15	***	-0.44	***	0.01	ns	-0.21	***	0.03	**	-0.22	***	-0.02	ns
STD Années Scol Mères	-0.02	*	0.00	ns	0.11	***	0.07	***	0.05	***	0.01	ns	0.05	***	0.04	***
% Classe Bas BEE	-0.13	***	-0.26	***	0.16	***	-0.21	***	0.02	ns	-0.16	***	0.03	**	-0.13	***
% Classe Haute CSS	0.21	***	0.00	ns	0.27	***	0.05	***	0.25	***	0.02	ns	0.24	***	0.03	**
% Classe BAS Cap Cult	-0.04	***	-0.18	***	0.40	***	-0.08	***	0.20	***	-0.09	***	0.19	***	-0.04	***
% Classe MOYEN Cap Cult	0.00	ns	0.08	***	0.03	**	0.13	***	0.00	ns	0.09	***	0.01	ns	0.08	***
% Classe HAUT Cap Cult	0.04	***	0.12	***	-0.39	***	-0.004	ns	-0.19	***	0.03	**	-0.18	***	-0.01	ns
Score Moyen Classe (discipline)	0.11	***	0.24	***	-0.43	***	0.12	***	-0.17	***	0.13	***	-0.20	***	0.06	***
Écart-Type Classe (discipline)	0.00	ns	0.07	***	-0.08	***	0.06	***	-0.06	***	0.03	**	-0.03	**	0.02	*
Années d'Expérience	-0.02	ns	-0.04	***	0.07	***	0.00	ns	0.01	ns	-0.05	***	0.01	ns	-0.07	***
Postitre	0.00	ns	0.02	ns	0.01	ns	-0.01	ns	0.02	*	0.00	ns	0.02	ns	0.00	ns
Post-grade	-0.01	ns	0.00	ns	-0.09	***	-0.03	**	-0.01	ns	0.05	***	-0.03	**	0.00	ns
Autre Travail	-0.02	ns	0.01	ns	-0.02	ns	-0.02	**	-0.02	*	0.01	ns	-0.02	*	-0.04	***
Niv Prepa Bas	0.00	ns	-0.06	***	0.07	***	-0.01	ns	0.05	***	-0.05	***	0.02	**	-0.02	*
Niv Prepa Moyen	0.00	ns	-0.02	ns	-0.01	ns	-0.02	ns	-0.03	**	0.02	*	0.01	ns	0.01	ns
Niv Prepa Haut	0.00	ns	0.06	***	-0.06	***	0.03	**	-0.03	**	0.04	**	-0.03	**	0.01	ns
Attentes: SupTECH	-0.02	*	0.09	***	-0.06	***	0.07	***	-0.02	**	0.03	**	-0.04	***	0.06	***
Attentes: UNIV	0.06	***	0.09	***	-0.27	***	-0.01	ns	-0.12	***	0.02	ns	-0.12	***	-0.05	***
Préoccupé DISCIPLINE	-0.02	ns	-0.04	***	0.01	ns	-0.04	***	0.01	ns	0.00	ns	-0.02	ns	0.01	ns
Préoccupé RÉSULTATS	-0.01	ns	0.01	ns	-0.02	*	-0.01	ns	0.00	ns	0.01	ns	-0.03	**	0.00	ns
Couvert Prog BASSE	-0.02	*	-0.06	***	0.02	ns	-0.03	***	0.01	ns	0.01	ns	-0.04	**	0.00	ns
Couvert Prog MOYENNE	-0.03	**	0.00	ns	-0.01	ns	0.00	ns	0.00	ns	0.00	ns	0.00	ns	0.00	ns
Couvert Prog HAUTE	0.05	***	0.06	***	-0.01	ns	0.03	**	-0.01	ns	0.00	ns	0.03	**	-0.01	ns
% Heures Prepa Cours	0.01	ns	0.01	ns	-0.08	***	-0.03	**	-0.03	**	-0.01	ns	-0.05	***	-0.01	ns
Trav Coopérat Bas	-0.02	ns	-0.05	***	0.07	***	-0.02	ns	0.04	***	-0.02	ns	0.05	***	-0.01	ns
Trav Coopérat Moyen	-0.01	ns	-0.05	***	0.06	***	0.00	ns	0.01	ns	-0.02	*	0.00	ns	0.01	ns
Trav Coopérat Haut	0.02	*	0.08	***	-0.10	***	0.01	ns	-0.04	***	0.03	**	-0.04	***	0.00	ns
NO Prats (toutes)	-0.02	*	-0.01	ns	-0.04	***	0.00	ns	-0.05	***	-0.02	*	-0.04	**	-0.02	*
Plusieurs PRATs (PM)	0.01	ns	-0.02	**												
PRATs Indiv (PM)	0.01	ns	0.03	***												
Plusieurs PRATs (PL / PSS)					0.00	ns	-0.04	***					-0.05	***	-0.01	ns
PRATs Indiv (PL / PSS)					-0.01	ns	0.03	***					0.03	**	0.01	ns
PRATs Group (PL / PSS)					0.05	***	0.02	ns					0.05	***	0.02	ns
Tradit PRATs (PScN)									-0.02	ns	0.03	**				
Plusieurs PRATs (PScN)									0.07	***	-0.01	ns				
Étab a Labo Scs	0.01	ns	0.12	***	-0.23	***	0.05	***	-0.11	***	0.09	***	-0.12	***	0.06	***
N°élèves/Prof	0.00	ns	0.07	***	0.02	**	0.10	***	0.04	***	0.06	***	0.03	**	0.06	***
MUNI NS	-0.03	***	-0.15	***	0.27	***	-0.07	***	0.12	***	-0.09	***	0.12	***	-0.06	***
MUNI S	0.03	***	0.05	***	-0.03	**	0.04	***	-0.03	**	0.02	*	-0.02	ns	0.00	ns
PSUBV NS NFP	0.01	ns	-0.02	*	0.09	***	0.01	ns	0.06	***	0.00	ns	0.06	***	0.00	ns
PSUBV NS FP	-0.04	***	0.02	*	-0.03	**	0.02	ns	0.00	ns	0.02	ns	-0.01	ns	0.05	***
PSUBV S NFP	0.02	*	0.05	***	-0.03	**	0.05	***	-0.03	**	0.05	***	-0.01	ns	0.05	***
PSUBV S FP	0.01	ns	0.14	***	-0.18	***	0.10	***	-0.08	***	0.11	***	-0.08	***	0.07	***
Privé PAYANT	0.03	**	-0.02	*	-0.23	***	-0.11	***	-0.10	***	-0.06	***	-0.11	***	-0.07	***

Source : élaboration de l'auteur. Lecture : entre la proportion d'élèves aidants en maths et les années d'expérience des enseignants de cette discipline, par classe, on constate une corrélation de Pearson de -0,02 qui n'est pas significative (à p<0,1). Il faut garder à l'esprit que plusieurs variables sont de type catégoriel. Nous avons aussi ajouté les proportions d'élèves par classe selon niveau de capital culturel (variable qui n'est pas incluse dans les modèles).

ANNEXE « G »

QUESTIONNAIRES UTILISÉS POUR LA CONSTRUCTION DES BASES DES DONNÉES ¹³¹

¹³¹ Cette partie considère les questionnaires des élèves, des parents et des enseignants en mathématiques. Ceci compte tenu que les questionnaires des enseignants de langage, sciences de la nature et de sciences sociales sont identiques à celui appliqué pour les professeurs de maths à deux questions près, spécifiques à chaque discipline (la n°16 et la n°20). Les questionnaires sont présentés dans leur format et leur langue originaux.

I. ¿Hasse visto en clases los contenidos que se preguntaron en la prueba SIMCE de Estudio y Comprensión de la Naturaleza?		Si	No
a) Relaciones químicas			
b) Animales y plantas			
c) Energía en los cambios de estado de la materia			
d) Reproducción sexual y asexual			
e) Evolución de las especies y selección natural			
f) Crecimiento celular			
g) Ciclo celular eucariótico			
h) Investigación científica (preguntas, hipótesis, experimentos, etc.)			

II. ¿Hasse visto en clases los contenidos que se preguntaron en la prueba SIMCE de Estudio y Comprensión de la Sociedad?		Si	No
a) Principios geográficos y económicos (países, continentes)			
b) Guerra fría			
c) Tratados de paz de 1945			
d) Organismos internacionales para mantener la paz en el mundo (ONU)			
e) Globalización y su impacto en nuestro país y el mundo			
f) Problemas que vive la población debido al crecimiento poblacional y del mundo			
g) Derechos y responsabilidades que cada uno tiene en la comunidad y el Estado			

III. Clases		Si	No
8. ¿Cuáles de las siguientes actividades ocurren con mayor frecuencia en tus clases? (Pueden seleccionarse más de dos actividades.)			
a) En Lenguaje y Comunicación			
b) En Educación Matemática			
c) En Estudio y Comprensión de la Naturaleza			
d) En Estudio y Comprensión de la Sociedad			

9. ¿Cuáles de las siguientes actividades ocurren con mayor frecuencia en tus clases? (Pueden seleccionarse más de dos actividades.)		Si	No
a) Pizarra			
b) Computador			
c) Calculadora			
d) Internet			
e) Lenguaje de señas			

10. ¿Con qué frecuencia utilizas tu libro de texto para trabajar en clases?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) En clases de Lenguaje y Comunicación						
b) En clases de Educación Matemática						
c) En clases de Estudio y Comprensión de la Naturaleza						
d) En clases de Estudio y Comprensión de la Sociedad						

11. ¿Con qué frecuencia utilizas tu libro de texto para hacer tareas?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) En Lenguaje y Comunicación						
b) En Educación Matemática						
c) En Estudio y Comprensión de la Naturaleza						
d) En Estudio y Comprensión de la Sociedad						

12. ¿Con qué frecuencia tu profesor de Lenguaje y Comunicación realiza las siguientes actividades?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) Enseña como estudiar						
b) Realiza varias veces la materia						
c) Explica de distintas formas la materia						
d) Realiza las tareas						
e) Explica los errores que hay en las tareas, pruebas y trabajos						
f) Me ayuda cuando pasa una materia difícil						
g) Me motiva para que estudie y me daque buenas notas						
h) Le da lo mejor de mí cuando me da las notas						

13. ¿Con qué frecuencia tu profesor de Educación Matemática realiza las siguientes actividades?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) Enseña como estudiar						
b) Realiza varias veces la materia						
c) Explica de distintas formas la materia						
d) Realiza las tareas						
e) Explica los errores que hay en las tareas, pruebas y trabajos						
f) Me ayuda cuando pasa una materia difícil						
g) Me motiva para que estudie y me saque buenas notas						
h) Le da lo mejor de mí cuando me da las notas						

14. ¿Con qué frecuencia estás de acuerdo o en desacuerdo con cada uno de las siguientes situaciones?		Siempre de acuerdo	A veces de acuerdo	Siempre en desacuerdo
a) Hago siempre los deberes				
b) Me siento bien				
c) Me siento extraño y fuera de lugar				
d) Le consigo bien a mis compañeros				
e) Me siento solo				
f) Soy tímido/a				
g) Me aburro				
h) Me puse un sobreentorno que me molesta				

15. ¿Con qué frecuencia estudias fuera del horario de clases?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) Para Lenguaje y Comunicación						
b) Para Educación Matemática						
c) Para Estudio y Comprensión de la Naturaleza						
d) Para Estudio y Comprensión de la Sociedad						

16. ¿Con qué frecuencia te ocurren cada uno de las siguientes situaciones?		Siempre	Frecuentemente	A veces	Raramente	Nunca
a) Me siento capaz de entender la materia más difícil que presenta el profesor						
b) Confió en que puedo hacer un excelente trabajo en pruebas y tareas						
c) Si me propongo aprender bien algo, puedo hacerlo						
d) Si resoltoro sacarme buenas notas, realmente puedo lograrlo						
e) Cuando estudio, me distraigo porque soy muy malo para estudiar						
f) Cuando estudio, hago esfuerzos, aunque la materia sea difícil						
g) Cuando estudio, todo lo estudio para aprender						
h) Cuando estudio y no entiendo algo, busco información adicional para entenderlo						

Bibliographie

ANNEXES

- Arrègle J.-L. ; 2003 ; *Les modèles hiérarchiques: principes et illustration* ; M@n@gement, Vol. 6, n°1 ; pp. 1-28
- Bernard, J.-M. ; Simon, O. ; Vianou, K. ; 2005 ; *Le redoublement : mirage de l'école africaine?* ; Programme d'Analyse des Systèmes Educatifs de la CONFEMEN ; 96 pp.
- Bickel, R.; 2007 ; *Multilevel analysis for applied research: It's just regression !* ; Guilford Press, NY ; 355 pp.
- Birch, E.R. ; Miller, P. ; 2006 ; *Student outcomes at University in Australia: a Quantile Regression approach* ; Australian Economic Papers ; Blackwell Publishing Ltd/University of Adelaide and Flinders University ; 17 pp.
- Bonett, D.; Price, R.; 2005 ; *Inferential Methods for the Tetrachoric Correlation Coefficient*; Journal of Educational and Behavioral Statistics, Vol.30, n°2 ;pp. 213-225
- Boudon, R. ; 2002 ; *Sociology That Really Matters* ; European Sociological Review, Vol.18, n°3 ; pp. 371-378
- Bressoux, P. ; 2008 ; *Modélisation statistique appliquée en sciences sociales* ; De Boeck, Paris ; 443 pp.
- Bressoux, P. ; Coustère, P. ; Leroy-Audouin, C. ; 1997 ; *Les modèles multiniveau dans l'analyse écologique : le cas de la recherche en éducation* ; Revue Française de Sociologie, Vol.38 ; pp. 67-96
- Buchinsky, M. ; 1998 ; *Recent Advances in Quantile Regression Models: A practical guide for empirical research* ; Journal of Human Resources, Vol.33, n°1 ; pp. 88-126
- Caille, J.P. ; 2001 ; *Scolarisation à deux ans et réussite de la carrière scolaire au début de l'école élémentaire* ; Éducation et Formations, n°60 ; pp. 7-18
- Chaix B. ; Chauvin P. ; 2002 ; *L'apport des modèles multiniveaux dans l'analyse contextuelle en épidémiologie sociale : une revue de la littérature* ; Revue d'épidémiologie et de santé publique, Vol.50, n°5 ; pp. 489-499
- Chen, C. ; 2005 ; *An Introduction to Quantile Regression and the QUANTREG Procedure* ; Paper n°213-30 ; SAS Institute Inc., Cary, NC ; 24 pp.
- CISIA-CERESTA ; 2001 ; SPAD-Base, Aide à l'interprétation ; 126 pp.
- Crahay, M.; 2003 ; *Peut-on lutter contre l'échec scolaire ?* ; De Boeck, Bruxelles ; 378 pp.
- Demeuse, M.; 2003 ; *Les analyses factorielles. Quelques applications d'analyses factorielles dans le domaine de l'éducation et de la formation* ; Séminaire IREDU (16/12/2003).
- Demeuse, M.; Henry, G.; 2004 ; *La fidélité des mesures, Chap. III* ; In Introduction aux Théories et aux Méthodes de la mesure en Sciences Psychologiques et en Sciences de l'Éducation, Demeuse, M.; pp. 149-162
- Durand, C. ; 2003 ; *L'analyse factorielle et l'analyse de fidélité notes de cours et exemples* ; Université de Montréal, Département de Sociologie ; 30 pp.
- Falissard, B.; 1998 ; *Comprendre et utiliser les statistiques dans les sciences de la vie (2^{ème} Ed.)* ; Masson, Paris ; 344 pp.
- Gujarati, D. ; 2004 ; *Économétrie (Trad. de la 4^{ème} Ed. américaine par Bernard Bernier (Basic Econometrics))* ; De Boeck, Bruxelles ; 1009 pp.

- Heckman, J.J.; 1979 ; *Sampleselectionbias as a specification error* ; Econometrica, Vol.47, n°1; pp. 153-162
- Hox, J.J. ; 2002 ; *Multilevelanalysis. Techniques and applications*. LEA, London ; 304 pp
- Koenker, R. ; 2005 ; *Quantile Regression* ; Cambridge University Press ; 349 pp.
- Koenker, R. ;Hallock K. ; 2001 ; *Quantile Regression: An Introduction* ; Journal of Economic Perspectives, Vol.15, n°4 ; pp. 143–156.
- Leblond, S. ; 2003 ; *Guide d'économétrie appliquée pour Stata* ; Université de Montréal ; 50 pp.
- MIDEPLAN ; 2007 ; *Documento N°2: Distribución del Ingreso e Impacto Distributivo del Gasto Social* ; Serie Análisis de resultados de la Encuesta de Caracterización Socioeconómica Nacional, CASEN 2006 ; 44 pp.
- Mingat, A. ; Jarousse, J.P. ; Richard, M. ; 1992 ; *La scolarisation maternelle à deux ans : effets pédagogiques et sociaux* ; Éducation et Formations, n°31 ; pp. 3-9
- Monseur, C ; Lafontaine, D. ; 2005 ; *Methodological Issues Raised by Equity Indicators Derived from Multilevel Analyses* ; Working Paper ; 24 pp.
- Panik, M.; 2009 ; *Regressions Modeling. Methods, Theory and Computation with SAS* ; CRC Press, NY ; 814 pp.
- Paul J.J. ; Troncin T. ; 2004 ; *Les apports de la recherche sur l'impact du redoublement comme moyen de traiter les difficultés scolaires au cours de la scolarité obligatoire* ; Rapport du Haut conseil de l'évaluation de l'école (HCEE), n°14 ; 44 pp.
- Raudenbush S.W. ;Bryk A.S. ; 1992 ; *Hierarchical Linear Models. Applications and data analysismethods* ;SagePublications, London ; 265 pp.
- Ray J.-C.; 2004 ; *Transferts sociaux et gains d'activité des jeunes adultes européens. Une application des modèles multiniveaux* ; Économie et Prévision, Vol.3-4, n°164 ; pp. 189-210
- Robin, J.-M.; 2003 ; *Éndogénéité et variables instrumentales dans les sciences sociales* ; *Actes des journées de méthodologie statistique* ; Document de Travail INSEE-Méthodes n°101 ; pp. 217-276
- SAS Institute ; 2005 ; *The QUANTREG Procedure (Experimental)* ; SAS Institute Inc., Cary, NC ; 70 pp.
- Schabenberger, O.; 2005 ; *Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models* ; Paper 196-30 ; SAS Institute Inc., Cary, NC ; 20 pp.
- Singer, J.; 1998 ; *Using Proc Mixed to fit multilevel models, hierarchical models and individual growth models* ; Journal of educational and behavioural statistics, Vol.24, n°4 ; pp. 323-355
- STATA ; 2003 ; *Base Reference Manual, Vol.3 (N-R) Release 8* ; Stata Press Publication, Texas ; 464 pp.
- Vivanco, M. ; 1996 ; *Analisismultivariable* ; Universidad de Chile, Departamento de Sociología, Serie Cuadernos de Trabajo N°3 ; 192 pp.
- Wooldridge, J.M. ; 2006 ; *Introductory Econometrics: a modern approach (3th Ed.)*; Mason ; 890 pp.