

COMMENT L'ÉCOLE AMPLIFIE LES INÉGALITÉS SOCIALES ET MIGRATOIRES ?

Inégalités scolaires : quels rôles jouent les cours privés ?

ARNAUD GALINIÉ & ARTHUR HEIM

Chargés de missions au Conseil National d'Évaluation du Système Scolaire

heimarthur@gmail.com



cnesco
conseil national
d'évaluation
du système scolaire

Ce document s'inscrit dans une série de contributions publiées par le Conseil national d'évaluation du système scolaire (Cnesco) dans le cadre de son rapport scientifique : **comment l'école amplifie les inégalités sociales et migratoires ?**

Il a bénéficié d'un processus de révision analytique et éditorial piloté par Nathalie Mons, présidente du Cnesco et professeure à l'université de Cergy Pontoise, et Jean-François Chesné, directeur scientifique du Cnesco, avec la participation de Dominique Goux, chercheuse au CREST et membre du Cnesco, Arnaud Galinié, Emily Helmeid, Arthur Heim, Alain Lopez et Philippe Romain, chargés de missions au Cnesco.

Ce travail s'appuie principalement sur l'analyse des données du panel d'élèves entrés en 6^e en 2007 (MENESR-DEPP) apparié à des données publiques disponibles sur le site insee.fr. Un graphique mobilise également les données des enquêtes PISA des cycles 2006, 2009 et 2012.

Il mobilise également des éléments des recherches exploratoires entreprises avec Claire Steinmetz, alors chargée de missions au Cnesco. Les auteurs souhaitent la remercier, notamment pour son travail de revue de littérature. Ils remercient également Marc Gurgand pour ses précieux commentaires sur ce travail, et Patrice Caro pour avoir produit les cartes des entreprises de soutien privé à partir de leur données.

Les opinions et arguments exprimés n'engagent que les auteurs de la contribution.

Pour citer cet article :

Galinié, A. & A. Heim, (2016). Inégalités scolaires : quels rôles jouent les cours privés ?. Contribution au rapport du Cnesco *Les inégalités scolaires d'origines sociales et ethnoculturelle*. Paris. Cnesco.

Disponible sur le site du Cnesco : <http://www.cnesco.fr>

Publié en Septembre 2016
Conseil national d'évaluation du système scolaire
Carré Suffren - 31-35 rue de la Fédération
75015 Paris

Table des matières

Résumé	8
Introduction	12
I Un phénomène mondial, d'ampleur inégale, relativement peu documenté	13
1 L'ampleur du <i>shadow education</i> dans le monde	13
2 Les cours privés dans le panel 2007	15
II Les déterminants du soutien privé	17
1 Qui prend des cours privés de soutien, et pourquoi?	17
2 Des caractéristiques socio-démographiques identifiables	20
III L'effet du soutien privé sur les performances des élèves	26
1 Les performances des élèves dans le panel 2007	26
2 Le soutien privé en 6 ^e	31
3 L'effet des cours de soutien en 4 ^e /3 ^e	46
IV Cours privés, trajectoires et orientations des collégiens	50
1 Des collégiens aux trajectoires diverses	50
2 Effet des cours de soutien sur les trajectoires	52
Conclusion	57
Annexes	57
Bibliographie	80

Table des figures

Figure 1	Distributions des performances entre 2008 et 2011 suivant la variable soutien scolaire	29
Figure 2	Évolution des performances moyennes des élèves entre 2008 et 2011 suivant le moment où ils suivent des cours de soutien privés	30
Figure 3	Différence conditionnelle de performances due au soutien scolaire en 2008, par quantile.	35
Figure 4	Nombre d'entreprises de soutien privé et performances moyennes	37
Figure 5	Trajectoires au collège.....	51
Figure A.1	Cours additionnels en mathématiques d'après PISA.....	60
Figure A.2	Histogramme des variables instrumentales en 2008	65
Figure A.3	Entreprises privées dans les communes de France fournissant du soutien scolaire privé en 2008	66
Figure A.4	Entreprises privées dans les communes de France fournissant du soutien scolaire privé en 2011	67
Figure A.5	Estimations par régressions quantiles en 2011	73
Figure A.6	Estimations par régressions quantiles en 2011 en tenant compte du niveau de 2008	74

Liste des tableaux

Table 1	Proportion d'élèves suivant des cours de soutien privés en 2008 et 2011.	16
Table 2	Effet des caractéristiques observables sur la probabilité de prendre des cours de soutien payant en 2008 et 2011	20
Table 3	Différence de performances en 2008 entre élèves suivant ou non des cours privés de soutien, avec ou sans variables de contrôle	33
Table 4	Estimation de l'effet des cours particuliers sur les performances des élèves en 2008 par variable instrumentale et Heckit	43
Table 5	Régressions instrumentales <i>placebo</i>	46
Table 6	Trajectoires en fin de troisième et soutien scolaire	52
Table 7	Différence de trajectoires entre élèves suivant ou non des cours en 2008, avec ou sans variables de contrôles	54
Table 8	Différence de trajectoires entre élèves suivant ou non des cours en 2011, avec ou sans variables de contrôles et échantillon restreint ou non	55
Table A.1	Pourcentage d'élèves étudiant chaque discipline dans le cadre de cours de soutien payants	61
Table A.2	Équation de sélection (<i>first stage</i>) des modèles 2SLS en 2008	68
Table A.3	Différence de performances en 2011 entre élèves suivant ou non des cours, avec ou sans variables de contrôle, performances passées et soutien passé.	71
Table A.4	Équation de sélection (<i>first stage</i>) des modèles 2SLS en 2011	75
Table A.5	Estimation de l'effet des cours particuliers sur les performances des élèves en 2011 par variable instrumentale et Heckit	78
Table A.6	Estimation de l'effet des cours particuliers sur les performances des élèves en 2011 par variable instrumentale et Heckit	79

Résumé

En 2013, la publication des résultats de l'enquête PISA fait l'effet d'une petite bombe dans les médias en pointant du doigt d'une part, l'importance des inégalités de performances entre les élèves scolarisés en France, et d'autre part, leur aggravation par rapport aux enquêtes passées. Les mécanismes qui transmettent les inégalités sociales aux inégalités scolaires sont multiples et ont fait l'objet de nombreuses études auxquelles les autres contributions au rapport du Cnesco font largement référence. Cependant, le recours au soutien scolaire payant est de loin celui le moins étudié, faute de données satisfaisantes pour en mesurer correctement les effets. Des problèmes plus conceptuels rendent également difficile l'étude de ce phénomène ; le soutien privé pouvant prendre des formes très différentes, être dispensé individuellement ou en groupe, par des intervenants de compétences et de qualifications variées, avec des méthodes très hétérogènes, se focalisant sur des sujets ou des axes de travail différents, ... rendant la définition du soutien privé et la comparaison des élèves extrêmement complexes.

Le soutien privé a pu être activement contrôlé, voire interdit dans certains pays¹ en réponse aux nombreuses critiques qui lui sont adressées. En effet, le soutien privé est très souvent accusé d'accentuer les inégalités sociales et de détériorer le système d'éducation public sans nécessairement parvenir à améliorer les performances scolaires des élèves qui y ont recourt (Dang et Rogers, 2008).

Ce rapport est, à notre connaissance, le premier à s'intéresser aux effets des cours particuliers sur les performances des élèves et sur leurs trajectoires scolaires en France. Il se fonde sur une analyse économétrique des données du panel d'élèves du second degré constitué par la DEPP en 2007. Ces données permettent de suivre le parcours d'une génération d'élèves entrés en 6^e en 2007 jusqu'à la première (pour l'instant), et contient de nombreuses informations sur leurs caractéristiques socio-démographiques ainsi que sur leurs performances. Il permet d'identifier leur profil et les particularités de leur environnement scolaire et familial.

Ainsi, les élèves qui prennent des cours de soutien scolaire au collège sont plutôt moins performants que ceux qui n'en prennent pas, mais l'appréciation de leur niveau par leurs parents détermine davantage leur probabilité de suivre des cours privés de soutien que leurs performances réelles. L'opinion de ces derniers sur la qualité des enseignants, de l'établissement et leur ambition pour leur enfant jouent également un rôle significatif. D'autre part, Les élèves qui suivent des cours privés de soutien sont plus souvent issus de familles plutôt favorisés, citadines et peu nombreuses. On note en outre que les familles d'origines immigrés ont plus souvent recours au soutien scolaire que les familles françaises d'origine, ce qui rejoint les travaux déjà existant sur la question (Brinbaum et Kieffer, 2005, 2009).

Si on compare les résultats, à la fin de l'année, des élèves qui ont eu des cours de soutien et de ceux qui n'en ont pas eu, on constate que les premiers sont plus faibles que les seconds. Mais cela s'explique

1. Notamment au Cambodge, à l'Île Maurice, mais aussi en Corée du Sud où son usage est pourtant massif (Bray, 2000).

largement par le fait que les élèves qui bénéficient de cours privés sont en moyenne plus faibles au départ et que les cours ne suffisent donc pas à compenser entièrement ces écarts. Cela ne signifie pas que les bénéficiaires n'ont pas tiré profit des cours. Si l'on compare maintenant des élèves bénéficiaires ou non de cours, mais de niveau initial observé et d'origine sociale semblables, on trouve des écarts de performance beaucoup plus faibles et dans certains cas nuls, ce qui impliquerait que les cours sont sans effet. Mais, dans ce cas, si la distribution de la plupart des scores de fin d'année est assez homogène entre les deux groupes, les mathématiques et le raisonnement en 6^e constituent une exception. Dans ces deux dimensions, l'écart de performances final entre les élèves qui prennent des cours et ceux qui n'en prennent pas - toujours conditionnellement aux caractéristiques initiales des élèves - est plus important pour les élèves initialement faibles que pour les élèves forts, ce qui suggère que les cours de soutien privés pourraient rendre la distribution des performances plus inégales dans ces deux dimensions.

Cependant, pour estimer tout à fait l'effet causal des cours privés, pour tout élève donné, il faudrait mesurer *ce qui se serait passé* pour cet élève s'il n'avait pas pris de tels cours. Or les caractéristiques que nous observons (niveau de départ, origine sociale) ne suffisent pas à prédire complètement les progrès potentiels des élèves. Parmi les élèves de caractéristiques données, ce sont peut-être précisément ceux qui ont des difficultés particulières, non mesurées dans les enquêtes, qui se tournent plus souvent vers les cours privés. On sous-estimerait alors le bénéfice des cours pour ces élèves. Afin de surmonter cette difficulté, nous avons utilisé une approche indirecte². Nous avons mesuré l'offre de cours privés dans chaque commune de notre échantillon : même des communes de caractéristiques proches peuvent avoir des offres de cours sensiblement différentes. Or on observe que les enfants qui, un peu par hasard, se trouvent dans des communes où l'offre est abondante ont plus de chance de prendre des cours. Si les cours privés ont des effets sur les progressions, à niveau de départ donné, les élèves qui se trouvent dans des communes bien dotées devraient progresser davantage puisqu'une plus grande proportion d'entre eux bénéficient de cours. Les estimations réalisées en suivant ce principe conduisent aux résultats suivants : les cours de soutien privés ont un impact positif et significatif en 6^e, notamment sur les compétences en mathématiques, en compréhension de texte et en "culture générale". Ainsi, si les élèves concernés sont au départ plutôt plus faibles, le fait de bénéficier de cours de soutien privés leur permet de progresser. Cependant, les capacités en raisonnement logique sont négativement affectées par les cours de soutien privés, ce que nous interprétons comme le produit d'un travail davantage axé sur la préparation d'exercices types, la reproduction de contenu pédagogique, etc., au détriment de résolution de problèmes ou d'autres activités développant la logique. En revanche, en 4^e/3^e, la plupart des estimations sont imprécises et n'excluent pas l'absence d'effet. À nouveau, seules les compétences en raisonnement logique sont négativement et significativement affectées par le soutien scolaire.

Au total, la sélection des élèves qui prennent des cours privés est telle qu'il est difficile de mesurer de façon certaine les effets moyens de tels cours. Les différentes méthodes que nous avons mises en œuvre donnent à penser que les effets sont probablement assez faibles, avec ponctuellement quelques effets positifs et négatifs, en 6^e seulement, et pour certaines compétences très scolaires.

Enfin, si les performances des élèves en fin collège ne semblent pas améliorées par les cours de soutien, l'accès à la seconde générale et technologique apparaît facilité. En effet, qu'ils aient ou non eu des difficultés

2. Nous utilisons des modélisations à *variable instrumentale*.

dans leur scolarité au collège, les élèves qui suivent des cours privés sont plus fréquemment en seconde générale et technologique que ceux qui n'en suivent pas, toutes choses égales par ailleurs. Cette analyse sur les trajectoires doit cependant être interprétée de façon descriptive puisque nous n'avons pu trouver de stratégie crédible permettant d'identifier une relation causale.

Ce rapport fournit ainsi une évaluation originale sur un sujet trop peu documenté et contribue à nourrir le débat sur l'efficacité des cours de soutien privés et leur impact sur les inégalités scolaires. Il comporte toutefois quelques insuffisances qui rendent nécessaires davantage de recherches sur la question. D'abord, il ne s'intéresse qu'aux élèves du second degré inférieur et ne permet pas d'inférer quoi que ce soit sur les caractéristiques des élèves qui prennent des cours plus tard dans leur scolarité, ni sur l'effet que cela peut avoir sur leur réussite future. Ensuite, l'offre de soutien scolaire n'est pas du tout étudiée dans ce rapport, faute d'informations sur ceux qui dispensent ces cours. La potentielle variété de leurs profils peut fragiliser la modélisation employée ici, qui repose sur un effet moyen homogène. Enfin, la méconnaissance du recours au soutien scolaire entre les deux enquêtes et tests fragilise nos estimations concernant son efficacité en fin de collège.

Inégalités scolaires : quels rôles jouent les cours privés ?

Introduction

Ce travail est divisé en 4 sections : la première définit et caractérise le phénomène de *shadow education*, la seconde analyse les facteurs influençant la demande de cours de soutien privés, la troisième évalue l'effet de ces cours sur les performances des élèves en 6^e et en 4^e/3^e, la dernière estime leur effet sur les trajectoires au début du lycée.

I Un phénomène mondial, d'ampleur inégale, relativement peu documenté

Le recours au soutien privé s'est fortement développé à partir des années 80-90, à la fois dans les pays en développement et dans les pays développés. Au Japon et en Corée du Sud, jusqu'à 9 élèves sur 10 prennent des cours de soutien privés à un moment de leur scolarité. En Europe, les familles dépensent chaque année plusieurs milliards d'euro dans le soutien scolaire privé (Bray, 2011), surtout en Europe du Sud et en Europe Centrale. En France, le pourcentage d'élèves qui ont recours au soutien scolaire est important et en forte croissance même si les estimations de la taille de ce marché sont très imprécises. Dans certains pays, la dépense des familles en soutien privé atteint le niveau de la dépense publique d'éducation comme en Turquie ou en Corée du Sud (où elle atteint 2.9 % du PIB (Dang et Rogers, 2008)).

1 L'ampleur du *shadow education* dans le monde

"L'éducation de l'ombre", ou *shadow education*, correspond au développement d'un système éducatif parallèle à celui de l'école traditionnelle. Il regroupe les activités éducatives ayant lieu en dehors de l'école (le soir, le week-end ou pendant les vacances scolaires) qui prennent la forme de soutien privé (le plus souvent payant), d'écoles de *bachotage* ou de préparation aux examens. Ce soutien privé peut être proposé individuellement, en petit groupe ou en classe entière. Les cours particuliers ont plusieurs fonctions : apporter de l'aide pour le rattrapage, aider les élèves en retard ou ayant des difficultés, éviter de redoubler, se perfectionner pour mettre toutes les chances de son côté afin de réussir des concours ou entrer dans une voie sélective (Glasman et Besson, 2004). Sur bien des aspects, le *shadow education* est indistinct de certains dispositifs de remédiation ou de tutorat proposé par le système scolaire, les collectivités locales, . . . Ainsi, de nombreuses études s'intéressant exclusivement au soutien privé recensent, malgré tout, les articles qui traitent de provision publique³, de tutorats ou de cours du soir (Dang et Rogers, 2008 ; Banerjee

3. Proposé par l'Etat, la localité, l'école, des organisations locales, . . . le plus souvent ces cours sont gratuits.

et al., 2007 ; Dang, 2007 ; Zimmer et al., 2010 ; Zhang, 2013 ; Hof, 2014). Cette littérature fournit un point de référence de l'effet de ce type de programme sur une population qui n'y souscrit pas forcément volontairement et qui n'en supporte pas le coût.

Le phénomène de *shadow education* s'est d'abord fortement développé en Asie de l'Est, et notamment au Japon et en Corée du Sud dans les années 1980-1990, puis en Afrique, en Amérique du Sud et en Europe de l'Est (Bray, 2013). S'il est aujourd'hui présent dans le monde entier, on peut cependant distinguer deux groupes de pays (Centre d'analyse stratégique, 2013) : d'un côté, les pays scandinaves, le Royaume-Uni, l'Allemagne et les États-Unis dans lesquels le recours au soutien scolaire reste assez mineur, de l'autre, les pays d'Asie du Sud-Est ou d'Europe de l'Est dans lesquels la demande en soutien scolaire est très importante, probablement en raison de l'attention accrue donnée à la réussite scolaire ou de la faiblesse de l'école traditionnelle. La France, elle, se situe entre ces deux modèles.

Il est assez difficile de connaître l'importance du soutien privé auprès des élèves français tant ce phénomène est difficile à définir. L'étude de marché de Melot (2007) estime qu'en France, 1/4 des élèves du secondaire inférieur et 1/3 du secondaire supérieur ont recours à du soutien scolaire, et ce taux peut atteindre jusqu'à 3/4 des élèves des grands lycées parisiens. D'après les estimations de Glasman et Besson (2004) et Melot (2005), le nombre d'élèves ayant recours à du soutien privé payant varierait entre 850 000 et 2 millions par an en France.

Le marché du soutien scolaire : ampleur et coût

Le marché du soutien scolaire représenterait 2 milliards d'euros en France (Ministère de l'Éducation nationale, 2006), ce qui est beaucoup plus élevé que dans les autres pays d'Europe tels que Chypre, l'Allemagne ou la Grèce (Bray, 2011). L'INRP (2006) décompose ce chiffre selon les différents types de soutien privé :

- **l'accompagnement scolaire** représenterait 100 millions d'euros, soit 5 % du marché (Ministère de l'Éducation nationale, 2006) ;
- **le soutien privé "déclaré"** représenterait plus de 600 millions d'euros (Melot, 2005) ou 800 millions d'euros d'après les chiffres d'Academia de 2004 (INRP, 2006), soit entre 30 et 40 % du marché ;
- **le soutien privé "informel"** se situerait entre 1,1 et 1,3 milliard d'euros, représentant 55 à 65 % du marché.

Le prix d'une heure de cours privé varie sensiblement d'un pays à l'autre et également en fonction du niveau scolaire^a :

- **Belgique** : Educadomo, qui se revendique comme le "premier réseau de cours privé à domicile" demande 75 € de droits d'inscription et 33 € par heure auxquels s'ajoute le coût de transport de l'intervenant.
- **Irlande** : Le coût moyen d'une heure de cours de soutien est d'environ 30 € mais ce coût varie suivant les localités, les matières et le niveau scolaire.
- **Portugal** : Une heure de cours privé coûte en moyenne 15 € de l'heure, soit deux fois moins que dans les pays précédemment cités.

- **Grande Bretagne** : Home Tutors UK, qui se revendique "leading tutoring agency" propose des cours de soutien pour des élèves de 3^e à 20 £, soit environ 25 € suivant le taux de change. L'intervenant facture entre 2 et 4 £ additionnels pour le transport.
- **France** : Une heure de cours dispensé par Acadomia coûte 33,50 € TTC, mais peut être déduite des impôts jusqu'à 50 %.

a. Ces estimations viennent de [Bray \(2011\)](#) sauf pour la France où ce sont les tarifs applicables en juillet 2015

Au niveau international, le temps consacré au soutien scolaire par les élèves du secondaire pour améliorer leurs résultats - que ce soit pour les élèves en difficulté ou pour conforter les meilleurs - varie selon les pays. D'après l'[OCDE \(2011\)](#), 26 % des élèves de 15 ans déclarent suivre jusqu'à quatre heures de soutien en mathématiques par semaine, 17 % en sciences, 16 % en langue de l'évaluation et 19 % dans les autres matières. Une part plus faible suit plus de quatre heures de soutien par semaine : 7 % en mathématiques, 5 % en langue d'enseignement et en sciences et 7 % dans les autres matières (d'après l'enquête PISA 2009). Plusieurs études présentent l'ampleur du soutien privé dans le monde mais il est assez difficile de les comparer parce qu'elles ne font pas toujours référence au même niveau de scolarité ou au même type de soutien ([Bray et Lykins, 2012](#) ; [Bray, 2011, 2013](#) ; [Bray et Kwok, 2003](#)).

Pour comparer l'évolution du soutien scolaire entre les pays de l'OCDE, nous avons représenté la part d'élèves qui déclarent suivre au moins une heure de cours supplémentaire par semaine en mathématiques à partir des enquêtes PISA. Nous présentons ces résultats dans le graphique A.1 de l'Annexe A.

Ces résultats surprennent par leur ampleur, quel que soit le pays de l'OCDE. En effet, le soutien en mathématiques concerne entre 1 élève sur 4 dans certains pays nordiques et près de 3 élèves sur 4 en Corée du Sud. Le classement des pays évolue sensiblement d'une enquête à l'autre alors que la question est restée la même. Cette évolution peut être due à des changements de politique scolaire ou des modifications des incitations à recourir à des cours de soutien privé. En France, Depuis 2005, une partie des heures de soutien privé peut faire l'objet d'une exonération ou d'un crédit d'impôts, créant ainsi une incitation monétaire forte pour les familles. Pourtant, il semble qu'une part plus faible d'élèves suive des cours de soutien en 2009 et 2012 par rapport à 2006. Le taux d'élèves qui déclarent suivre au moins une heure de cours supplémentaire en mathématiques est passé d'environ 55 % en 2006 à 37 % en 2009 et 2012.

2 Les cours privés dans le panel 2007

La section précédente a montré à quel point le recours au soutien privé était massif dans de nombreux pays - y compris en France - et combien les estimations étaient variables suivant les sources. Nous disposons de données sur un échantillon représentatif de 35 000 élèves entrés en 6^e en 2007, dont la famille a été interrogée deux fois à quatre ans d'intervalle. Ainsi, en 2008 et 2011, les parents des élèves de l'échantillon ont répondu à la question :

Votre enfant prend-il des cours particuliers ou de soutien scolaire payants dans l'une des disciplines suivantes ?

Environ 10 % des élèves de 6^e prennent des cours payants de soutien privés⁴. Quatre ans plus tard, ils

4. [Caille \(2010\)](#) avait présenté ce chiffre mais ne disposait pas encore de l'enquête famille 2011.

sont 14 %. Cependant, seuls 3,5 % des élèves prennent à la fois des cours en 2008 et 2011, **près de 2/3 des élèves qui prenaient des cours en 2008 n'en prennent plus en 2011.**

A contrario, **3/4 des élèves qui prennent des cours en 2011 n'en prenaient pas en 2008.** On perçoit ici l'avantage d'un suivi longitudinal par rapport à des données en coupe qui permettent de montrer que le recours au soutien privé aux différents moments de la scolarité, ne concerne pas les mêmes élèves ; élément que nous nous efforcerons de garder à l'esprit dans la suite de l'analyse.

Ce constat nous amène à nous interroger sur les raisons qui peuvent pousser un élève ou ses parents à souscrire à des cours de soutien privés à différents moments, avec des interruptions, ou à ne plus y souscrire après une ou deux années consécutives. Nous donnons des pistes d'interprétations basées sur les résultats de la littérature scientifique internationale dans la section suivante.

Le tableau A.1 de l'Annexe B nous montre que, en 2008, plus de 40 % des élèves qui prennent des cours de soutien travaillent le français, les mathématiques ou suivent des cours de soutien plus généraux⁵. En 2011, plus de 2/3 des élèves qui prennent des cours le font en mathématiques⁶. Un peu plus d'un quart des élèves qui prennent des cours de soutien travaillent le français, et un peu plus d'un élève sur cinq travaille sa première langue vivante ou les méthodes de travail. Ainsi, d'une part, ce ne sont pas les mêmes élèves qui prennent des cours en 2008 et 2011 et d'autre part, pas les mêmes disciplines qui sont travaillées dans ces cours. On peut donc s'attendre à des effets différents en raison de l'âge des élèves, de leurs profils et des domaines qui sont travaillés.

En 2008, 40 % des élèves qui prennent des cours n'en prennent qu'une heure par semaine, et 30 % en prennent deux heures, en moyenne. En 2011, ils sont environ 45 % à ne prendre qu'une heure et 33 % à prendre des cours deux heures par semaine.

Les cours de soutien scolaire au collège ne représentent pas un gros volume horaire. À titre de comparaison, en France, les élèves âgés de 15 ans en 2012 déclarent passer en moyenne environ 5h hebdomadaires aux devoirs à la maison (OCDE, 2015).

Tableau 1 – Proportion d'élèves suivant des cours de soutien privés en 2008 et 2011.

	2008	2011
Pas de soutien scolaire	90,6 %	85,9 %
soutien scolaire	9,4 %	14,1 %
Soutien en 2008 seulement	6,0 %	
Soutien en 2011 uniquement	10,7 %	
Cours en 2008 et 2011	3,5 %	
"Jamais" de cours	79,9 %	
N	32 629	29 502

Sources : Panel 2007 (MENESR-DEPP), enquêtes famille, calculs des auteurs, pondérations enquêtes famille respectives.

Lecture : en 2008, 90,6 % des élèves entrés en 6^e en 2007 ne prennent pas de cours de soutien privé.

5. La véritable modalité du questionnaire est "Aide au travail scolaire toutes disciplines", en 2011 la question portait sur la "méthode de travail".

6. Les cours ne sont pas mutuellement exclusifs, c'est pourquoi sommer ces pourcentages peut dépasser 100.

II Les déterminants du soutien privé

Les caractéristiques des élèves qui souscrivent à des cours de soutien peuvent varier grandement avec les raisons qui motivent ce choix. Dans cette section, nous décrivons les principales motivations et caractéristiques de ces familles analysées dans la littérature et décrivons en détails celles des élèves du panel 2007.

1 Qui prend des cours privés de soutien, et pourquoi ?

A Élèves faibles ou performants ?

De manière générale, on aurait tendance à penser que les élèves en difficulté sont ceux qui souscrivent le plus à du soutien privé. Cette croyance se vérifie parfois : à Hong Kong, [Bray et al. \(2014\)](#) montrent que les élèves de classes équivalentes à la 3^e et à la terminale se déclarant parmi les moins performants ont plus tendance à suivre du soutien privé que les autres élèves. Cependant, la probabilité d'y souscrire est plus élevée pour les élèves dans les meilleurs établissements. De même, [Zhang \(2013\)](#) montre qu'en Chine, le soutien scolaire concerne principalement les lycéens les moins performants et les élèves des établissements ayant le plus de ressources. Au final, de nombreux travaux de recherche montrent que les élèves ayant déjà de bons résultats souscrivent davantage à des cours, afin de maintenir leur niveau, de se perfectionner ou de préparer des concours ou des examens ([Bray et Kwok, 2003](#) ; [Bray, 2011](#)).

On perçoit ici deux rôles très différents associés au soutien privé : un rôle de remédiation poussant les familles qui le peuvent à faire aider leur enfant pour lui permettre de rattraper son retard, et un rôle de perfectionnement où les familles intensifient l'investissement dans l'éducation de leur enfant de façon à conforter leur avance. Le cas de la France est décrit plus en détail dans les sections suivantes.

Les travaux de Jean-Paul Caille (2010) sur le soutien scolaire en 6^e

Le rapport de [Caille \(2010\)](#) présente les différents déterminants de la demande en soutien privé à partir des premières vagues d'enquête du panel 2007 et s'intéresse aux caractéristiques individuelles et familiales des élèves de 6^e. Alors que des recherches étrangères ont montré que les cours de soutien étaient fortement suivis par les élèves performants, les données traitées par Jean-Paul Caille montrent que les 3/4 des élèves qui en bénéficient sont ceux en difficulté (le niveau de difficulté est celui perçu par les parents). Ainsi, parmi les élèves jugés comme ayant de grosses difficultés par leurs parents, 19 % suivent du soutien privé contre 2,8 % pour les élèves jugés excellents. Plus l'élève est en difficulté, plus il suivra des cours de soutien fréquemment et dans plusieurs matières. Cependant, la tendance à souscrire à des cours privés dépend aussi des caractéristiques familiales de l'élève. En effets, les enfants de chefs d'entreprise sont trois fois plus nombreux à suivre des cours privés que ceux d'ouvriers non qualifiés (pour les élèves considérés comme ayant des difficultés). Le niveau de vie de la famille révèle les inégalités entre les plus favorisés et les moins favorisés : les parents se situant parmi les 10 % de revenus les plus élevés sont deux fois plus nombreux à envoyer leur enfant en soutien privé que ceux se situant parmi les 10 % les plus faibles, toujours en considérant les élèves ayant des difficultés. Par ailleurs, les élèves dont les parents ne sont pas issus de l'immigration ou qui vivent en milieu rural sont moins nombreux à suivre du soutien privé comparé aux autres familles.

De plus, Caille mène une analyse *toutes choses égales par ailleurs* afin de déterminer l'impact de chacune des caractéristiques de l'élève sur la probabilité d'avoir recours au soutien privé. Il montre alors qu'un élève jugé "bon" par ses parents a une probabilité moins forte (-6,5 points) de suivre du soutien privé qu'un élève jugé en difficulté. Les élèves dont les parents ont une profession libérale, sont chefs d'entreprise, artisans ou commerçants, ont une probabilité plus élevée de suivre des cours privés par rapport à un ouvrier qualifié. Cependant, le niveau d'études des parents a beaucoup moins d'importance que les autres caractéristiques. Le niveau d'études du père a peu d'effet comparé à celui de la mère : avoir une mère ayant un diplôme supérieur ou égal au baccalauréat diminue la probabilité de suivre du soutien privé comparé à un niveau CAP/BEP. Ce dernier résultat peut traduire le niveau de capital humain de la famille : celles ayant un niveau d'éducation élevé sont plus à même de venir en aide à leurs enfants. La probabilité de suivre du soutien diminue lorsque le rang dans la fratrie augmente. Cependant, la taille de la famille a très peu d'impact : la probabilité augmente légèrement lorsque l'enfant n'a ni frère ni sœur. Les enfants d'immigrés ont plus tendance à suivre des cours de soutien que les enfants dont les parents sont nés en France. Enfin, les élèves vivant en milieu rural ont moins tendance à prendre des cours de soutien que les citadins. Le sexe de l'élève n'a pas d'effet sur la probabilité de suivre des cours privés.

B Le rôle de l'origine sociale

Les familles d'origines sociales favorisées ont plus facilement accès aux cours de soutien privés et se saisissent davantage de cette opportunité (Bray, 2011). Ce constat n'est pas surprenant et se retrouve dans la plupart des pays où des recherches existent. Ainsi, en Irlande, Smyth (2008, 2009) montre que la part d'élèves qui suit des cours privés est deux fois supérieure (près de 60 %) pour les enfants d'origine très favorisée⁷ par rapport aux enfants de parents peu qualifiés⁸. Les enfants inscrits dans des écoles payantes ont également plus souvent recours aux cours de soutien privés. Buchmann et al. (2010) montrent qu'aux États-Unis, les élèves issus de familles plus favorisées (en terme de revenu et de niveau d'éducation des parents) ont significativement plus de chances de suivre des cours privés pour se préparer aux concours d'entrée à l'université. Les élèves issus de minorités ethniques ont également une probabilité plus élevée de prendre des cours de soutien.

Le lien entre revenu, ou richesse, de la famille et soutien scolaire peut être difficile à identifier voir, s'il l'est, complètement faussé par les raisons qui poussent les familles à recourir au soutien scolaire. En effet, les enfants favorisés réussissant généralement mieux que les enfants d'origine défavorisée, la corrélation entre revenu et soutien apparaît négative si le but du soutien est de remédier aux difficultés des élèves. En revanche, à niveau scolaire donné, la corrélation est bien de signe positif⁹.

7. *Higher profession.*

8. *unskilled.*

9. C'est exactement ce que nous observons dans les données du panel 2007.

C Une responsabilité du système scolaire traditionnel

Lorsque les familles sont déçues par l'organisation du système éducatif¹⁰, elles peuvent choisir d'inscrire leurs enfants dans des établissements privés. Une autre solution consiste à compléter l'enseignement public par celui des cours privés de soutien afin de compenser, à moindre coût, les lacunes du système (Davies, 2004).

Bray et Lykins (2012) montrent à partir de questionnaires que dans plusieurs pays d'Asie, certains élèves choisissent de suivre des cours de soutien privés à cause du manque d'efficacité de l'école traditionnelle. Par exemple, au Sri Lanka, des élèves interrogés ont répondu qu'ils suivaient des cours privés parce que l'école ne propose pas assez d'exercices, que le programme n'est pas entièrement couvert ou qu'ils ne comprennent pas totalement ce qui leur est enseigné. L'école traditionnelle ne répondant pas aux besoins des familles, ces dernières se tournent vers le soutien privé.

L'existence d'examens en fin de cursus ou d'entrée à des universités peut également pousser les élèves à souscrire à des cours de soutien afin d'améliorer leurs chances de réussite. Ce type d'examen pouvant être très compétitif, certains élèves souscrivent à des préparations intensives à ces épreuves dans de nombreux pays, et notamment au Japon (Ono, 2007) ou aux États-Unis (Buchmann et al., 2010 ; Zimmer et al., 2010). Ce phénomène est également très répandu en Turquie (Tansel et Bircan, 2006) et en Russie (Loyalka et Zakharov, 2014) pour l'entrée à l'université.

En lien avec les examens, les cursus scolaires sont, dans de nombreux cas, décrits comme surchargés, encourageant donc les élèves à approfondir ou terminer le programme à l'aide de cours supplémentaires (Bray, 2011, p.36).

D D'autres motivations non scolaires ?

Si les élèves sont les premiers concernés par leur réussite et leur accès à des universités prestigieuses, la décision de prendre des cours de soutien émane généralement des parents mais n'est pas nécessairement motivée par la seule réussite de leur enfant. Comme le notent Gauci et Wetz (2009) à Malte :

"Les parents envoient leurs enfants dans des cours privés de façon à pouvoir se dire qu'ils font le maximum pour les aider. Ainsi, il y a probablement des élèves qui suivent des cours de soutien privés alors qu'ils n'en ont pas vraiment besoin."

Les cours privés de soutien peuvent également être un moyen de faire garder "productivement" les enfants lorsque l'emploi du temps des parents est particulièrement chargé. Ainsi, Haag et Van Kessel (2000) observent qu'au Luxembourg, les cours de soutien privés sont particulièrement attractifs pour les parents qui n'ont pas assez de temps pour s'occuper de leurs enfants durant la journée. Glasman (2010) va même plus loin en faisant l'hypothèse que le soutien privé peut être un moyen "d'acheter la paix à la maison" en externalisant l'aide aux devoirs des adolescents.

Enfin, il se peut aussi que les cours de soutien apparaissent en substitut d'un soutien familial plus traditionnel (en vivant plus éloigné de ses grands parents, par exemple) (Bray, 2011).

10. entendu ici comme public, fourni par l'État.

2 Des caractéristiques socio-démographiques identifiables

À partir des données du Panel 2007, nous disposons d'un large ensemble de caractéristiques des élèves, de leur famille et de leur environnement qui peuvent nous permettre de conforter les hypothèses présentées précédemment. Pour ce faire, nous modélisons l'impact de ces différentes caractéristiques sur la probabilité de suivre des cours de soutien privé en 6^e et en 4^e ou 3^e à l'aide de régressions logistiques. Les modèles sont très semblables à ceux de Caille (2010) que nous avons présentés dans l'encadré de la section précédente. Les résultats sont présentés dans le tableau 2. L'Annexe C décrit la méthode et les indicateurs présentés.

Tableau 2 – Effet des caractéristiques observables sur la probabilité de prendre des cours de soutien payant en 2008 et 2011

Variable	Modalités actives	(2008)		(2011)	
		$e^{\hat{\beta}}$	(\hat{SE})	$e^{\hat{\beta}}$	(\hat{SE})
Sexe	Fille	0.901**	(0.045)	0.997	(0.052)
Caractéristiques de la ville :	Rurale	0.680***	(0.085)	0.836	(0.112)
Ref : Commune de 50 000 à 100 000 hab.	≤ 5 000 hab.	0.710***	(0.084)	0.708***	(0.093)
	5 à 10 000 hab.	0.812*	(0.096)	1.120	(0.142)
	10 à 20 000 hab.	0.849	(0.102)	0.924	(0.124)
	20 à 50 000 hab.	0.855	(0.100)	1.060	(0.135)
	100 à 200 000 hab.	1.061	(0.127)	1.668***	(0.213)
	200 000 à 2 million hab.	1.104	(0.105)	1.396***	(0.150)
	Paris	1.414***	(0.139)	1.836***	(0.205)
Éducation de la mère	Inconnue	1.617***	(0.196)	0.967	(0.139)
Ref. Bac	Au plus primaire	0.864	(0.089)	0.743**	(0.094)
	Collège	1.061	(0.097)	0.809**	(0.083)
	CAP BEP	1.124	(0.085)	0.866*	(0.068)
	Sup. court	0.912	(0.082)	1.095	(0.087)
	Sup. long	1.147	(0.118)	1.053	(0.096)
Éducation du père	Inconnue	1.015	(0.107)	0.952	(0.112)
Ref. Bac	Au plus primaire	1.046	(0.112)	0.898	(0.113)
	Collège	0.937	(0.099)	0.994	(0.105)
	CAP BEP	0.889	(0.077)	0.905	(0.076)
	Sup. court	0.834	(0.094)	0.949	(0.094)
	Sup. long	1.100	(0.123)	1.020	(0.103)
Origine géographique des parents¹¹	Mixte française	1.260***	(0.091)	1.171**	(0.092)
Ref : deux parents nés en France	Mixte étrangère	2.055***	(0.314)	1.759***	(0.331)
	Europe	1.262	(0.251)	0.951	(0.229)
	Moyen Orient	1.793***	(0.327)	0.927	(0.244)
	Maghreb	2.167***	(0.261)	1.566***	(0.238)

Suite page suivante

<i>Suite de la page précédente</i>					
	Reste Afrique	1.248	(0.241)	1.300	(0.292)
	Amérique	1.846*	(0.654)	0.690	(0.422)
	Asie	2.834***	(0.605)	1.266	(0.396)
Secteur établissement	Privé	1.220***	(0.076)	1.478***	(0.088)
Statut boursier	Non boursier	1.079	(0.072)	1.170*	(0.099)
Redoublement	À déjà redoublé	1.147**	(0.075)	0.804***	(0.062)
Educ. prioritaire	ZUS	1.121	(0.096)	1.191	(0.136)
	RRSS	1.265***	(0.108)	1.202*	(0.119)
PCS chef de famille Ref : employé(e)	Agriculteur	0.943	(0.190)	1.030	(0.207)
	Commerçant, artisan, etc.	1.906***	(0.166)	1.873***	(0.186)
	Cadres et prof. libérales	1.419***	(0.140)	1.551***	(0.159)
	Prof intermédiaires	1.022	(0.091)	1.100	(0.107)
	Ouvriers	0.880*	(0.067)	0.932	(0.085)
	Chômeurs et inactifs	0.810	(0.200)	0.193	(0.203)
	Instits, profs et scientifiques	0.991	(0.190)	0.950	(0.176)
taille de la fratrie Ref : enfant unique	un(e) frère/sœur	0.767***	(0.065)	0.730**	(0.093)
	deux frères et sœurs	0.673***	(0.059)	0.615***	(0.080)
	3 frères et sœurs	0.663***	(0.068)	0.471***	(0.068)
	4 frères et sœurs	0.740**	(0.093)	0.546***	(0.090)
	5 ou plus	0.545***	(0.075)	0.443***	(0.080)
Structure familiale Ref : 2 parents	Alternée ou 1 parent	0.974	(0.078)	1.061	(0.085)
	Recomposée	0.865	(0.083)	0.942	(0.097)
	Autre	1.069	(0.235)	1.221	(0.255)
Avis sur l'établissement Ref : tout à fait satisfait	Plutôt satisfait	1.186***	(0.066)	1.269***	(0.075)
	Pas du tout satisfait	1.178	(0.147)	1.478***	(0.197)
Avis sur l'élève Ref : excellent élève	Grosses difficultés	5.347***	(0.798)	5.395***	(0.836)
	Un peu en difficulté	4.509***	(0.591)	5.273***	(0.702)
	Bon élève	1.814***	(0.225)	2.607***	(0.332)
Avis sur les enseignants Ref : plutôt pas satisfait	Tout à fait satisfait	0.845*	(0.082)	0.679***	(0.072)
	Plutôt satisfait	0.802***	(0.064)	0.768***	(0.059)
	Pas du tout satisfait	1.077	(0.210)	1.182	(0.238)
Aide d'un proche Ref : jamais ou presque	De temps en temps	0.827**	(0.073)	0.972	(0.064)
	Régulièrement	1.042	(0.086)	1.471***	(0.097)
Orientation envisagée par les parents Ref : activité pro jeune	Bac pro.	1.257**	(0.140)	1.716***	(0.187)
	Bac général ou techno	1.255**	(0.118)	2.689***	(0.304)
	Bac L ou ES	1.805***	(0.296)	3.473***	(0.434)

Suite page suivante

Suite de la page précédente

	Bac S	1.349**	(0.158)	2.394***	(0.313)
	Ne sait pas	1.147	(0.096)	1.630***	(0.263)
Performance en 6^e ref bottom 20%	4 ^e quintile	0.738***	(0.049)	0.834**	(0.070)
	3 ^e quintile	0.568***	(0.045)	0.707***	(0.063)
	2 ^e me quintile	0.413***	(0.037)	0.534***	(0.050)
	top 20%	0.307***	(0.035)	0.394***	(0.043)
Situation d'emploi de la mère Ref : active	Au chômage	0.689	(0.235)	0.866	(0.107)
	En formation	0.995	(0.512)	1.317	(0.501)
	Inactive	0.841	(0.208)	0.830**	(0.072)
Situation d'emploi du père Ref : active	Au chômage	1.172	(0.392)	0.666***	(0.099)
	En formation	1.257	(0.725)	3.659*	(2.558)
	Inactive	0.899	(0.217)	0.918	(0.123)
Constant		0.043***	(0.011)	0.022***	(0.007)
Pseudo R^2		0.110		0.120	
Observations		23661		15722	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses.

Sources : Panel d'élèves entrés en 6^e en 2007, Odds ratios obtenus à partir de régressions logistiques non pondérées.

Les modèles ont été construits en suivant la méthode de *forward regression* en excluant au final les variables ayant trop de réponses manquantes (e.g. type d'emploi des parents, heure de retour du travail le soir).

A Le soutien privé : plutôt pour les élèves faibles

Pour les élèves entrés en 6^e en 2007, le niveau de performance en mathématiques à l'entrée de la 6^e joue un rôle important sur la probabilité de prendre des cours de soutien privés qui persiste en 2010. En 2008, un élève qui figure parmi les 20 % les meilleurs en mathématiques à l'entrée en 6^e a une probabilité relative¹² 3 fois plus faible de prendre des cours privés de soutien qu'un élève parmi les 20 % les moins bons, toutes choses égales par ailleurs. Un élève se situant autour de la moyenne (entre les 40 et 60 %) a, quant à lui, 75 % de chance en moins de prendre des cours, à autres caractéristiques données. En 2011, l'impact des performances initiales perdure même si son effet est un peu atténué. **Un élève plutôt faible en mathématiques à l'entrée en 6^e a donc une probabilité plus élevée de prendre des cours de soutien en 6^e et 4 ans plus tard.**

Cette probabilité est, davantage encore, affectée par le jugement des parents sur le niveau de leur enfant. En effet, à caractéristique donnée (et donc en particulier, à niveau en mathématiques constant), un élève dont les parents pensent qu'il a de grosses difficultés a une probabilité 5 fois plus élevée de prendre des cours qu'un élève dont les parents pensent qu'il est un bon élève. **La perception des parents du niveau**

11. Nous utilisons la variable indiquant le lieu de naissance des parents

12. Nous avons renversé le ratio de façon à rendre l'interprétation plus simple. Ainsi, l'Odd ratio d'un élève parmi les 20 % les meilleurs par rapport à un élève parmi les 20% les moins bons est de 0.307. Le ratio inverse vaut donc $\frac{1}{0.307} = 3.26$.

de l'enfant affecte fortement la probabilité de suivre des cours privés de soutien, surtout en 2011, année où la majorité de notre échantillon passe le brevet des collèges.

Les données que nous présentons indiquent que, au collège, les cours de soutien privés sont utilisés en priorité par des élèves plutôt en difficulté, à la fois au regard de leurs parents, mais aussi plus objectivement au regard de leur niveau de mathématiques à l'entrée en 6^e.

B Des familles plutôt favorisées, citadines et peu nombreuses

Lorsqu'on s'intéresse aux caractéristiques des parents, on constate tout d'abord que, à autres caractéristiques données, la probabilité de prendre des cours ne varie pas avec le niveau d'étude atteint par le père de l'enfant, et très peu avec celui de la mère, uniquement en 2011. On constate ainsi qu'un enfant de 4^e ou 3^e dont la mère a un niveau d'étude inférieur au bac (ou non renseigné dans l'enquête famille) a une probabilité plus faible de prendre des cours de soutien par rapport à un élève dont la mère a un niveau bac. Le fait d'avoir des parents qui ont fait des études supérieures n'a pas d'impact sur le recours au soutien privé lorsqu'on contrôle pour les autres caractéristiques. En revanche, la probabilité de prendre des cours varie significativement suivant certaines catégories socio-professionnelles de la personne de référence¹³. En effet, toutes choses égales par ailleurs, la probabilité de suivre des cours de soutien privés est environ 90 % plus élevée pour un enfant d'artisan, de commerçant, . . . , que pour un enfant d'employé(e), en 2008 comme en 2011. Celle d'un enfant de cadres est 42 % plus élevée en 2008 et 55 % plus élevée en 2011 que celle d'un enfant d'employé(e), toutes choses égales par ailleurs.

Ainsi, les enfants d'origines très favorisées ont une probabilité plus forte de suivre des cours privés de soutien. A contrario, les enfants dont les parents ont des revenus assez modestes ont une probabilité légèrement plus faible de prendre des cours privés de soutien en 2011 comme permet de l'approximer la variable "boursier".

Le statut d'emploi joue un rôle différent en 6^e en 4^e/3^e, ce qui peut également s'expliquer par l'aggravation de la crise économique en 2011¹⁴. Ainsi, un élève dont le père est au chômage en 2011 a une probabilité de suivre des cours privés de soutien de 50 % plus faible qu'un élève dont le père a un emploi. La probabilité de suivre des cours de soutien privés est 20 % plus faible pour les élèves dont la mère est inactive que pour ceux dont la mère est en emploi.

Par ailleurs, il n'y a pas de différence de recours au soutien privé, que l'enfant vive avec ses deux parents, avec un seul, dans une famille recomposée, . . . En revanche, la taille de la fratrie est fortement liée à la probabilité de suivre des cours privés de soutien. En 2008 comme en 2011, le recours au soutien est significativement plus élevé pour un enfant unique que pour une famille ayant plusieurs enfants¹⁵. Ce résultat se retrouve dans la littérature scientifique internationale. Plusieurs études sur les pays asiatiques montrent que moins la famille est nombreuse, plus les parents investissent dans l'éducation de leur enfant via le soutien privé (De Castro et De Guzman, 2010 ; Liu, 2012 ; Bray et al., 2014). La même tendance est retrouvée par Tansel et Bircan (2006) en Turquie. Il contredit toutefois partiellement les résultats de Caille

13. La personne de référence est définie selon la méthodologie de l'INSEE : si les responsables sont le père et la mère, la PCS de référence est celle du père, si l'élève vit avec un de ses deux parents, la PCS est celle du parent avec laquelle vit l'enfant. Enfin, si l'enfant vit avec une autre personne, c'est la PCS de cette personne qui est retenue.

14. Le taux de chômage au sens du BIT est passé de 7.1 % à 8.8 % entre le troisième trimestre de 2008 et le troisième trimestre de 2011 (insee.fr).

15. la définition utilisée ici inclut les demi-frères et demi-sœurs vivant avec l'élève.

(2010), qui trouve que le rang dans la fratrie importe davantage que sa taille la probabilité de prendre des cours de soutien.

La commune de résidence de l'élève joue également un rôle important sur la probabilité de suivre des cours. De façon synthétique, **les élèves vivant dans des communes rurales ou de moins de 5 000 habitants ont une probabilité plus faible de prendre des cours de soutien que les élèves vivant dans des villes moyennes (i.e. de 50 à 100 000 habitants), toutes choses égales par ailleurs**. A contrario, les élèves vivant à Paris¹⁶ ou dans des villes de plus de 100 000 habitants, ont une probabilité de suivre des cours privés de soutien plus élevée. Ce constat capture probablement un effet de contrainte de l'offre, moins abondante dans les petites villes et les campagnes. Il peut aussi refléter d'autres caractéristiques propres aux familles citadines.

C L'origine géographique des parents

La probabilité de suivre des cours privés de soutien varie significativement avec le pays de naissance des parents. On constate tout d'abord que les élèves dont au moins un des parents est né à l'étranger¹⁷ ont une probabilité de prendre des cours de soutien supérieure ou égale aux élèves dont les deux parents sont nés en France, aux deux dates considérées. Pour certaines origines, la probabilité de suivre des cours de soutien est particulièrement élevée par rapport aux élèves dont les parents sont nés en France. En 6^e, les enfants d'origine maghrébine ont une probabilité plus de deux fois plus élevée de prendre des cours que les enfants dont les parents sont nés en France, les enfants d'origine asiatique ont une probabilité 2.8 fois plus élevée, toutes choses égales par ailleurs. Les enfants d'origines mixtes ont également une probabilité de prendre des cours privés de soutien plus importante que les enfants dont les parents sont tous deux nés en France, particulièrement élevée pour les parents tous deux nés à l'étranger.

Ainsi, les enfants d'origine immigrée ont plus de chances (ou au moins autant) de suivre des cours privés de soutien que des enfants dont les deux parents sont nés en France. Ces résultats sont conformes à la littérature sur l'investissement éducatif des familles issues de l'immigration (Brinbaum, 2002 ; Brinbaum et Kieffer, 2009).

D Aspiration des familles, appréciation et investissement

Nous incluons plusieurs variables mesurant l'appréciation des parents sur l'enseignement reçu par leur enfant, notamment la satisfaction des parents vis-à-vis des enseignants et de l'établissement fréquenté. On voit ainsi que, à autres caractéristiques données (et notamment à avis sur le niveau de l'élève constant), plus les parents ont une opinion négative sur l'établissement ou les enseignants, plus la probabilité de prendre des cours est élevée par rapport à un enfant dont les parents ont un opinion très positive¹⁸. On interprète ce résultat comme une volonté de compenser les manques et faiblesses de l'offre générale d'éducation par du soutien privé, financé par les familles (Davies, 2004). Les élèves qui sont inscrits dans des établissements privés ont aussi une probabilité plus forte de suivre des cours privés de soutien que les élèves inscrits dans le

16. La modalité Paris est significatives en 2008 et 2011, les modalités grandes villes de plus de 100 000 hab. ne sont significatives qu'en 2011.

17. Nous distinguons les familles mixtes dont l'un des parents est né en France, de familles mixtes dont les deux parents sont nés à l'étranger.

18. La significativité n'est toutefois pas atteinte pour toutes les modalités à toutes les dates.

public ayant les mêmes caractéristiques. Cette probabilité est même beaucoup plus forte en 2011, indiquant une intensification de l'investissement dans l'éducation pour ces familles.

Cette hypothèse est corroborée par la fréquence de l'aide au travail scolaire accordée par un proche¹⁹. En 6^e, à autres caractéristiques données, un élève qui reçoit de temps en temps de l'aide à la maison a une probabilité 20 % plus élevée de bénéficier de cours privés qu'un élève qui n'est pas ou presque jamais aidé, indiquant une substitution du soutien privé à l'aide familiale. En 2011, le fait d'avoir une aide régulière à la maison augmente de 50 % la probabilité de suivre des cours privés payants par rapport à un élève qui n'est jamais aidé ou presque, toute chose égale par ailleurs. **Le soutien privé en 2011 (principalement en 3^e) est plus une intensification de l'investissement des familles, et moins un substitut.** Les attentes en matière de soutien scolaire en 4^e/3^e ne sont plus les mêmes qu'en 6^e et visent certainement à permettre une orientation vers les filières les plus valorisées, ce que nous suggèrent les résultats de l'effet de l'orientation envisagée par les parents pour l'élève sur la probabilité de prendre des cours.

En effet, en 2008, la probabilité de prendre des cours de soutien est plus élevée pour les parents qui souhaitent que leur enfant ait un bac plutôt que la possibilité d'entrer rapidement dans la vie active²⁰. Parmi ceux-ci, les élèves dont les parents souhaitent qu'ils passent un bac L ou ES plutôt qu'entrer dans la vie active ont la probabilité relative la plus élevée et ceux dont les parents veulent que leur enfant passe un bac professionnel ont la même probabilité relative que ceux dont les parents souhaitent que leur enfant ait un bac général ou technologique²¹, soit 25 % plus élevée environ.

En 2011, les estimations sont pratiquement deux fois plus élevées pour toutes les modalités de cette variable indiquant ainsi que l'orientation voulue est davantage déterminante en 2011 qu'en 2008. À titre d'illustration, la probabilité qu'un élève dont les parents souhaitent qu'il aille en terminal L ou ES est 3.5 fois plus élevée qu'un élève dont les parents souhaitent qu'il entre rapidement dans la vie active, toutes choses égales par ailleurs.

19. L'aide à la maison peut être accordée par sa mère, son père, un frère ou une sœur, un camarade de classe, une autre personne de la famille.

20. Nous avons regroupé ici les parents qui souhaitent que leur enfant entre directement dans la vie active à 16 ans, fasse un apprentissage ou un CAP.

21. Pour des raisons de taille d'effectif, nous avons groupé les parents qui souhaitent que leur enfant ait un bac général, peu importe la série et ceux qui veulent que leur enfant ait un bac technologique.

III L'effet du soutien privé sur les performances des élèves

Cette section analyse l'effet du soutien privé sur cinq mesures de performances des élèves en 6^e et en 4^e/3^e. Elle se divise en trois parties : la première présente les indicateurs mobilisés et une série de statistiques descriptives, la seconde propose plusieurs estimations des différences de performances entre les élèves qui prennent des cours de soutien en 6^e et détaille les méthodologies employées. La troisième reproduit ces estimations pour les cours de soutien privés en 4^e/3^e et discute davantage la validité des méthodes employées.

1 Les performances des élèves dans le panel 2007

A Les variables mobilisées dans l'analyse

Nous disposons de 5 mesures cognitives, comparables dans le temps et disponibles à deux dates pour un échantillon représentatif d'élèves entrés en 6^e en 2007. Nous présentons brièvement leurs composants dans l'encadré ci-dessous.

Les mesures de performance du Panel 2007 (MENESR-DEPP)

Trois évaluations standardisées ont été mises en place au cours des années de collèges des élèves du panel. Elles sont très bien décrites dans [Ben Ali et Vourc'h \(2015\)](#). Nous reprenons certaines de leurs explications dans ce qui suit.

Nous exploitons les résultats de 5 épreuves, les autres ne pouvant pas être comparées dans le temps. La variable **Lexique 6** est le résultat d'un test de *mémoire encyclopédique*. Il consiste à faire associer le mot "le plus proche" à l'item testé. Il s'appuie sur le contenu des manuels scolaires l'année concernée dans différentes disciplines telles que les mathématiques, les sciences, la technologie, l'histoire-géographie, les sciences de la vie et de la terre. C'est la seule variable dont le contenu est directement lié aux programmes de l'année en cours.

La variable **Phrases lacunaires** est une épreuve de compréhension basée sur des textes à trous. Elle vise à évaluer à la fois la mémoire sémantique et la richesse du lexique.

La variable **Maths** est une épreuve de 45 items répartis dans plusieurs champs : calcul mental, problème, calcul d'horaire et d'unité, géométrie et logique.

La variable **Compréhension** mesure les capacités en lecture silencieuse des élèves. Elle repose sur trois textes d'une cinquantaine de mots sur lesquels ils sont ensuite interrogés, en gardant le texte sous les yeux. Il s'agit "*d'un test de compréhension avec des questions nécessitant une inférence*" ([Ben Ali et Vourc'h, 2015, p.214](#)).

La variable **Raisonnement** mesure le raisonnement logique de l'élève sur des cartes de Chartier (RCC). L'élève doit identifier, à partir des caractéristiques d'une série de cartes (couleurs, valeurs) celle qui continue une suite proposée.

Ces épreuves visent à évaluer les capacités de raisonnement logique à partir de "situations problèmes" utilisant comme support des cartes à jouer. L'élève doit alors identifier les caractéristiques de la carte (couleur et valeur) pour compléter une série. La tâche est considérée comme relevant du facteur général d'intelligence (facteur g) et plus précisément de l'intelligence fluide (capacité générale à

établir des relations entre des éléments), considérée comme étant relativement indépendante des connaissances acquises (Chartier, 2012).

Ces variables sont standardisées de façon à ce qu'en 2008, chaque score soit de moyenne nulle et d'écart type 1 et comparable dans le temps.

L'étude de Ben Ali et Vourc'h (2015) analyse l'effet de nombreuses caractéristiques individuelles, du contexte scolaire, du contexte social et familial, de l'environnement culturel et matériel sur ces mêmes variables. Dans la mesure où nous construisons des modèles semblables dans la suite de ce travail, il ne nous semble pas utile de présenter l'effet *toutes choses égales par ailleurs* de ces variables sur les performances. Nous ne faisons que commenter certains des résultats de Ben Ali et Vourc'h (2015) que nous trouvons particulièrement intéressants et invitons le lecteur intéressé à s'y reporter directement pour une analyse complète.

Si plusieurs caractéristiques et variables de contexte influencent significativement les performances et leurs progressions, la dimension qui explique le plus leur évolution est le niveau à l'entrée en 6^e. Ce résultat semble indiquer que **l'avenir scolaire d'un élève est déjà en grande partie déterminé lorsqu'il rentre dans le second degré**. La progression dans certaines compétences n'est d'ailleurs pas du tout affectée par les caractéristiques de l'élève et de son environnement ; notamment les compétences en raisonnement logique et en compréhension (traitement de phrase lacunaire). À partir de la 6^e, la progression dans ces dimensions n'est plus significativement impactée par l'origine sociale de l'élève. En revanche, l'évolution en mathématiques et mémoire encyclopédique est affectée par le niveau d'éducation et la position sociale du responsable de l'élève. Plus il possède un niveau de formation bas, moins la progression de l'élève est rapide dans ces domaines. De même, l'indice social impacte positivement la progression indiquant un accroissement des inégalités scolaires d'origine sociale au collège, pour ces dimensions scolaires.

B Évolution des performances et soutien scolaire

La figure 1 présente les distributions des scores en 2008 et 2011, selon que l'élève prend des cours de soutien ou non.

La distribution des scores des élèves qui prennent des cours de soutien en 2008 est globalement plus basse que celle de ceux qui n'en prennent pas. Cependant, les deux groupes progressent *a priori* de façon comparable entre les deux dates, dans toutes les disciplines. Ce résultat semble indiquer que les élèves qui prennent des cours sont plutôt moins bons au départ, mais progressent au même rythme que ceux qui n'en prennent pas. En 2011, le premier quartile²² des élèves qui ne suivent pas de cours privés de soutien se situe au niveau de celui de l'élève médian²³ en 2008. Pour les élèves qui suivent des cours particuliers, le premier quartile est toujours plus élevé que le niveau de l'élève médian en 2008.

Il faut toutefois garder à l'esprit le fait que la majorité des élèves qui prenaient des cours en 2008 n'en prennent plus en 2011 et vice-versa. Ainsi, les distributions que nous comparons ici ne concernent pas les mêmes élèves. La nature longitudinale des données dont nous disposons nous permet de distinguer l'évolution des performances moyenne des élèves dans les différentes dimensions testées suivant le moment

22. Niveau atteint par les 25 % d'élèves les plus faibles représenté par le bas de la boîte à moustache.

23. Élève dont le niveau, au milieu de la distribution du score, est représenté par le trait au milieu de la boîte à moustache.

où ils prennent des cours de soutien scolaire. Les résultats sont présentés dans la figure 2.

On constate alors immédiatement que les élèves qui ne prennent pas de cours en 6^e ou en 4^e/3^e ont des performances nettement supérieures aux élèves qui en prennent. Viennent ensuite les élèves qui ne prennent pas de cours en 2008 mais en prennent en 2011. Ceux-ci ont initialement des performances légèrement plus faibles que ceux qui ne prennent pas de cours en 2008, statistiquement identiques en connaissance encyclopédique (lexique 6) et en compréhension de texte (phrase lacunaire). Pour les trois autres dimensions, ces élèves sont, en moyenne, plus faibles de 10 % d'un écart-type²⁴. Cependant, leur progression semble moins rapide que celle des élèves qui ne prennent pas de cours, comme si, en la sentant ralentir, ils souscrivaient à des cours payants. Les élèves qui prennent des cours aux deux dates sont nettement moins bons en 2008 et progressent visiblement moins vite que les deux groupes précédemment décrits. Ils sont cependant meilleurs que les élèves qui prennent des cours de soutien en 2008 mais n'en prennent plus ensuite. Il est difficile d'interpréter la différence entre ces deux groupes. D'un côté, on peut penser que le groupe qui ne prend plus de cours en 2011 n'a pas bénéficié de ceux de 2008 et c'est pourquoi ces élèves abandonnent. D'un autre côté, on peut se dire que ceux qui continuent étant meilleurs, ils ont bénéficié de ces cours et c'est pourquoi ils continuent.

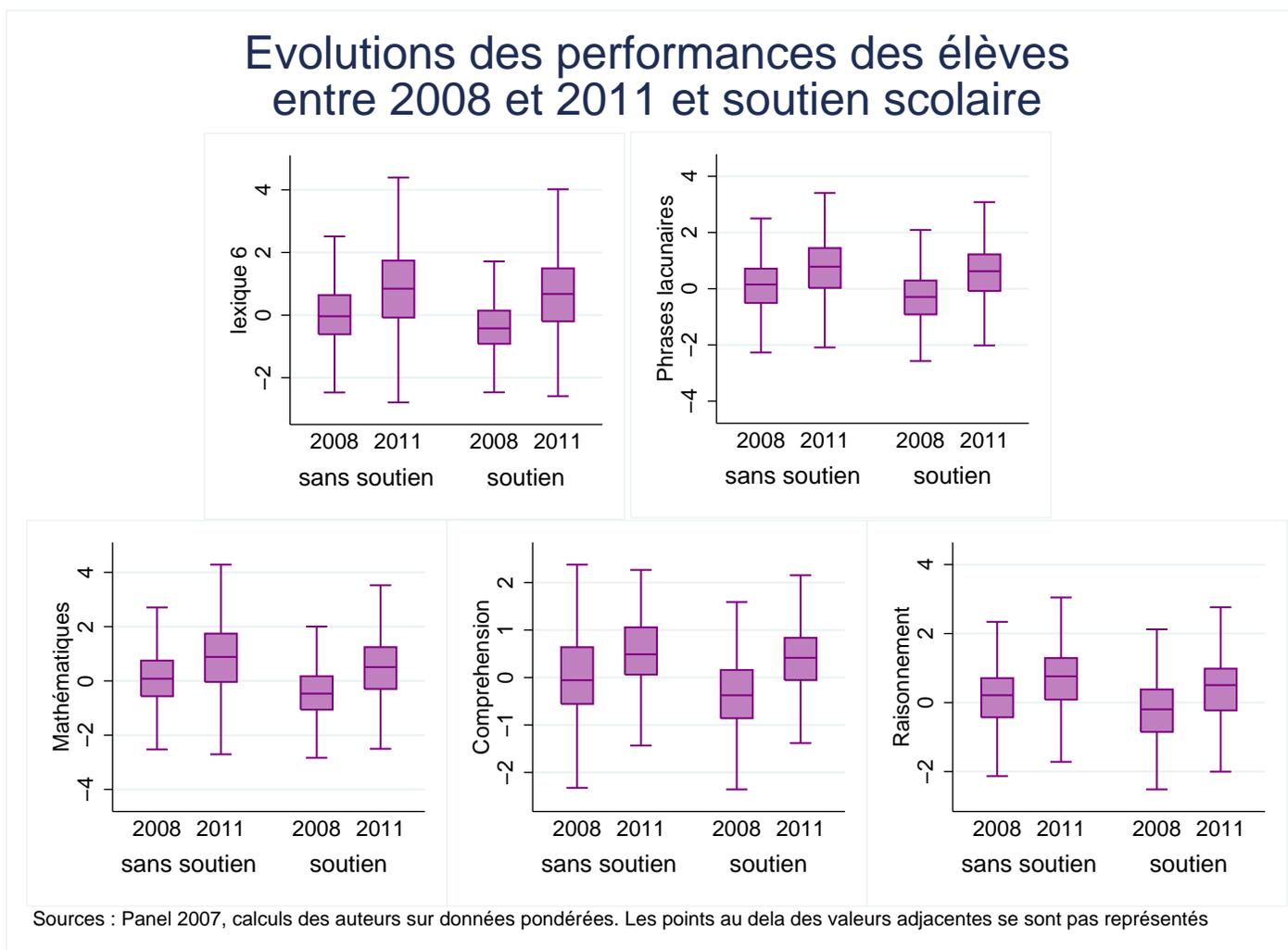
Le recours au soutien scolaire à différentes dates, ou son abandon, est le signe d'un processus de sélection dynamique difficilement explicable à partir des informations dont nous disposons. On ne peut toutefois pas négliger ce processus et considérer qu'il est le fruit du hasard. Pour en tenir compte, nous sommes contraints de faire certaines hypothèses et de nous restreindre à des analyses année par année, contrôlant par le niveau passé. En effet, il eut été tentant de profiter de la dimension panel de nos données et d'estimer l'effet des cours privés en utilisant un *modèle à effets fixes*. Cette méthode, qui a l'avantage de neutraliser l'effet des caractéristiques inobservables mais stables dans le temps, fait l'hypothèse que les modifications du choix de recourir à des cours privés ne sont pas liées à l'erreur du modèle, alors que ce sont peut-être précisément des chocs sur le potentiel²⁵ qui conduisent à modifier la stratégie de recours au soutien privé.

Nous analysons successivement l'effet des cours de soutien en 6^e sur les performances puis ceux en 4^e/3^e et proposons plusieurs modélisations pour tenir compte de cette sélection dynamique.

24. Comme les variables sont standardisées de façon à avoir une moyenne nulle et un écart-type de 1 en 2008, les différences s'interprètent en pourcentage d'écart-type.

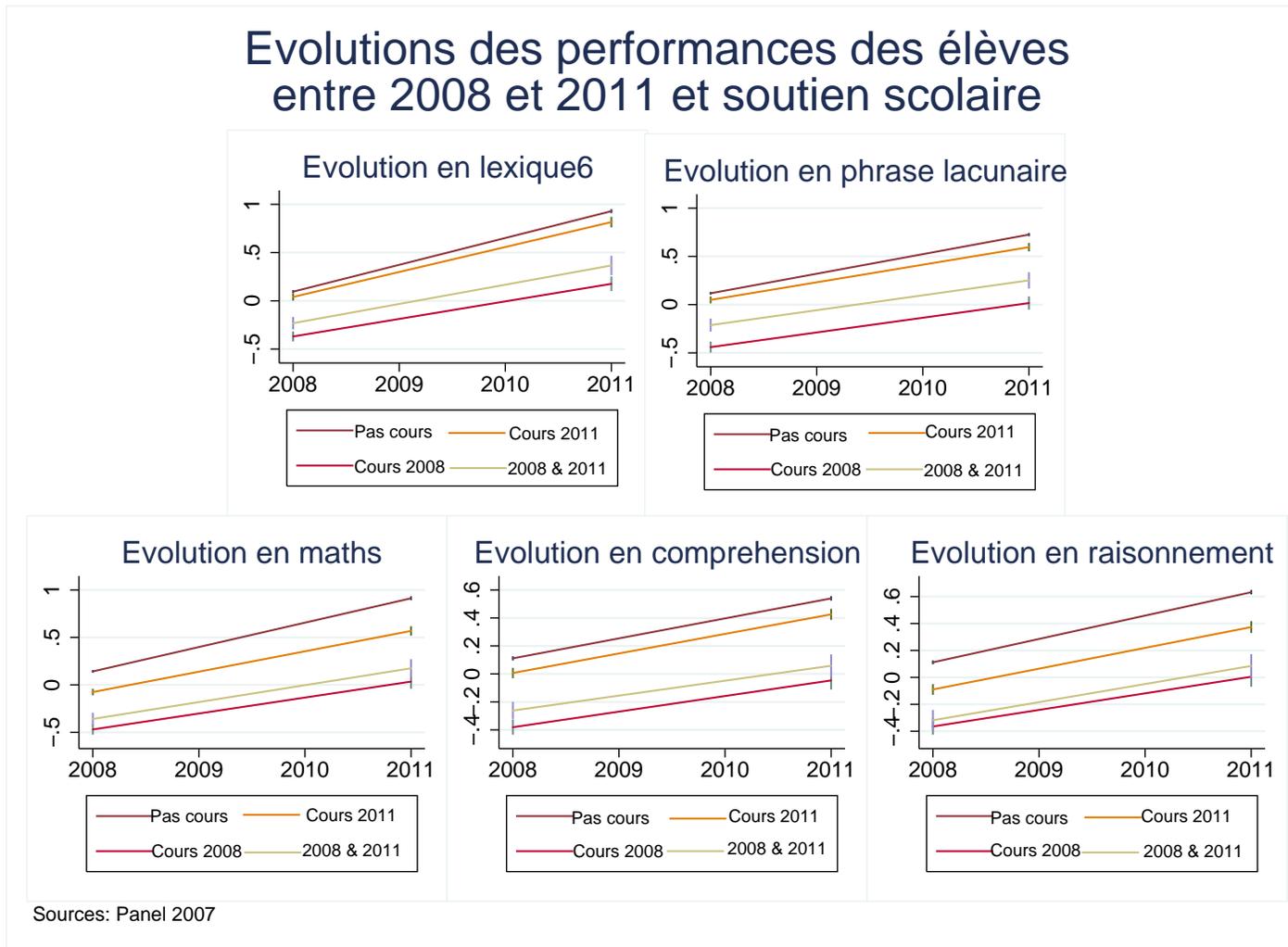
25. Changement dans la capacité à progresser de l'élève, chocs extérieurs qui affectent durablement ses performances, divorce, etc.

Figure 1 – Distributions des performances entre 2008 et 2011 suivant la variable soutien scolaire



Lecture : Les graphiques présentent la distribution des scores dans les différentes dimensions testées en 2008 et 2011 pour les élèves qui, à ces dates, suivent des cours de soutien. Les boîtes représentent l'intervalle interquartile ; la barre du milieu, la médiane. Les "pattes" s'étendent jusqu'à 1.5 fois l'intervalle interquartile.

Figure 2 – Évolution des performances moyennes des élèves entre 2008 et 2011 suivant le moment où ils suivent des cours de soutien privés



Lecture : Les graphiques présentent la progression moyenne des élèves dans les différentes dimensions cognitives testées en 2008 et 2011 pour les quatre groupes d'élèves identifiés : ceux qui prennent des cours en 2008, mais pas en 2011, ceux qui en prennent en 2011 mais pas en 2008, ceux qui n'en prennent ni en 2008 ni en 2011 et ceux qui n'en prennent pas à ces dates.

2 Le soutien privé en 6^e

Dans cette section, nous analysons d'abord la différence moyenne de performances entre les élèves qui prennent des cours privés de soutien et les autres, en éliminant les différences dues aux caractéristiques socio-démographiques de l'élève et à celles de son environnement. Dans un second temps, nous nous intéressons à l'hétérogénéité entre les élèves de chaque groupe en fonction de leurs performances. Enfin, comme la souscription à des cours privés de soutien dépend de facteurs qui sont pour nous inobservables, la comparaison que nous réalisons ici, bien que très informative, ne nous permet pas d'identifier de relation causale entre le soutien privé et les mesures de performance. Pour y parvenir, nous présentons un modèle exploitant la contrainte de l'offre de soutien privé.

A Les écarts moyens de performance

a) Modélisation

Dans la section précédente, nous avons décrit les distributions des performances des élèves suivant qu'ils prennent des cours de soutien ou non. Ici, on compare ces performances en moyenne, d'abord de façon *brute* - sans tenir compte des différences entre les élèves - puis en contrôlant pour un ensemble large de caractéristiques observables. Les méthodes que nous proposons n'imposent pas de restriction sur la manière dont les caractéristiques agissent sur les performances et le soutien scolaire.

Plus précisément, nous présentons trois indicateurs pour chaque dimension testée :

1. la différence moyenne *brute* entre les élèves qui prennent des cours privés de soutien et les autres ;
2. la *différence conditionnelle moyenne* de performances obtenue par des régressions des performances sur un ensemble de caractéristiques estimées séparément pour chaque groupe et pondérées par l'inverse de la probabilité de prendre des cours de soutien (méthode 1). Cette dernière est estimée par une régression logistique de la variable soutien sur les mêmes caractéristiques²⁶.
3. la *différence conditionnelle moyenne* de performance obtenue en appariant les élèves qui prennent des cours aux élèves qui n'en prennent pas mais dont la probabilité estimée de suivre des cours est la plus proche²⁷ (méthode 2).

Pour réaliser ces estimations, nous ne retenons pas toutes les variables utilisées dans les régressions logistiques du tableau 2, car certaines sont susceptibles d'être de mauvaises variables de contrôle pour la question qui nous intéresse (Angrist et Pischke, 2009, sec 3.2.3). En effet, certaines peuvent être partiellement déterminées par des caractéristiques inobservables ou encore être davantage une conséquence qu'un déterminant du soutien scolaire. Nos modèles incluent donc : le genre de l'élève, l'éducation de ses parents, leur origine migratoire, une variable binaire si l'élève a déjà redoublé, la PCS du chef de famille, la structure de la famille, le nombre de frères et de sœurs de l'élève et la fréquence de l'aide de proches. Nous contrôlons également davantage le niveau initial en utilisant deux polynômes d'ordre deux des tests d'entrée en 6^e en français et mathématiques. Enfin, nous ajoutons deux variables mesurant le niveau de

26. Cet estimateur est souvent appelé "*Wooldridge double-robust estimator*" (Wooldridge, 2007).

27. Méthode de matching sur score de propensité ; elle a notamment été appliquée à cette question par Thongphat (2012) sur des données thaïlandaises.

richesse de la commune. Nous prenons pour cela le logarithme du revenu fiscal moyen dans la commune de l'établissement et le logarithme de la somme des revenus fiscaux des foyers de la commune²⁸.

En contrôlant par le niveau initial des élèves, nous construisons un modèle dit à *valeur ajoutée*. Ce type de modélisation est fréquemment employé en économie de l'éducation, en particulier dans les recherches sur l'effet enseignant (Kim et Lalancette, 2013). Toutefois, l'utilisation des variables du niveau scolaire passé comme contrôle pose un problème non trivial dans l'estimation de ces modèles, que nous décrivons dans l'Annexe D. En effet, en raison du lien entre niveau de performance initial et soutien scolaire, les résultats de ces estimations sont biaisés par l'omission d'un paramètre inobservable. Cependant, comme le montre Bouguen (2015) - dont nous avons adapté la démonstration à notre problème - le biais qui affecte nos estimations est négatif. Les résultats que nous proposons ci-dessous sous-estiment donc le véritable effet des cours privés de soutien.

b) Résultats

Comme attendu, les élèves qui prennent des cours particuliers en 2008 ont des scores beaucoup plus faibles que les autres (entre -40 et -50 % d'un écart-type, suivant les mesures), mais cette différence est très largement expliquée par les caractéristiques de l'élève et de sa famille. L'estimation des différences conditionnelles moyennes aboutie à des écarts bien moindres. Toutes les estimations sont environ 4 fois plus faibles lorsqu'on tient compte des caractéristiques des élèves et de leur famille, soit environ 10 % d'un écart-type. Ces résultats ne mesurent pas parfaitement l'effet moyen des cours de soutien. Des caractéristiques que nous ne pouvons prendre en compte diminuent ces estimations, qui sous-évaluent le véritable effet. Aussi, l'impact des cours de soutien a de grandes chances d'être nul ou positif, mais probablement pas négatif. En somme, toutes choses égales par ailleurs, il n'y a qu'une faible différence de performances entre les élèves qui prennent des cours de soutien en 6^e et les autres.

Toutefois, cette différence moyenne peut dissimuler de l'hétérogénéité suivant le niveau de l'élève. On analyse ces différences par des régressions quantiles.

28. d'après INSEE : *Revenus imposables et montant des impôts*

Tableau 3 – Différence de performances en 2008 entre élèves suivant ou non des cours privés de soutien, avec ou sans variables de contrôle

Variable	Méthode	différence	(erreur type)	P-value
lexique 6	différence brute	-0,404***	(0,021)	0,000
	méthode 1	-0,056*	(0,032)	0,075
	méthode 2	-0,08*	(0,048)	0,095
Phrases lacunaires	différence brute	-0,465***	(0,023)	0,000
	méthode 1	-0,076***	(0,021)	0,000
	méthode 2	-0,092***	(0,033)	0,005
Maths	différence brute	-0,529***	(0,022)	0,000
	méthode 1	-0,098***	(0,023)	0,000
	méthode 2	-0,113***	(0,033)	0,001
Compréhension	différence brute	-0,431***	(0,022)	0,000
	méthode 1	-0,11***	(0,027)	0,000
	méthode 2	-0,117***	(0,034)	0,001
Raisonnement	différence brute	-0,43***	(0,025)	0,000
	méthode 1	-0,105***	(0,031)	0,001
	méthode 2	-0,125***	(0,036)	0,001

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses.

Lecture : Le groupe d'élèves qui prend des cours de soutien en 2008 a, en moyenne, des performances en lexique plus faible de 0,4 écart-type que le groupe qui n'en prend pas. Toutefois, lorsqu'on retire la part de l'écart due aux caractéristiques observables de l'élève et de sa famille, la différence de performances n'est plus que de 5.6 à 8 % d'un écart-type suivant la méthode, non significative au seuil de 5 %.

De façon plus précise, on estime la probabilité de suivre des cours en 2008 par une régression logistique. Dans la méthode 2, à chaque élève recevant des cours de soutien, on identifie l'élève qui a la probabilité de suivre des cours la plus proche parmi ceux qui n'en prennent pas pour constituer un groupe de contrôle. Pour la méthode 1, on estime des régressions séparées des performances sur les caractéristiques socio-démographiques des élèves pour les deux groupes, en pondérant les élèves par l'inverse de leur probabilité de prendre des cours. On estime ensuite la différence moyenne entre les prédictions des deux régressions.

B Hétérogénéité : différence de performances par quantile de niveau

Pour tester si la différence entre les élèves qui prennent ou non des cours dépend du niveau de l'élève, on analyse les écarts de performances à plusieurs endroits des distributions des scores à partir de régressions quantiles. La figure 3 représente ces différences pour chaque dimension considérée, par quantile, et en tenant compte des autres caractéristiques des élèves. Autrement dit, chaque point violet correspond à l'estimation de la différence de performances entre élèves ayant des résultats scolaires similaires (appartenant au même quantile sur la distribution des résultats), suivant ou non des cours de soutien privé.

La régression quantile et l'hétérogénéité de l'effet du soutien privé

Les régressions quantiles peuvent être vues comme une extension des modèles linéaires classiques d'espérance conditionnelle, comme les moindres carrés, à des modèles de médiane conditionnelle où d'autres quantiles. L'un des avantages les plus évident est que contrairement aux moindres carrés, les régressions quantiles sont beaucoup moins sensibles aux valeurs extrêmes. Mais leurs avantages ne s'arrêtent pas là. En effet, le principal intérêt des régressions quantiles - dans le contexte des cours de soutien - est de comparer **la distribution** des scores des élèves qui prennent des cours avec ceux de ceux qui n'en prennent pas. Elles comparent ainsi les valeurs correspondantes à chaque quantile, c'est-à-dire, de performance pour lesquelles un pourcentage q d'élève se situe en dessous et $1-q$ au dessus dans les deux groupes. De façon très schématique, si les coefficients des régressions quantiles sont négatifs au début de la distribution (par exemple au premier décile), cela signifie que les 10 % des élèves les plus faibles qui prennent des cours de soutien sont moins performants que les 10 % des élèves les plus faibles qui n'en prennent pas. Dès lors, la distribution des scores est plus *étalée* pour les élèves qui prennent des cours, donc plus inégale.

L'interprétation des coefficients demande toutefois d'être précautionneux sur le vocabulaire choisi. Le fait de trouver un coefficient négatif pour le premier décile n'implique pas qu'un élève peu performant aurait été meilleur s'il n'avait pas pris de cours. **Cela signifie que les élèves peu performants parmi ceux qui prennent des cours sont moins bons que les moins performants de ceux qui ne prennent pas de cours.** Car rien n'indique que le fait de prendre des cours ne modifie pas le rang des élèves dans la distribution des performances.

Ces estimations sont à comparer avec les résultats des régressions OLS standards qui donnent l'espérance conditionnelle de Y en X : $Y_i = \alpha + \beta T_i + \mathbf{X}_i' \rho + \varepsilon_i$. On représente graphiquement la pente $\hat{\beta}_q$ pour chaque quantile q et l'intervalle de confiance associé (les erreurs standards sont obtenues après 50 bootstrapping) ainsi que le coefficient OLS correspondant (Cameron et Trivedi, 2005, p.89).

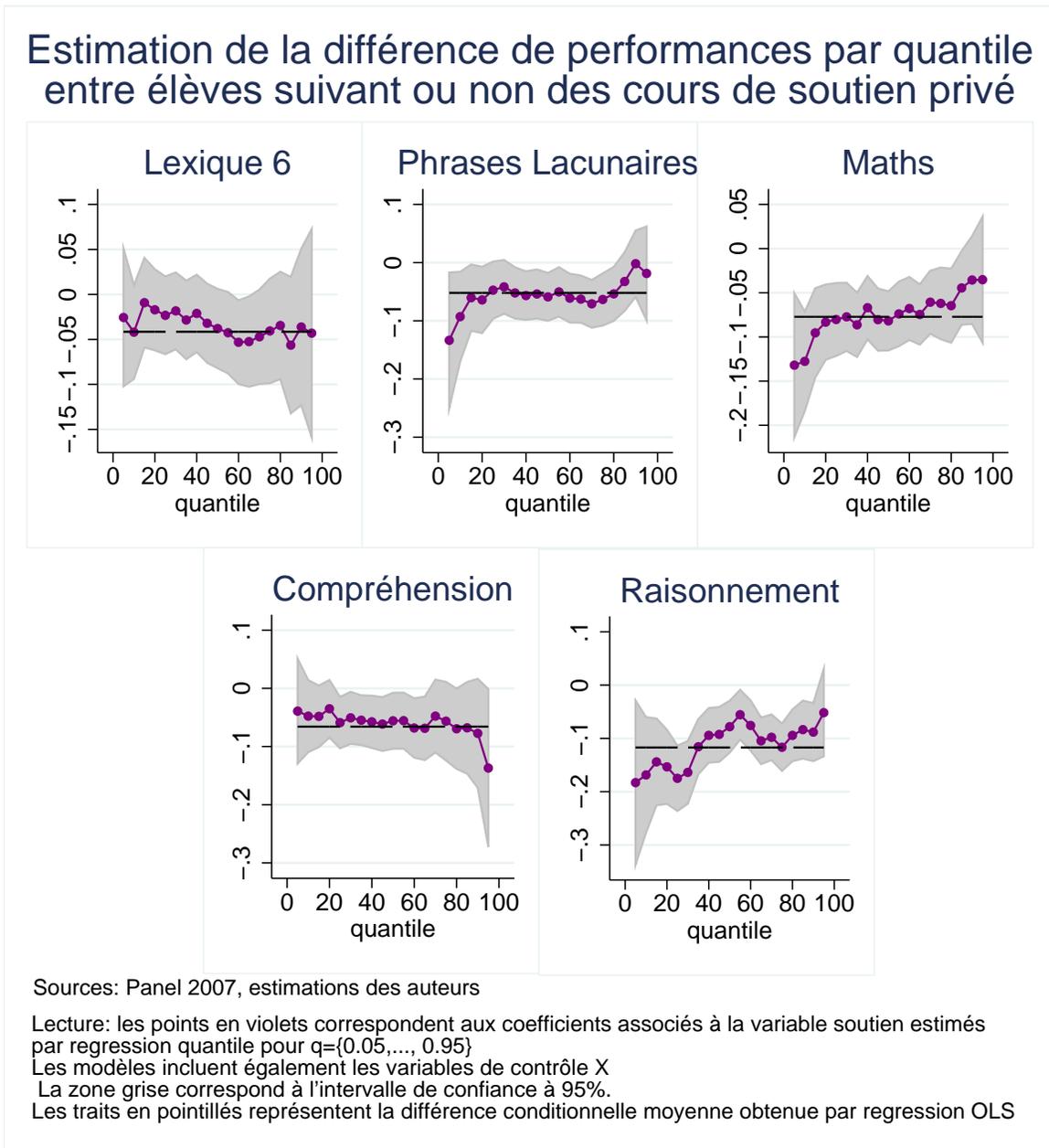
De façon assez surprenante, l'écart entre les élèves qui prennent des cours et ceux qui n'en prennent pas est assez homogène, et le peu de variation que l'on observe est souvent mince et rarement significativement différent de 0. Deux domaines se distinguent cependant clairement : les mathématiques et le raisonnement.

En effet, la différence de résultats scolaires entre les élèves qui prennent des cours de soutien et ceux qui n'en prennent pas, conditionnellement aux caractéristiques socio-démographiques du modèle²⁹, est négative mais elle est d'autant moins élevée que l'élève est bon en mathématiques. Pour les 10 % d'élèves les plus performants, il n'y a plus de différence imputable au soutien privé. La différence moyenne obtenue par régression dans la section précédente³⁰ dissimule donc de l'hétérogénéité en fonction du niveau de l'élève. Les performances en mathématiques des élèves faibles qui prennent des cours de soutien sont donc inférieures à celles des élèves qui n'en prennent pas. En abusant un peu dans le langage de l'inférence causale, on pourrait dire que l'introduction du soutien privé *rend* plus inégale la distribution des scores en mathématiques en 6^e.

29. identiques aux modèles de la section précédente.

30. représentée ici par les pointillés noirs.

Figure 3 – Différence conditionnelle de performances due au soutien scolaire en 2008, par quantile.



On fait le même constat pour la dimension raisonnement même si la forme des estimations est difficilement interprétable. Il semble que la différence conditionnelle de performances en raisonnement entre les élèves qui prennent des cours et ceux qui n'en prennent pas soit globalement négative, mais plus faible pour les élèves médians et les 10 % des élèves les meilleurs.

Les autres domaines : phrase lacunaire, lexique 6 et compréhension sont assez linéaires à tout niveau de la distribution et souvent très proche de 0.

Il apparait ainsi que les compétences en mathématiques et en raisonnement sont assez différentes des autres mesures cognitives dont nous disposons. Ce qui n'est que peu étonnant. Les mathématiques sont souvent beaucoup travaillées dans les cours privés et font parfois appel aux capacités de raisonnement.

À ce stade, on ne peut pas dire que les cours de soutien ont, ou n'ont pas d'effet sur les performances des élèves. On sait simplement que les caractéristiques socio-démographiques des élèves neutralisent une grande partie des différences de performances entre les élèves soutenus et les autres, et que la distribution des scores de mathématiques et raisonnement est plus inégale avec du soutien privé.

Mais la décision de prendre ou non des cours privés dépend probablement d'autres caractéristiques, inobservables dans cette base, qui biaisent la comparaison. C'est pourquoi nous proposons une stratégie visant à neutraliser ce biais de sélection.

C Une estimation par variable instrumentale

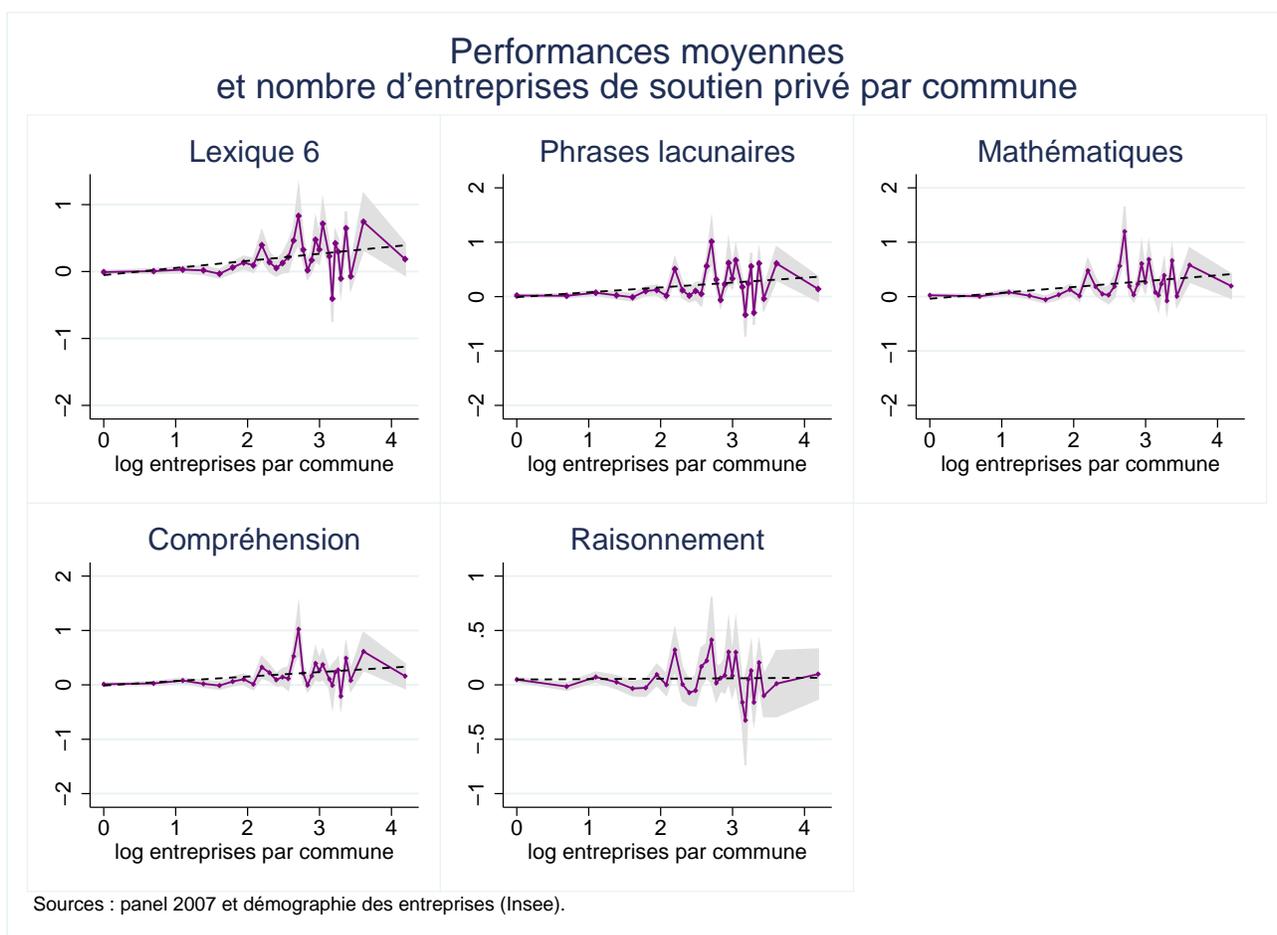
a) Données et intuitions

Si les parents continuent de payer des cours de soutien à leur enfant, c'est qu'ils attendent - et probablement observent - des bénéfices en terme de performances ou de trajectoires scolaires. Puisque le soutien concerne majoritairement des élèves déjà en difficulté, il est probable que les comparaisons précédentes, même à caractéristiques socio-démographiques constantes, sous-estiment l'effet des cours privés sur les performances. Pour mesurer *l'impact causal* des cours de soutien, on doit pouvoir identifier une valeur *contrefactuelle*, c'est-à-dire, une estimation de ce qui se serait passé pour un élève s'il n'avait pas pris de cours de soutien. Or il serait naïf de simplement substituer les performances des élèves qui ne prennent pas de cours pour estimer cette valeur. Car même en appariant les élèves de chaque groupe en fonction de leurs caractéristiques observables, on s'attend à ce que ces élèves diffèrent malgré tout, en raison de caractéristiques inobservables qui affectent à la fois les performances et la probabilité de prendre des cours. C'est pourquoi, pour identifier cette relation causale, on propose une série d'estimations par variable instrumentale.

Notre stratégie d'identification est basée sur une idée très simple : la présence en plus ou moins grand nombre ou l'absence d'établissements offrant des cours privés les rendent plus ou moins accessibles et donc affectent la probabilité pour un élève de suivre des cours. On observe ainsi que les enfants qui, un peu par hasard, se trouvent dans des communes où l'offre est abondante, ont plus de chance de prendre des cours. Si les cours privés ont des effets sur les progressions, à niveau de départ donné, les élèves qui se trouvent dans des communes bien dotées devraient progresser davantage puisqu'une plus grande proportion d'entre eux bénéficient de cours. La figure 4 l'illustre en représentant graphiquement les corrélations entre le nombre d'entreprises de soutien scolaire dans une commune et les performances moyennes des élèves scolarisés dans cette commune. Dans la plupart des dimensions, on constate que plus il y a d'entreprises, plus les élèves sont bons, en moyenne. Les estimations sont moins précises pour les villes ayant beaucoup d'entreprises car il y a peu de commune dans cette situation.

La présence en plus ou moins grand nombre d'entreprises de soutien privé facilite l'accès à ces services. Cependant, à caractéristiques individuelles et familiales données, il n'y a pas de raison pour que le nombre d'entreprises de soutien scolaire dans la commune affecte les performances scolaires autrement que par l'accès à ces cours. Sous réserve que cette hypothèse soit valide (ce que nous défendons plus loin), cette variable constitue un instrument plausible. Tout se passe alors comme si nous comparions deux élèves, par ailleurs identiques, dont l'un peut souscrire facilement à des cours de soutien et l'autre avec plus de mal,

Figure 4 – Nombre d'entreprises de soutien privé et performances moyennes



Lecture : les lignes pourpres présentent les performances moyennes des élèves dans les 5 disciplines testées par rapport au nombre d'entreprises de soutien scolaire présentes dans la commune, exprimé en logarithme. Les droites pointillées représentent l'ajustement linéaire de cette relation. À l'exception de la variable raisonnement, la pente de cette droite est positive et significative, indiquant que plus il y a d'entreprises de soutien privé dans la commune, plus les performances moyennes des élèves de cette commune sont élevées.

parce-qu'il y a moins d'entreprises offrant ce service dans sa commune. L'instrument permet d'identifier une valeur contrefactuelle basée sur la contrainte locale d'accès aux cours. En réalisant cette comparaison sur l'ensemble de l'échantillon, nous identifions l'effet moyen des cours privés de soutien sur la performance des élèves. Les modèles que nous proposons supposent un effet constant. Les graphiques de la figure 4 représentent la *forme réduite* du modèle.

À partir des données INSEE, nous avons apparié le nombre d'entreprises proposant des cours de soutien scolaire dans la commune et le département où l'élève est scolarisé en 2008 et en 2011. Nous avons retenu le nombre d'entreprises qui, en 2008³¹, ont un numéro de SIREN et dont l'activité principale est :

- 85.60, 85.60Z : Activités de soutien à l'enseignement
- 85.52 et 85.52Z : Enseignement culturel
- 85.59 : Enseignements divers c'est-à-dire :

31. Les instruments sont les mêmes pour l'analyse en 2011 avec les données de l'année correspondante.

- | | |
|---|--|
| 1. 85.59.11 : Services des écoles de langues | 3. 85.59.13 : Services d'enseignement professionnel n.c.a. |
| 2. 85.59.12 : Services des organismes de formation informatique | 4. 85.59.14 : Cours de soutien scolaire à domicile |
| | 5. 85.59.19 : Services d'enseignement n.c.a. |

Nous aurions voulu exclure les entreprises dont l'activité principale n'est pas explicitement le soutien scolaire (notamment les détails des codes 85.59), mais les données dont nous disposons ne le permettent pas. Nous avons également fait des estimations en retirant les entreprises 85.52. Les résultats sont qualitativement similaires mais les erreurs standards sont élargies. La distribution des instruments est présentée en Annexe E. Comme de nombreuses communes n'ont pas d'entreprise de ce type (pour plus de 45 % des observations que nous utilisons), nous ajoutons le nombre d'entreprises dans le département comme une proxy de la disponibilité proche de cours privé.

Dans l'idéal, on pourrait affiner cette stratégie avec une variable capturant la distance à la ville la plus proche ayant une entreprise de soutien, mais nos données ne nous permettent pas de produire un tel indicateur.

L'offre de cours privés sur le territoire français est représentée dans les deux cartes en Annexe E, figures A.3 et A.4. Que ce soit en 2008 ou en 2011, ces cartes montrent que les entreprises de soutien privé sont concentrées dans les grandes villes : Paris, Lyon, Toulouse... et plus généralement dans les zones urbaines. Certains territoires, plus ruraux, sont dépourvus d'entreprises de soutien scolaire. L'Yonne, par exemple, en 2008, n'a que 2 à 4 entreprises de soutien privé à Troyes et moins de deux dans deux autres villes. Les élèves de ce département ont un accès plus difficile à ces services qu'un élève scolarisé dans le Finistère, où des entreprises sont installées un peu partout dans le département. Notons qu'entre 2008 et 2011, le nombre d'entreprises offrant des services de soutien scolaire a considérablement augmenté ; le plan Borloo (2005), favorisant le développement des services à la personne, a probablement contribué à cet essor. Par exemple, leur nombre a plus que doublé à Toulouse, passant de 65 en 2008 à 132 en 2011 (maximum en France). L'identification va alors résider dans la comparaison d'élèves aux caractéristiques identiques, scolarisés dans des villes semblables en taille et en revenu moyen de sa population, mais dont l'une (Troyes, par exemple) n'a que quelques entreprises de soutien privé (2 à 4) et l'autre (Saint-Dié des Vosges) davantage (4 à 6).

Ce type d'instrument est assez classique même si ce sont plus souvent des indicateurs de distances qui sont utilisés en économie de l'éducation. Dans l'évaluation de l'effet des cours privés de soutien sur les performances en fin de lycée, c'est également ce que Zhang (2013) utilise sur des lycéens de la province de Shandong en Chine. Il obtient des résultats contrastés et hétérogènes dans les différentes disciplines testées (mathématiques, chinois et anglais) et sur le test d'entrée à l'université³². En effet, les impacts calculés sur les performances en mathématiques et en chinois ne sont généralement pas significatifs, ceux en anglais sont généralement positifs et significatifs alors que pour les scores au test d'entrée à l'université, les effets sont négatifs et significatifs dans toutes les spécifications. Les résultats varient sensiblement lorsqu'ils séparent l'échantillon en deux groupes : rural et urbain. On note qu'à l'image de nos données, les estimations par régressions linéaires indiquent un effet négatif ou nul dans ces domaines, suggérant un biais de sélection de même signe (les élèves les moins performants ont recours au soutien privé). À notre connaissance, la seule autre étude publiée dans une revue à comité de lecture qui utilise des variables instrumentales (ou d'autres

32. National College Entrance Exam.

modèles à exclusion restriction) est Dang (2007). Dans cet article, l'auteur utilise le coût horaire des cours privés de soutien comme instrument et obtient des résultats positifs sur les performances des élèves au Vietnam. Cependant, les différences de niveau de vie dans les différentes régions peuvent probablement expliquer les différences tarifaires. Dès lors, les familles riches vivant dans des quartiers favorisés ont de fortes chances de se voir proposer des tarifs plus élevés, ce qui met en doute la validité de l'instrument. Hof (2014) cherche à borner³³ l'effet des cours de soutien sur les performances des élèves âgés de 15 ans et exploite l'éducation des parents comme *variable instrumentale monotone* (Manski et Pepper, 2000) pour restreindre l'intervalle incluant l'effet des cours de soutien. En exploitant les données PISA 2009 en Suisse, elle obtient des bornes de l'effet des cours de soutien³⁴ plutôt positives sur les performances en lecture, suivant les hypothèses³⁵.

b) Validité de la stratégie

De façon générale, un bon instrument doit posséder plusieurs propriétés dont certaines ne sont pas testables statistiquement. La première est que l'instrument doit suffisamment influencer le soutien scolaire pour que la démarche soit valide. Cette propriété est validée dans notre cas indépendamment des spécifications et des variables de contrôle ajoutées, comme le montre le tableau A.2 de l'Annexe F. Les instruments sont significativement³⁶ différents de 0, de signes et de magnitudes attendus. Les estimations indiquent qu'une augmentation d'1 % du nombre d'entreprises de soutien privé dans la commune augmente la probabilité de prendre des cours de soutien pour un élève de 6^e d'environ 1 point de pourcentage³⁷, toutes choses égales par ailleurs. Le fait d'avoir 1 % d'entreprises de soutien privé supplémentaire dans la commune fait passer la probabilité de prendre des cours de soutien de 9 à 10 %³⁸.

Cadre conceptuel et modélisation

Pour bien saisir ce que nous proposons dans ce travail, il est utile de présenter simplement les modèles sous-jacents afin de saisir les enjeux de ce type de stratégie. Nous reprenons donc la notation standard initialement proposée par Rubin (1974) avec des performances potentielles Y_j indicées $j = \{0; 1\}$ suivant qu'un élève prend ou non des cours de soutien en 2008, un ensemble de caractéristiques \mathbf{X} observables incluant les performances passées. Le modèle théorique autorise des paramètres différents entre les deux groupes (donc une équation pour chaque). La réalisation du modèle est représentée par l'équation 1c :

$$Y_0 = \mu_0 + \mathbf{X}\beta_0 + \varepsilon_0, \quad E(\varepsilon_0|\mathbf{X}) = 0, \quad \text{élèves sans soutien} \quad (1a)$$

33. *Non-parametric bounds analysis* à la Manski (1990, 1997).

34. pas nécessairement privés et payants ici puisque le questionnaire PISA ne permet pas de les distinguer d'autres formes de tutorats.

35. Le modèle basé sur un *monotone treatment response* excluant donc la possibilité d'un effet négatif du tutorat borne l'effet dans un intervalle strictement positif. Les autres hypothèses (monotone treatment selection et monotone instrumental variable) n'excluent pas l'absence d'effet de l'intervalle.

36. Le test de significativité jointe des instruments (F-test) est systématiquement différent de 0 et supérieur à 10 pour toutes les spécifications, indiquant une puissance de test suffisante.

37. Les estimations varient suivant les spécifications entre 0.97 et 1.5.

38. La probabilité de base pour l'élève possédant les modalités de références est de 9.0 %.

$$Y_1 = \mu_1 + \mathbf{X}\beta_1 + \varepsilon_1, \quad E(\varepsilon_1|\mathbf{X}) = 0, \quad \text{élèves avec soutien} \quad (1b)$$

$$Y = Y_0 + w(Y_1 - Y_0) \quad (1c)$$

où w est une variable binaire indiquant le soutien scolaire. En substituant les deux premières équations dans la troisième, on obtient :

$$Y = \mu_0 + (\mu_1 - \mu_0)w + \mathbf{X}\beta_0 + w(\mathbf{X}\beta_1 - \mathbf{X}\beta_0) + \varepsilon_0 + w(\varepsilon_1 - \varepsilon_0) \quad (2)$$

On peut également définir à ce stade une équation de sélection dans le soutien scolaire telle que :

$$w = \theta_0 + \mathbf{X}\theta_1 + \epsilon \quad (3)$$

où ϵ est un terme d'erreur.

Pour faire lien avec les estimations précédentes par matching ou IPWRA, nous avons estimé l'équation 2 en supposant que :

$$\varepsilon_0 = \varepsilon_1 = \varepsilon, \quad \beta_0 \neq \beta_1, \quad E(\varepsilon|\mathbf{X}, w) = 0 \quad (4)$$

C'est-à-dire, un impact différent des caractéristiques observables mais un effet commun des caractéristiques inobservables dans les deux groupes et surtout, un processus de sélection **exogène** conditionnellement à \mathbf{X} .

Cependant, l'hypothèse d'exogénéité conditionnelle des cours de soutien est peu crédible, même conditionnellement à \mathbf{X} . Dès lors, si $E(\varepsilon|\mathbf{X}, w) \neq 0$, l'estimation de l'équation 2 par moindre carré n'identifie pas l'impact causal des cours de soutien.

Si en revanche il existe un instrument ou un vecteur d'instruments valide Z tel que le système puisse se réécrire :

$$\begin{aligned} Y &= \mu_0 + (\mu_1 - \mu_0)w + \mathbf{X}\beta_0 + w(\mathbf{X}\beta_1 - \mathbf{X}\beta_0) + \varepsilon_0 + w(\varepsilon_1 - \varepsilon_0) \\ w &= \theta_0 + \eta Z + \mathbf{X}\theta_1 + \epsilon \end{aligned} \quad E(Z|\mathbf{X}, \epsilon) = 0 \quad (5)$$

On peut alors exploiter la variation de w induite par l'instrument Z pour identifier l'effet des cours particuliers sur les performances des élèves.

Toutefois, en raison de la nature binaire de la variable w , les modèles "standards" de variable instrumentale ne sont pas parfaitement adaptés. C'est pourquoi nous adoptons 3 méthodes d'estimations de l'effet des cours particuliers :

- Des estimations 2SLS standards
- Des estimations utilisant les prédictions d'un probit comme instrument dans un modèle 2SLS permettant d'obtenir un modèle plus efficace qu'un 2SLS sans tomber dans la *régression interdite* (Angrist et Pischke, 2009, p.144)
- Des estimations d'un modèle de sélection endogène à la Heckman (appelé ensuite Heckit),

plus précises mais reposant sur des hypothèses structurelles fortes.

Les différentes spécifications sont basées sur le modèle (2) et font l'hypothèse que :

$$\varepsilon_0 = \varepsilon_1 = \varepsilon, \quad \beta_0 = \beta_1 = \beta, \quad E(\varepsilon|\mathbf{X}, w) \neq 0 \quad (6)$$

c'est-à-dire celles de caractéristiques communes pour les deux groupes mais d'une sélection endogène.

Enfin, le modèle Heckit estime l'effet du traitement par l'équation :

$$E(Y|w, \mathbf{X}, Z) = \mu_0 + \alpha w + \mathbf{X}\beta + \rho_1 w \frac{\phi(q\theta)}{\Phi(q\theta)} + \rho_0(1-w) \frac{\phi(q\theta)}{1-\Phi(q\theta)} \quad (7)$$

où α est l'effet moyen des cours, ρ_1 et ρ_0 la corrélation entre les termes d'erreurs des deux outcomes potentiels 1a et 1b et l'erreur de l'équation de sélection dans les cours privés, $\phi(\cdot)$ et $\Phi(\cdot)$ sont les densités et fonctions de répartition d'une loi normale, respectivement.

Le modèle Heckit est le plus simple et performant ; il est estimé par maximum de vraisemblance, mais cette relative efficacité se fait au prix d'hypothèses très fortes sur la distribution des résidus. En effet, le modèle impose une distribution tri-normale des erreurs des équations d'outcomes potentielles (1a et 1b) et de l'erreur de l'équation de sélection (5).

Les estimations tiennent compte de la potentielle corrélation intra-établissement en calculant les erreurs types *clusterisé* au niveau de l'établissement.

La seconde propriété d'un instrument est qu'il ne doit influencer les performances qu'à travers son effet sur la probabilité de suivre des cours privés de soutien. Si le nombre d'entreprises de soutien privé a un impact, même indirect, sur les performances autrement que par le soutien scolaire, alors l'estimation par variable instrumentale est invalide³⁹. **Cette hypothèse cruciale n'est pourtant pas testable, statistiquement.** On ne peut que discuter sa validité.

Les menaces sur notre stratégie résident principalement dans les raisons qui poussent les entreprises de soutien scolaire à s'implanter en nombre dans une ville. Si ces dernières s'implantent parce que les familles de ce quartier sont plus riches ou plus disposées à recourir à leurs services (information que nous ne pouvons ni vérifier ni infirmer), alors que les performances des élèves sont liées à ces caractéristiques ou, plus simplement, parce que la population d'élèves y est d'un niveau scolaire particulièrement faible, alors la validité de notre stratégie d'identification peut potentiellement être mise en péril.

Il est évident que la distribution des entreprises de soutien scolaire n'est pas aléatoire sur le territoire et le fait qu'il y ait des zones de concentration est lié au marché potentiel pour ces entreprises. La présence d'établissements prestigieux, de classes préparatoires aux grandes écoles ou, plus généralement, d'établissements de l'enseignement supérieur est probablement un moteur important de l'évolution du marché, en agissant à la fois sur la demande, et l'offre (par exemple, Acadomia recrute des intervenants à partir du niveau licence). Cependant, nous pensons que ce n'est pas une menace pour notre stratégie empirique dans la mesure où l'identification repose sur l'effet marginal du nombre d'entreprises, c'est-à-dire, ce qu'une entreprise supplémentaire dans la commune apporte sur la probabilité de suivre des cours de soutien pour

39. Les résultats sont alors distordus au point qu'on peut ne pas rejeter, ou au contraire rejeter à tort, et de façon imprévisible, l'hypothèse nulle d'absence d'effet causal (Berkowitz et al., 2008).

un élève donné, indépendamment du nombre initial.

Cette stratégie impose toutefois que cet effet marginal soit constant sur toute la distribution du nombre d'entreprises, ce qui peut être discutable⁴⁰. Par ailleurs, le peu de variation entre les différentes spécifications qui intègrent un ensemble plus ou moins important de caractéristiques des élèves (notamment leur niveau initial), de leur famille (PCS, éducation des parents, aide à la maison, taille de la famille,...) et de leur environnement (notamment la taille de la ville et des indicateurs de revenu) atteste d'une certaine robustesse de cette stratégie d'identification.

c) Résultats

Le tableau 4 présente les résultats des estimations par variables instrumentales et les compare avec les estimations *naïves* obtenues par des modèles de régressions linéaires simples (OLS) qui sont présentés dans la première colonne. Pour tous les domaines testés, ces derniers sont compris entre -5 % et -11 % d'un écart-type dans la dimension testée.

Le premier constat que l'on peut faire est que toutes les estimations par variables instrumentales sont positives, souvent significativement différentes de 0, à l'exception de la variable *Raisonnement* pour lesquels les résultats sont négatifs et parfois significatifs. Les résultats des estimations sont assez variables suivant les spécifications mais les différentes méthodes fournissent des estimations cohérentes. Ensuite, la magnitude et les erreurs-standards des estimations par 2SLS sont beaucoup plus élevées que les autres⁴¹. Une des faiblesses de ce modèle est que la première équation estime un modèle linéaire en probabilité qui ne garantit pas des prédictions dans l'intervalle [0;1]. Or ce sont précisément ces prédictions qui sont utilisées dans la seconde équation à la place de la variable soutien instrumentée. On focalise donc notre attention sur le modèle Probit 2SLS, plus efficace que le 2SLS et moins contraint que le heckit (Angrist et Pischke, 2009 ; Cerulli, 2015).

La première colonne de chaque méthode fournit des résultats positifs souvent élevés en contrôlant pour un minimum de variables (i.e. la taille de la ville et les performances en mathématiques et en Français à l'entrée en 6^e). La seconde contrôle pour tout un ensemble de caractéristiques socio-démographiques de l'élève et de sa famille et produit des résultats très proches de la première. La troisième contrôle également le niveau de revenu dans la commune où l'élève est scolarisé, ce qui diminue la magnitude des coefficients dans presque toutes les spécifications et diminue également la précision des estimations. Ce résultat peut s'expliquer de deux façons : statistiquement d'une part, l'ajout de cette variable affaiblit le lien entre l'instrument et la variable soutien⁴². Plus spécifiquement, le nombre d'entreprises de soutien privé est positivement corrélé avec le revenu moyen et le revenu total des habitants de la commune, ce qui est attendu, mais provoque de la multi-colinéarité dans le modèle⁴³ de sélection ce qui augmente les erreurs standards et tend à rendre plus instable le modèle. D'autre part, de façon plus conceptuelle, l'ajout de ce

40. On pourrait par exemple s'attendre à des effets de seuils, notamment entre 0 et 1, et probablement un effet marginal qui tend vers 0 quand le nombre d'entreprises est très important.

41. Les erreurs standards sont deux à trois fois plus élevées que les modèles probit-2SLS et heckit

42. Ce qui se voit par le F-test partiel qui diminue dans le premier étage mais reste suffisamment important, selon la règle recommandé par Stock et al. (2002) (Voir tableau en Annexe F.) ce qui a pour conséquence directe de biaiser l'estimation vers le coefficient OLS (Cameron et Trivedi, 2005)

43. La variable de revenu dans la commune est la seule à avoir un facteur d'inflation de variance (VIF) supérieur à 5 dans la première équation. Aucune variable n'a un VIF de plus de 3 dans les autres spécification.

Tableau 4 – Estimation de l'effet des cours particuliers sur les performances des élèves en 2008 par variable instrumentale et Heckit

	OLS	2SLS (1)	2SLS (2)	2SLS (3)	Probit-2SLS (1)	Probit-2SLS (2)	Probit-2SLS (3)	Heckit (1)	Heckit (2)	Heckit (3)
lexique 6	-0.041** (0.019)	2.963*** (0.634)	2.443*** (0.706)	1.501** (0.711)	1.512*** (0.389)	0.560** (0.243)	0.342 (0.262)	1.077*** (0.036)	0.952*** (0.038)	0.946*** (0.039)
Phrases lacunaires	-0.052*** (0.019)	1.264*** (0.482)	0.565 (0.539)	0.573 (0.636)	0.269 (0.337)	0.295 (0.245)	0.501* (0.277)	0.085 (0.080)	0.121* (0.065)	0.128** (0.065)
Maths	-0.077*** (0.016)	1.576*** (0.449)	1.144** (0.511)	0.907 (0.583)	1.321*** (0.342)	0.822*** (0.227)	0.804*** (0.247)	0.252*** (0.049)	0.188*** (0.049)	0.161*** (0.049)
Compréhension	-0.066*** (0.020)	0.766 (0.471)	0.196 (0.562)	0.316 (0.665)	0.974*** (0.375)	0.545** (0.273)	0.695** (0.308)	1.050*** (0.032)	1.024*** (0.033)	1.037*** (0.033)
Raisonnement	-0.117*** (0.026)	-0.413 (0.536)	-0.382 (0.646)	0.419 (0.766)	-0.518 (0.441)	-0.535* (0.321)	-0.460 (0.354)	-0.134*** (0.038)	-0.124*** (0.039)	-0.124*** (0.040)
Contrôles	Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui
Revenu/commune	non	Non	Non	Oui	Non	Non	Oui	Non	Non	Oui
Observations	18554	19674	19343	18554	19674	19343	18554	19674	19343	18554

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes ajustées par établissement (*cluster*)

Tous les modèles utilisent comme instrument le nombre d'entreprises enregistrées dans la démographie des entreprises (INSEE) en 2008. Ces bases ont été appariées au Panel 2007 sur la base du code commune de l'établissement scolaire où étudie l'élève durant l'année 2007-2008. Les spécifications contiennent donc le nombre d'entreprises de soutien scolaire dans la commune et dans le département, exprimé en logarithme.

Les modèles (1) ont pour seules variables de contrôles les polynômes d'ordre deux du niveau en mathématiques et en Français à l'entrée en 6^e et le nombre d'habitants de la commune de l'établissement (regroupé en classes, voir la 1st stage en Annexe F). Les modèles deux ajoutent à ces variables le sexe de l'élève, l'éducation de ses parents, leur origine géographique, s'il a déjà redoublé, la PCS du chef de famille, la structure familiale, le nombre de frères et de sœurs de l'élève et la fréquence de l'aide des parents. Les modèles 3 ajoutent le logarithme du revenu moyen dans la commune de l'établissement et le logarithme du revenu total (basé sur le revenu net déclaré des foyers fiscaux de la commune de l'établissement d'après INSEE : *Revenus imposables et montant des impôts*.)

contrôle peut potentiellement faire disparaître l'équation de sélection si la corrélation initialement observée entre le nombre d'entreprise et le soutien scolaire capture en réalité, de façon approximative, la richesse de la commune. Nous rejetons cette hypothèse en constatant que les coefficients de l'équation de sélection ne sont pas diminués par l'ajout de ces variables de contrôle et restent significatifs à 1 % bien que les erreurs standards soient beaucoup plus élevées.

Les estimations pour la variable *Lexique 6* sont positives et significatives. Suivant les spécifications, la magnitude de l'effet des cours de soutien varie autour d'un écart-type. Dans le tableau 3, nous avons montré que la différence moyenne de performance entre les élèves qui prennent des cours de soutien et les autres était de -40 % d'un écart-type. Ainsi, l'effet *causal* des cours de soutien privés tels qu'estimé ici par les stratégies de variables instrumentales compense largement l'écart de performance initial.

Les estimations sur les *Phrases lacunaires* sont très imprécises et rarement significatives. Les cours de soutien en 6^e ne semblent pas améliorer la compréhension de texte, telle que mesurée par ce test.

Comme pour la variable *Lexique 6*, les estimations de l'effet des cours de soutien sur les performances en *Mathématiques* en 6^e sont positives et significatives dans toutes les spécifications. Les estimations les plus conservatrices fournies par les *Heckit* indiquent un effet des cours privés augmentant les performances en mathématiques d'environ 20 % d'un écart-type. Les *probits 2SLS* fournissent des estimations presque quatre fois plus élevées. L'effet sur les performances en mathématiques semble donc important.

Les estimations sur la variable *Compréhension* sont moins précises, non significatives pour les *2SLS* et seulement significatives à 5 % pour les *probits 2SLS*. Les estimations des *Heckit* sont très positives et significatives. Il semble donc que les cours de soutien impactent positivement la compréhension de texte des élèves en 6^e.

Enfin, et de façon assez surprenante, l'effet des cours privés de soutien sur les capacités de raisonnement des élèves semble négatif même si, comme pour la variable *Compréhension*, la significativité n'est pas toujours atteinte. Les capacités de raisonnement et de logique des élèves semblent donc être négativement affectées par les cours privés de soutien en 6^e.

d) Discussion

Les élèves qui prennent des cours de soutien en 2008 ont des performances nettement plus faibles que les élèves qui n'en prennent pas. Cependant, cette différence disparaît presque complètement dans toutes les disciplines dès qu'est pris en compte l'effet des caractéristiques socio-démographiques des élèves. Mais l'absence de différence, à caractéristiques observables données, ne signifie pas que les cours privés n'ont pas d'effet sur les performances des élèves. S'il existe des facteurs inobservables qui poussent les élèves à prendre des cours privés, alors la comparaison des deux groupes ne mesure pas l'effet des cours, mais mélange cet effet avec celui des caractéristiques inobservables. Pour contourner ce problème, on exploite l'existence de variations *exogènes* dans la probabilité de prendre des cours de soutien scolaire induite par la contrainte d'offre. On compare ainsi des élèves, par ailleurs tout à fait semblables, dont les uns peuvent facilement souscrire à des cours de soutien et les autres avec plus de mal, parce-qu'il y a moins d'entreprises offrant ce service dans leurs communes. La relative abondance de l'offre *décontraint* les disponibilités des tuteurs privés et peut avoir un effet concurrence des prix sur les cours, les rendant plus accessibles. Cette variation du nombre d'entreprises de soutien privé ne devrait pas affecter les performances des élèves autrement qu'en

leur permettant d'y souscrire plus facilement. Si cette hypothèse est valide, l'exploitation de cet *instrument* mesure l'effet causal des cours privés de soutien sur les performances.

En dehors du résultat sur le raisonnement, ces conclusions sont finalement assez attendues dans la mesure où les élèves qui prennent des cours de soutien le font principalement en français et en mathématiques. Les tests de *Lexique 6* s'appuient sur les contenus de manuels scolaires et correspondent aux programmes de l'année qui sont probablement travaillés dans les cours de soutien. L'effet négatif sur le raisonnement est plus complexe à interpréter. Peut-être s'agit-il d'un effet de préparation aux contrôles⁴⁴ ; les cours de soutien entraîneraient uniquement les élèves aux exercices types, sans développer réellement leurs capacités de raisonnement. Les items de ce test visent à évaluer l'intelligence fluide (Cattell, 1971), qui repose sur une capacité à établir des relations entre des éléments d'information et en inférer des corrélations dans des situations peu familières. Ce test est peu sensible aux acquisitions, au langage et à la culture de manière générale (Chartier, 2012). Le soutien scolaire semble donc faire perdre en partie cette capacité aux élèves, probablement en les rendant moins autonome dans leur habileté à résoudre des problèmes. Pour cette interprétation, nous ne pouvons, malheureusement, que spéculer.

Les caractéristiques socio-démographiques des élèves qui prennent des cours sont significativement différentes de celles des élèves qui n'en prennent pas. En particulier, et à autres caractéristiques données, ils sont en difficulté à l'entrée en 6^e mais sont davantage issus de milieux sociaux favorisés, citadins, inscrits dans le privé, et plutôt d'origines étrangères.

L'usage de cours de soutien payant leur permet alors d'améliorer significativement leurs performances en mathématiques et leur mémoire encyclopédique, au prix d'une moindre capacité de raisonnement. En terme de magnitude, les estimations les plus conservatrices montrent que les cours privés de soutien compensent au moins l'écart de performance initial.

Soutien scolaire et performances passées

Pour se convaincre davantage de la validité de la stratégie par variable instrumentale, nous avons voulu nous assurer que le soutien scolaire en 6^e n'a pas d'effet sur les performances passées des élèves. Si c'était le cas, cela indiquerait que les relations identifiées ne reflètent pas un effet causal des cours sur les performances mais davantage un effet de sélection où les entreprises s'installent là où les élèves sont les moins performants. Cette démarche s'apparente à un test de falsification. Malheureusement, nous ne disposons pas de mesures des performances en primaire. Nous utilisons donc le redoublement comme proxy de ces performances. Les rapports du Cnesco (2015) et l'ouvrage d'Heim et al. (2015) ont montré que le redoublement en primaire concerne les élèves les plus faibles et vise à remédier aux difficultés scolaires. Dès lors, plus la probabilité d'avoir redoublé est grande, plus la probabilité que l'élève ait des performances faibles est élevée. Nous adoptons alors la même stratégie de variable instrumentale que précédemment en ayant cette fois pour variable dépendante le redoublement primaire. Toutefois, la nature binaire de cette variable nous contraint à utiliser d'autres méthodes d'estimations que nous ne détaillons pas. Nous présentons les effets marginaux de façon à ce que les estimations soient directement interprétables en point de probabilité.

44. *teaching to the test* (Firestone et al., 2004).

Quelle que soit la méthode utilisée, les estimations sont imprécises et proches de zéro. La significativité n'est jamais atteinte, même à des seuils de risque élevés. Les cours privés de soutien scolaire n'ont pas d'effet sur les performances passées. Les entreprises de soutien scolaire ne semblent donc pas s'installer dans des zones où les élèves sont moins performants.

Tableau 5 – Régressions instrumentales placebo

Variable dépendante	Méthodes	Effets marginaux	(SE)	P-value	Nombre d'observations
Redoublement en primaire	IV-Probit	0.116	(0.18)	0.508	32037
	Probit-2sls	0.014	(0.15)	0.926	32037
	Biprobit	0.031	(0.04)	0.481	32037

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Les erreurs standard présentées sont robustes et ajustées par établissement (*cluster*)

Note : les trois estimations contiennent les mêmes variables de contrôle que les modèles (2) du tableau 4 à l'exception des performances d'entrée en 6^e (car elles sont postérieures au redoublement primaire) et le retard à l'entrée en 6^e.

3 L'effet des cours de soutien en 4^e/3^e

Cette section reproduit exactement les mêmes modèles que la précédente avec les données de l'année 2011. Cependant, comme les élèves qui prennent des cours de soutien en 2011 sont différents de ceux de 2008, nous proposons plusieurs estimations restreignant l'échantillon aux élèves qui ne prenaient pas de cours en 2008. En outre, nous proposons des variantes tenant également compte du niveau de performances en 2008. Ces modélisations à *valeur ajoutée* sont biaisées négativement (Annexe D) et fournissent donc une estimation basse de l'effet des cours privés.

A Comparaison toutes choses égales par ailleurs des performances

Le tableau A.3 de l'Annexe G présente l'estimation de la différence de performances entre les élèves qui prennent ou non des cours en 2011, puis en contrôlant par les caractéristiques socio-démographiques de la même manière que dans la section A. Ces indicateurs sont présentés pour l'échantillon complet, pour l'échantillon d'élèves qui ne prenaient pas de cours de soutien en 2008, avec ou sans contrôle du niveau de performance en 2008.

L'écart brut de performances entre les deux groupes est nettement plus faible qu'en 2008, à l'exception notable des mathématiques et du raisonnement, où les différences sont respectivement de -36 % et -28 % d'un écart-type (tableau A.3). Les différences brutes dans les autres dimensions sont deux fois plus faibles, soit environ -15 % d'un écart-type. Ainsi, l'écart entre les élèves qui prennent des cours et les autres est nettement moins important en 4^e/3^e qu'en 6^e. On peut interpréter ce résultat de deux façons. Soit les élèves qui prennent des cours sont de relativement moins mauvais élèves qu'en 2008, soit les différences moyennes

entre les groupes ont été gommées par un usage des cours particuliers entre les deux dates que nos données ne nous permettent pas de contrôler. Cet effet de "contagion" est une menace sérieuse à l'identification de l'effet des cours particuliers puisque nous ne sommes pas en mesure de renseigner le passé des élèves en la matière. Il se peut que certains élèves du groupe aient suivi des cours de la 5^e à la 3^e, que d'autres aient seulement débutés en 3^e,... Les différences brutes dans l'échantillon qui exclut les élèves qui prenaient des cours en 2008 sont légèrement plus faibles qu'avec l'échantillon complet, ce qui va dans le sens d'une sélection de meilleurs élèves dans les cours privés mais n'exclut pas la contagion des groupes.

Lorsqu'on tient compte des caractéristiques socio-démographiques observables des élèves, la différence est diminuée, mais beaucoup moins qu'en 2008. La seule dimension pour laquelle la différence disparaît complètement est la mémoire encyclopédique (*Lexique 6*). Les autres mesures restent significativement différentes de 0 et négatives. Les modèles à valeur ajoutée contribuent à réduire les différences de performances mais elles restent significatives. Les résultats pour l'échantillon complet ou restreint aux élèves qui ne prenaient pas de cours en 2008 sont tout à fait similaires une fois qu'on contrôle les caractéristiques des élèves et de leur environnement, ce qui indique que l'effet de sélection dynamique est assez bien corrigé par l'ajout de ces variables de contrôle.

B Hétérogénéité des différences en fonction du niveau de performance

Comme pour le soutien en 2008, nous avons estimé des régressions quantiles pour identifier l'hétérogénéité de la différence de performances entre les élèves qui prennent ou non des cours de soutien en 2011, conditionnellement à leurs caractéristiques socio-démographiques. Nous présentons les résultats des modèles avec tous les élèves⁴⁵, en contrôlant ou non par le niveau de performance en 2008. Les résultats sont représentés graphiquement dans les figures A.5 et A.6 de l'Annexe H.

Contrairement à 2008, les pentes des régressions quantiles sont assez différentes pour chaque dimension testée. Pour la variable *Lexique 6* et la variable *Compréhension*, les coefficients des régressions quantiles ne sont jamais significativement différents de 0, ce qui correspond à peu près à la valeur estimée par les modèles de régression linéaire. Les allures un peu étranges des graphiques (surtout *Lexique 6*) ne peuvent pas être interprétées car les variations ne sont pas significatives.

Les estimations pour la variable *Phrases lacunaire* forment une courbe en "U inversé", indiquant que la différence entre les élèves qui prennent des cours privés et ceux qui n'en prennent pas est plus importante parmi les moins performants en compréhension écrite que pour les autres (i.e. les 30 % d'élèves les plus faibles). Au delà, la différence n'est pas significative. Les résultats de la variable *Raisonnement* sont assez semblables même s'ils sont négatifs et significatifs pour tout niveau de quantile mais très imprécis pour les 5 % les plus faibles et les 5 % les plus forts. Il semble toutefois que la différence soit légèrement plus importante pour les élèves les meilleurs.

Enfin, les estimations pour les mathématiques sont globalement négatives mais sont légèrement différentes lorsqu'on contrôle par le niveau en 2008. Dans tous les cas, l'écart imputable au cours de soutien est maximal pour les élèves les plus faibles et relativement stable autour de 10 % d'un écart-type pour les élèves au milieu de la distribution. Toutefois, lorsqu'on contrôle pour le niveau de performance en 6^e, il semble que la différence de performance soit plus ténue (et même non significative) pour les 20 % d'élèves

45. i.e. en gardant ceux qui prenaient des cours en 2008.

les meilleurs.

C Estimations par variables instrumentales

a) Équation de sélection et puissance de test

Nous avons estimé les mêmes modèles qu'en 2008 en ajoutant ou non une variable de contrôle pour le niveau de performance en 2008 et présentons ces résultats en Annexe J. Avant de les discuter, il est utile de s'intéresser à ceux de l'équation de sélection (i.e; first stage) présentés dans l'Annexe I. Comme pour le modèle de 2008, les instruments sont significativement différents de 0, de signes et de magnitudes attendus : une augmentation d'1 % du nombre d'entreprises de soutien privé dans la commune augmente la probabilité de prendre des cours de soutien pour un élève de 1 à 2 points de pourcentage environ⁴⁶, toutes choses égales par ailleurs. Le modèle est globalement significatif comme l'indique la F-stat globale. Le test de significativité jointe des instruments est systématiquement différent de 0. Cependant, le 3^e modèle a une F-stat. d'à peine 5, ce qui fait craindre un biais élevé dû à des instruments faibles dans cette spécification. Nous avons discuté dans la partie C de la section précédente des raisons probables de cette perte de puissance. Il semble qu'ici, la multi-colinéarité entre le nombre d'entreprises et le revenu de la commune soit si importante qu'elle diminue très fortement la puissance du F-test.

Ainsi, l'effet minimal détectable par nos estimations se trouve largement augmenté. Dès lors, à moins que la valeur des coefficients ne soit très élevée, nous risquons de ne pas réussir à identifier d'effets significatifs.

b) Résultats

Lorsqu'on regarde les tableaux de l'Annexe J, c'est effectivement le constat que nous pouvons faire. Dans les deux tableaux, la vaste majorité des estimations par variables instrumentales fournit des résultats non significatifs, très imprécis et très variables suivant les spécifications et les méthodes d'estimations. Seules les dimensions de compréhension de texte et de raisonnement semblent négativement affectées par les cours privés de soutien en 2011.

En effet, le premier tableau, qui présente les résultats des modèles sans contrôler le niveau de performance en 2008, fournit des résultats de signes et de niveaux de significativité variables suivant les modèles pour les dimensions *Mathématiques*, *Phrases lacunaires* et *Lexique 6*. Le modèle 2SLS standard est visiblement le plus instable et souffre particulièrement de ce problème de multi-colinéarité. Les estimations par *probit 2SLS* et *Heckit* sont plus conservatrices en magnitude, mais rarement significatives.

Les estimations de l'effet des cours de soutien sur la compréhension de texte sont significatives pour les modèles 2SLS et *probit 2SLS*. Les résultats les plus conservateurs du *probit 2SLS* indiquent un effet négatif d'environ 40 % d'un écart type (pour les trois spécifications). Les estimations des 2SLS sont beaucoup plus élevées en magnitude mais aussi beaucoup plus imprécises.

L'ajout des performances en 2008 dans l'ensemble des variables de contrôle modifie peu les résultats que nous venons de présenter. En effet, l'impact sur les variables *Mathématiques*, *Phrases lacunaires* et *Lexique 6* n'est pas précisément estimé et presque jamais significatif même si en tendance, les coefficients sont assez généralement négatifs. La magnitude des coefficients associés à la variable *Compréhension* dans

46. Les estimations vont de 0.87 à 1.9 suivant les spécifications.

les modèles *probit 2SLS* reste la même (autour de 35 % d'un écart-type), celle sur le raisonnement est beaucoup plus variable.

Le coefficient associé aux variables retardées (score 2008) est systématiquement positif et significatif, compris entre 0 et 1. Si la stratégie de variable instrumentale permet d'estimer correctement l'effet des cours de soutien, le biais de variable omise qui affecte le score initial⁴⁷ affecte toujours l'estimation de l'effet de ce score. De l'Annexe D, nous savons que ce biais atténue la valeur du coefficient. Il nous est donc difficile de savoir si les trajectoires des bons et des plus mauvais élèves divergent, évoluent parallèlement ou convergent.

c) Discussion

En employant exactement la même stratégie qu'en 2008, c'est-à-dire en exploitant la contrainte d'offre de cours privés dans la commune où l'élève est scolarisé, nous obtenons des résultats plus instables, moins informatifs et plus imprécis. Nous parvenons toutefois à identifier un effet plutôt négatif sur les capacités de compréhension de texte des élèves et sur leur capacité de raisonnement. Les autres dimensions testées ne semblent pas affectées par les cours de soutien en 2011. Cependant, le manque de puissance statistique réduit par la multi-colinéarité dans l'équation de sélection nous fait craindre d'avoir, pour certaines variables, accepté à tort l'hypothèse d'absence d'effet. Le biais que ce phénomène induit fait en effet tendre les coefficients estimés vers ceux des modèles de régressions linéaires classiques.

D'autre part, l'impossibilité de contrôler le passé des élèves en matière de soutien scolaire est très certainement responsable des erreurs de mesures et probablement aussi de la plus forte multi-colinéarité. Ce problème génère un potentiel effet de *contagion* entre les groupes et biaise nos estimations en raison d'erreurs de mesures de la variable soutien. Celle-ci ne classe pas correctement les élèves, les cours de soutien affectent simultanément les deux groupes et atténuent les différences. Ce phénomène est quelque part similaire à un problème d'endogénéité puisqu'il produit les mêmes effets : il atténue les estimations de la variable mal mesurée et peut biaiser les autres coefficients du modèle si ceux-ci sont corrélés à la variable mal mesurée (Treiman, 2009, p. 258-261). Toutefois, notre stratégie de variable instrumentale a peu de chances de corriger à la fois le problème de sélection endogène et d'erreur de mesure. En effet, le nombre d'entreprises de soutien scolaire dans une commune augmente entre les deux périodes, mais ne varie pas énormément d'une année à l'autre. Aussi, les coefficients des instruments estimés par nos modèles de sélection ont toutes les chances d'être similaires d'une année à l'autre⁴⁸. C'est pourquoi, même si la probabilité de prendre des cours pour un élève donné était indépendante du temps⁴⁹, nos estimations par variable instrumentale ne pourraient pas corriger ces erreurs de mesure. On peut alors se dire que ce problème pouvait déjà exister pour l'année 2008. Nous pensons cependant que l'entrée en 6^e marque un tel changement dans les méthodes de travail que l'effet des cours de soutien est moins dépendant du soutien passé qu'en 4^e ou 3^e, où les élèves peuvent capitaliser leurs acquis des cours précédents, ou au contraire, avoir moins de gains au soutien scolaire du fait que les méthodes de travail du collège sont déjà au moins partiellement acquises.

47. Notamment parce que nous ne tenons pas compte du soutien passé.

48. C'est en effet le cas pour 2008 et 2011.

49. Si on supposait que le fait de démarrer, d'abandonner ou de continuer à prendre des cours privés intervenait aléatoirement et sans mémoire, comme une chaîne de Markov.

Au final, il est difficile de conclure sur l'efficacité du soutien scolaire en 4^e/3^e tant nos estimations sont imprécises. Il semble toutefois moins bénéfique qu'en 6^e et semble toujours affecter négativement les capacités de raisonnement des élèves.

IV Cours privés, trajectoires et orientations des collégiens

Si les cours de soutien privés affectent les performances des élèves, ils peuvent également avoir un impact sur leurs trajectoires et leurs décisions d'orientation. En exploitant les questionnaires *orientation* complétés par les chefs d'établissements et en utilisant les données administratives dans le panel 2007, nous proposons une évaluation de l'effet des cours privés de soutien sur l'accès en second cycle supérieur.

1 Des collégiens aux trajectoires diverses

L'évolution des parcours scolaires dans l'enseignement secondaire français suit des tendances claires. Les trajectoires des collégiens sont de moins en moins heurtées, les redoublements, de moins en moins fréquents et l'accès en seconde générale et technologique en progression. Le nombre d'interruptions précoces de la scolarité a également diminué significativement dans les dernières décennies (Caille, 2014).

Dans le numéro 85 d'Éducation & Formations, Jean-Paul Caille (2014) compare les parcours scolaires des élèves du panel 1995 et 2007 en distinguant 7 trajectoires au collège :

- Parcours linéaire de la 6^e à la seconde GT
- Parcours linéaire de la 6^e au second cycle pro.
- Accès difficile⁵⁰ en seconde GT
- Orientation subie en pro.
- Orientation consentie en pro.
- Grand retard scolaire au collège
- Sortie précoce de formation initiale.

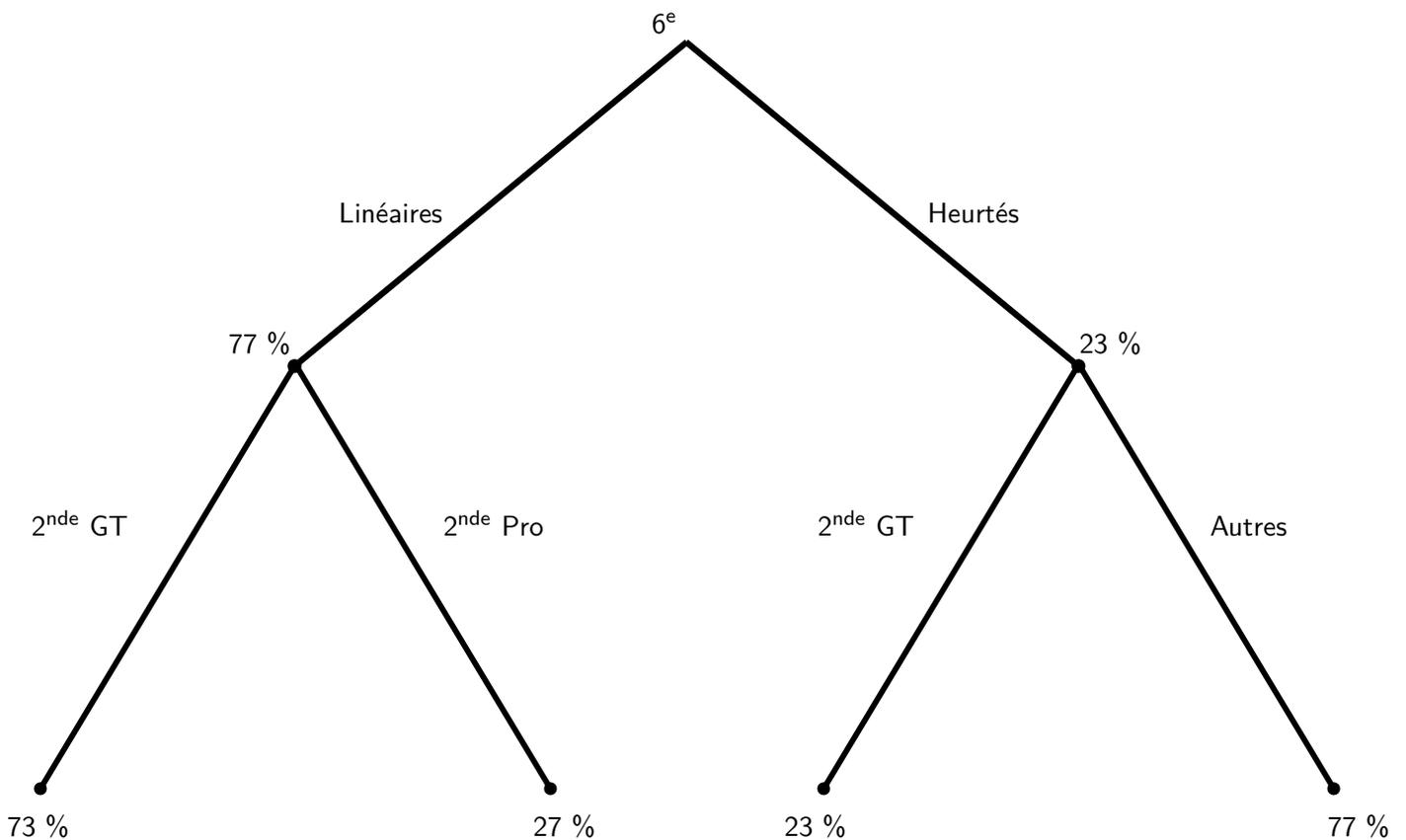
Pour la génération entrée en 6^e en 2007, 77.8 % des élèves accèdent à une seconde (GT ou pro.) sans redoubler et sans contestation de leurs vœux d'orientation par le conseil de classe. Seuls 58.2 % des élèves entrés en 6^e en 1995 avaient des parcours linéaires. La majorité des élèves (56.6 %) accède désormais directement à la seconde générale et technologique alors que pour la génération entrée en 6^e en 1995, 47.0 % y accédaient.

Dans l'analyse qui suit, nous n'avons pas défini des trajectoires aussi fines. Nous proposons une arborescence à deux niveaux où les élèves ont des parcours linéaires ou heurtés, et peuvent ensuite accéder à la seconde générale ou non (Figure 5). Plus précisément, les élèves qui ont un parcours linéaire peuvent ensuite entrer en 2^{nde} GT ou en seconde pro ; les élèves qui ont un parcours heurté peuvent entrer avec difficultés en 2^{nde} GT ou avoir un autre parcours, *i.e.* une 2^{nde} pro avec difficultés, un grand retard ou une sortie précoce du système scolaire.

Plus de 3/4 des élèves du panel ont une trajectoire linéaire et environ 1/4 ont des difficultés à atteindre la seconde. Parmi les élèves qui ont des parcours linéaires, les 3/4 vont en seconde générale et technologique

50. L'orientation en seconde se fait au terme d'un redoublement ou d'une orientation consentie ou imposée (Caille, 2014).

Figure 5 – Trajectoires au collège



Lecture : 77 % des élèves entrés en 6^e en 2007 ont un parcours linéaire jusqu'en seconde. Parmi eux, 73 % accèdent à une seconde générale et technologique.

et 1/4 s'oriente vers la voie professionnelle. A contrario, pour les élèves aux parcours heurtés, seul 1/4 parvient à rejoindre une seconde générale.

Le tableau 6 croise les trajectoires précédemment définies avec le soutien scolaire en 2008 et 2011. Les parcours des élèves qui y ont recours en 2011 ne sont pas plus souvent heurtés que ceux des élèves qui n'y ont pas recours. En revanche, les collégiens qui prenaient des cours en 2008, qu'ils continuent ou non à en prendre en 2011, ont plus souvent des parcours heurtés que les autres. Ainsi, 35.6 % des élèves qui prenaient des cours en 6^e ont un parcours heurté au collège, contre 21 % en moyenne. **Les difficultés scolaires des élèves qui prennent des cours de soutien en 2008 - discutées dans la section III - semblent donc affecter négativement les parcours scolaires.**

Parmi les trajectoires linéaires, la proportion d'élèves qui accède à une 2^{nde} GT est de 5 points plus élevée pour ceux qui prennent des cours en 3^e⁵¹ que pour ceux qui n'en prennent pas. Lorsqu'ils en prenaient en 2008, cette proportion est 20 points plus faible. On ne constate pas de différence pour les élèves qui prennent à la fois des cours en 6^e et en 3^e. **Les cours de soutien en 3^e semblent favoriser l'accès à la 2^{nde} GT pour les élèves qui ont un parcours linéaire.**

Ce constat est également valable pour ceux dont le parcours est heurté ; la part d'élèves en 2^{nde} GT est

51. Les élèves qui ont des parcours linéaires sont en 3^e en 2011.

15 points plus élevée lorsqu'ils prennent des cours en 2011, qu'ils en ai pris ou non en 6^e. En revanche, les élèves qui n'ont pris des cours qu'en 6^e ne sont pas plus souvent en 2nde GT que ceux qui n'en ont jamais pris.

Les deux dernières lignes ne considèrent pas la dimension *heurté/linéaire* mais uniquement l'accès en seconde GT. On voit alors que les élèves qui ont pris des cours de soutien en 6^e uniquement sont, en proportion, nettement moins susceptibles d'accéder à une 2nde GT que les autres ; les élèves qui ne prennent des cours qu'en 2011 sont légèrement plus fréquemment en 2nde GT.

Tableau 6 – Trajectoires en fin de troisième et soutien scolaire

Trajectoires		Pas de soutien	Soutien en 2011	Soutien en 2008	Soutien en 2008 et 2011	Total	Obs.
Heurté		19,5 %	22,1 %	35,6 %	30,5 %	21,0 %	5972
Linéaire		80,5 %	78,0 %	64,4 %	69,5 %	79,0 %	22412
Linéaire	2 nd e GT	74,7 %	79,3 %	53,1 %	72,9 %	74,1 %	16615
	2 nd e Pro	25,3 %	20,7 %	46,9 %	27,1 %	25,9 %	5797
Heurté	2 nd e GT	21,8 %	36,9 %	21,8 %	34,2 %	24,1 %	1437
	Autres	78,3 %	63,1 %	78,2 %	65,9 %	75,9 %	4535
2nde GT		64,4 %	69,9 %	42,0 %	61,1 %	63,6 %	18052
Autres		36,6 %	30,1 %	58,0 %	38,9 %	36,4 %	10332

Sources : Panel 2007, calculs des auteurs

Lecture : 79 % des élèves entrés en 6^e en 2007 ont pu accéder à la seconde sans redoublement ni contestation de leur choix d'orientation. Parmi les élèves qui n'ont pas pris de cours en 2008 et 2011, 80,5 % ont eu une trajectoire linéaire. Les différences des distributions marginales avec les statistiques de la figure 5 sont dues à l'erreur d'échantillonnage de l'enquête famille. Les pondérations ne modifient pas ces résultats.

En somme, les élèves qui prennent des cours de soutien en 2011 ont généralement plus de chance d'accéder à une 2nde GT, qu'ils aient ou non un parcours heurté. Ceux qui prenaient des cours en 2008 semblent souffrir de leurs difficultés scolaires, ce qui se répercute sur leur orientation. Cette analyse descriptive ne tient toutefois pas compte des différences entre les élèves ; ce que nous nous efforçons de faire dans la section suivante.

2 Effet des cours de soutien sur les trajectoires

Caille (2014) analyse les trajectoires scolaires et identifie un ensemble de caractéristiques observables affectant significativement les parcours des élèves. Dans la suite de cette section, nous présentons un modèle similaire estimant la probabilité d'être dans une trajectoire en focalisant notre attention sur le soutien privé. L'effet des autres variables n'est pas présenté dans ce rapport ; nous invitons le lecteur intéressé à se reporter à Caille (2014) ou Vrignaud (2015) pour une synthèse de ces résultats.

Notre analyse de l'effet du soutien privé sur les trajectoires au collège est moins ambitieuse que celle sur les performances car la méthode employée précédemment ne peut être appliquée à cette question. En effet, la validité de la stratégie de variable instrumentale repose sur l'hypothèse que l'instrument - dans notre cas, le nombre d'entreprises de soutien privé dans la commune et le département - n'a d'effet sur

la variable d'intérêt qu'à travers son impact sur la probabilité de suivre des cours de soutien. Dans le cas des trajectoires, les choix d'orientation des familles dépendent au moins en partie de l'offre des lycées à proximité. Un élève sera moins tenté par un cursus spécifique si celui-ci lui impose d'être interne ou d'allonger considérablement son temps de transport. Or, le nombre d'entreprises de soutien privé a de fortes chances d'être plus élevé là où le nombre de lycéens et d'étudiants est important. Les variables instrumentales sont donc corrélées à l'offre éducative locale ; le nombre d'entreprises de soutien privé n'affecte donc pas que la probabilité d'en suivre, mais aussi, indirectement, les choix d'orientation.

Nous présentons donc une analyse *toutes choses égales par ailleurs* sans conclure sur un effet causal.

Éléments de méthode

Nous modélisons la probabilité de suivre une trajectoire éducative et présentons les trois mêmes indicateurs que dans la section A, à savoir :

- *l'écart de probabilité moyen* d'avoir une trajectoire donnée, entre élèves qui prennent des cours de soutien privé et les autres, sans variable de contrôle^a.
- *l'écart de probabilité moyen* d'avoir une trajectoire donnée, estimé par des modèles *probit* séparés pour chaque groupe, incluant un ensemble de variables de contrôle et pondérés par l'inverse de la probabilité de suivre des cours de soutien^b
- *l'écart de probabilité moyen* d'avoir une trajectoire donnée, en comparant les élèves qui prennent des cours avec ceux qui n'en prennent pas mais dont la probabilité d'en prendre est la plus proche.

Les variables de contrôle introduites dans ces modèles sont presque identiques à celles utilisées dans la section A, c'est-à-dire : le genre de l'élève, l'éducation des parents, leur origine géographique, la CSP du chef de famille, la structure de la famille, le nombre de frères et sœurs, la fréquence de l'aide apportée par les proches et deux polynômes d'ordre deux des tests d'entrée en 6^e en français et mathématiques.

On estime également ces modèles en restreignant l'échantillon aux élèves qui ne prenaient pas de cours en 2008.

a. indiquée dans le tableau par "Différence brute".

b. le propensity score.

A Soutien en 6^e et accès en seconde

Le tableau 7 présente les résultats de ces estimations pour le soutien scolaire en 2008. Sans contrôles, les élèves qui prennent des cours de soutien privés en 6^e ont une probabilité 13 points plus faibles de connaître un parcours heurté par rapport à ceux qui n'en prennent pas. Néanmoins, cette différence est largement expliquée par les caractéristiques socio-démographiques de l'élève. La différence n'est plus que de 4 points dès lors que nous intégrons les variables de contrôles.

Parmi les élèves qui ont une trajectoire linéaire au collège, la probabilité d'être en seconde GT plutôt qu'en seconde Pro est 14 points plus faible lorsqu'ils prennent des cours de soutien privé. Cette différence est entièrement due aux caractéristiques des élèves et de leurs environnements. Dit autrement, à niveau

scolaire, situation familiale,... donnés, la probabilité d'aller en seconde générale est la même qu'ils aient ou non pris des cours de soutien privé en 6^e.

Parmi les élèves qui ont un parcours heurté, il n'y a que très peu d'écart d'accès à la seconde générale (3 points de pourcentage) entre ceux qui prennent des cours et les autres ; cette différence disparaît avec l'inclusion des variables de contrôle.

Au final, la probabilité d'accéder en seconde GT, peu importe le parcours⁵², est 15 points plus faible pour un élève qui prend des cours de soutien. Cependant, la prise en compte de ses caractéristiques fait disparaître cette différence. Les cours de soutien en 6^e ne semblent pas avoir d'impact sur la probabilité d'accéder à une seconde GT. En revanche, les élèves qui prennent des cours risquent davantage d'avoir un parcours heurté au collège.

Tableau 7 – Différence de trajectoires entre élèves suivant ou non des cours en 2008, avec ou sans variables de contrôles

Variable	Méthode	Beta	(SE)	P-value
Linéaires/Heurtés	Différence brute	-0,13***	(0,009)	0,000
	Méthode 1	-0,045***	(0,01)	0,000
	Méthode 2	-0,04***	(0,012)	0,002
Linéaires 2nde GT/ Linéaires 2nde Pro.	Différence brute	-0,142***	(0,012)	0,000
	Méthode 1	-0,011	(0,010)	0,280
	Méthode 2	-0,021	(0,015)	0,147
Heurtés 2nde GT/ Autres heurtés	Différence brute	0,03**	(0,015)	0,047
	Méthode 1	0,026	(0,018)	0,136
	Méthode 2	0,023	(0,023)	0,309
2nde GT/ Autres	Différence brute	-0,15***	(0,01)	0,000
	Méthode 1	-0,016*	(0,01)	0,099
	Méthode 2	-0,010	(0,011)	0,384

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses.

Source : Panel 2007

Lecture : Le groupe d'élèves qui prend des cours en 2008 a une probabilité plus faible d'avoir un parcours linéaire de 13 points. Lorsqu'on inclue les variables de contrôle, cet écart n'est plus que de 4 points.

a) Cours de soutien en 2011

Les résultats des cours de soutien en 2011 sont plus contrastés, comme le montre le tableau 8. Il est important de garder à l'esprit que d'une part, le recours au soutien privé en 2011 peut être le résultat et non la cause de difficultés à passer dans la classe supérieure, à avoir l'orientation souhaitée en premier vœux,... et d'autre part, il n'exclut pas d'avoir pris des cours de soutien entre 2008 et 2011. Ces résultats

52. On ne tient pas compte ici du caractère linéaire ou heurté des trajectoires.

présentent des différences, toutes choses égales par ailleurs, mais ne mesurent pas directement l'effet des cours de soutien privé.

Les élèves qui prennent des cours de soutien en 2011 sont, en proportion, 3.3 points de moins à avoir un parcours linéaire. L'ajout de variable de contrôle ne change pas ce résultat. Il est probable que les cours en 2011 soient effectivement la conséquence d'un redoublement au collège.

Pour les élèves qui ont eu un parcours linéaire, la proportion de ceux qui accèdent à une seconde GT est 4.4 points plus élevée pour les élèves qui prennent des cours que pour les autres. Cet écart est légèrement réduit par l'inclusion des variables de contrôle. Il n'y a pas de différence notable lorsqu'on restreint l'échantillon aux élèves qui n'ont pas pris de cours en 6^e.

Tableau 8 – Différence de trajectoires entre élèves suivant ou non des cours en 2011, avec ou sans variables de contrôles et échantillon restreint ou non

Variable	Méthode	Échantillon complet	Hors soutien en 2008
Linéaires/heurtés	Différence brute	-0,033*** (0,007)	-0,025*** (0,008)
	Méthode 1	-0,033*** (0,009)	-0,015 (0,01)
	Méthode 2	-0,043*** (0,014)	-0,019 (0,012)
Linéaires 2 nd e GT / Linéaires 2 nd e Pro	Différence brute	0,044*** (0,008)	0,045*** (0,009)
	Méthode 1	0,032*** (0,01)	0,028*** (0,01)
	Méthode 2	0,028** (0,013)	0,022* (0,013)
Heurté 2 nd e GT / Autres heurtés	Différence brute	0,140*** (0,016)	0,156*** (0,019)
	Méthode 1	0,073*** (0,02)	0,080*** (0,022)
	Méthode 2	0,11*** (0,029)	0,062** (0,025)
2 nd e GT / Autres	Différence brute	0,05*** (0,008)	0,055*** (0,009)
	Méthode 1	0,031*** (0,009)	0,030*** (0,010)
	Méthode 2	0,023* (0,013)	0,033** (0,013)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses.

Source : Panel 2007

Lecture : Le groupe d'élèves qui prend des cours en 2011 a une probabilité plus faible d'avoir un parcours linéaire de 3.3 points.

Pour les élèves qui ont un parcours heurté, la part de ceux qui accèdent à une seconde générale est 14

points plus élevés dans le cas où ils prennent des cours de soutien. L'ajout des contrôles réduit cet écart mais il reste significatif et important.

Au final, les élèves qui prennent des cours de soutien en 2011 ont plus souvent accès à la seconde GT que les élèves qui n'en prennent pas. La restriction de l'échantillon aux seuls élèves qui ne prenaient pas de cours en 6^e n'affecte que très marginalement les estimations, ne révélant pas de spécificité chez ces derniers.

Vu ces résultats et les estimations des caractéristiques des élèves ayant recours au soutien privé (voir section II), on peut légitimement penser que les cours privés en 2011, c'est-à-dire en 3^e pour les élèves à l'heure - visent à accéder plus facilement à une seconde générale.

À caractéristiques socio-démographiques données, il semble que cette stratégie soit payante même si nos estimations ne peuvent tenir compte des différences de motivation, de l'implication des parents, du projet de l'élève,...

Conclusion

Dans ce rapport, nous proposons une évaluation originale de l'effet des cours privés de soutien sur les performances des élèves et leur orientation en fin de collège, pour nourrir le débat sur le *shadow education* et son impact sur les inégalités scolaires.

Le soutien privé au collège concerne un élève sur dix en 6^e et près d'un sur cinq trois ans plus tard. Ces élèves sont, en moyenne, moins bons que leurs pairs qui n'y ont pas recours et sont dotés de caractéristiques socio-démographiques particulières. Ils sont plus souvent issus de l'immigration, d'origine sociale élevée, ambitionnant d'obtenir le baccalauréat, plutôt enfants unique et citadins, . . . Ce profil spécifique reflète certaines formes d'inégalités plus généralement présentes dans la société. Une grande partie de l'écart de performances (la différence est divisée par quatre) entre ceux qui prennent des cours et les autres est directement imputable à l'environnement familial et éducatif. Toutes choses égales par ailleurs, les élèves qui prennent des cours ont des performances environ 10 % plus faibles que leurs pairs qui n'en prennent pas.

Cependant, ce résultat ne permet pas de conclure sur l'effet des cours de soutien sur les performances. Car même en tenant compte d'un ensemble large de variables de contrôle, il existe d'autres caractéristiques inobservables, qui poussent les uns à prendre des cours et les autres à ne pas y souscrire, et affectent leurs performances. Pour identifier leur effet, nous avons profité des différences de facilité d'accès aux cours de soutien entre les communes où sont scolarisés les élèves. Tout se passe alors comme si nous comparions deux élèves, par ailleurs identiques, dont l'un peut souscrire facilement à des cours de soutien et l'autre avec plus de mal, simplement parce qu'il y a moins d'entreprises offrant ce service dans sa commune. En réalisant cette comparaison sur l'ensemble de l'échantillon, nous identifions l'effet moyen des cours privés de soutien sur la performance des élèves. On montre alors que les cours privés en 6^e améliorent les compétences des élèves en mathématiques et en compréhension de textes ainsi que leur mémoire encyclopédique. Aussi, si ces élèves sont au départ plutôt moins bons, le fait de bénéficier de cours privés de soutien leur permet de progresser. Aux vues de leurs profils spécifiques, le soutien scolaire semble permettre à ceux qui peuvent en bénéficier de rattraper leur retard.

En revanche, leurs capacités de raisonnement semblent en pâtir. Ce résultat nous apparaît être le produit d'un travail davantage axé sur la préparation d'exercices types, la reproduction de contenu pédagogique, . . . , au détriment de résolution de problèmes ou d'autres activités développant la logique. Ce constat persiste lors de la seconde prise d'information trois ans plus tard. De plus, les cours de soutien semblent moins bénéfiques en fin de collège⁵³ qu'en 6^e, même si quelques faiblesses dues à des données lacunaires et une méthodologie moins adaptée rendent ces résultats moins précis.

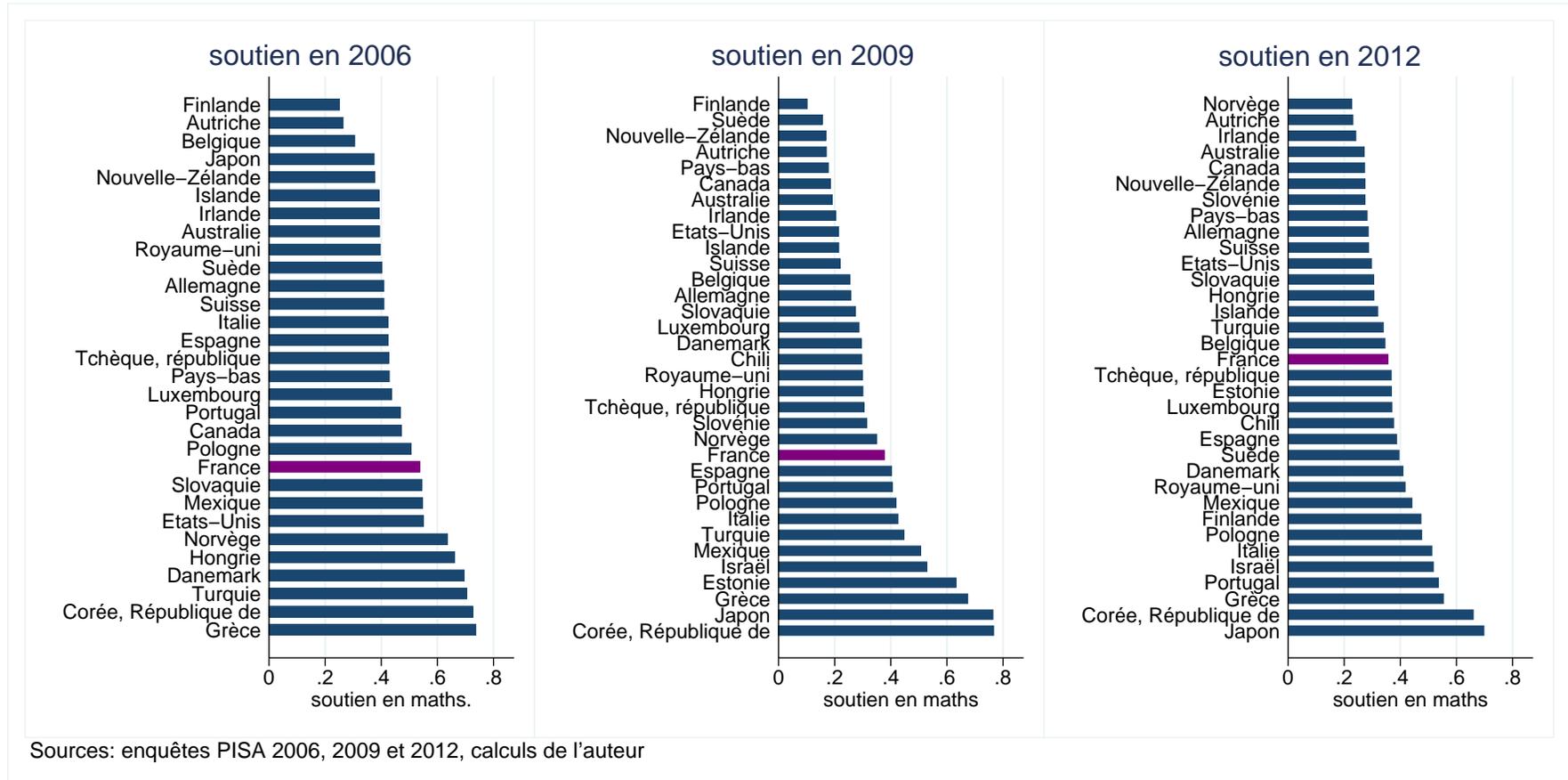
Enfin, si les performances des élèves ne semblent pas améliorées par des cours de soutien en fin de collège, l'accès à la seconde générale et technologique apparaît facilité. En effet, qu'ils aient ou non eu des difficultés dans leur scolarité au collège, les élèves qui suivent des cours privés sont plus fréquemment en seconde générale et technologique que ceux qui n'en suivent pas, toutes choses égales par ailleurs. Cette orientation reflète probablement le caractère prestigieux de la voie générale par rapport à la voie professionnelle que le soutien privé permet à certains d'éviter.

53. Les élèves à l'heure sont en 3^e, ceux qui ont redoublé en 4^e.

Annexes

Annexe A Les cours additionnels en mathématiques d'après PISA

Figure A.1 – Cours additionnels en mathématiques d'après PISA



Notes : le graphique présente le pourcentage d'élèves de chaque pays déclarant suivre au moins une heure de cours, en dehors du programme scolaire, en mathématiques. Ces statistiques sont construites à partir des variables st31q05 en 2006, st32q02 en 2009 et st55q02 en 2012. Les proportions sont obtenues en utilisant les pondérations. Lecture : En 2006, 56 % des élèves âgés de 15 ans en France déclarent suivre au moins une heure de cours de mathématiques en dehors du temps scolaire. Ce cours peut être proposé par l'établissement, la collectivité territoriale ou souscrit de façon privée. Les données ne permettent pas de distinguer les différentes formes de soutien.

Annexe B Disciplines et durée des cours de soutien scolaire dans le panel 2007

Tableau A.1 – Pourcentage d'élèves étudiant chaque discipline dans le cadre de cours de soutien payants

% Soutien	2008	2011
Français	43,7 %	26,6 %
Maths	44,1 %	68,7 %
Physique-chimie	NA	6,7 %
Langue vivante 1	27,9 %	23,8 %
Autre(s) matière(s)	8,1 %	7,9 %
Méthode ou soutien global	42,5 %	22,8 %
Total	9,38 %	14,13 %
Durée moyenne du soutien		
Pas de soutien	90,6 %	85,9 %
1h hebdomadaire	3,7 %	6,3 %
2h hebdomadaire	3,0 %	4,6 %
3h hebdomadaire	0,9 %	1,1 %
4h ou plus	0,9 %	0,9 %
Inconnu	0,9 %	1,2 %
N	32629	29502

Sources : Panel 2007 (MENESR-DEPP), enquêtes famille, calculs des auteurs, pondérations enquêtes famille respectives.

Lecture : en 2008, parmi les 9.38 % d'élèves qui prenaient des cours de soutien privés payants, 43.7 % d'entre eux prenaient des cours de soutien en français et 44.1 % en mathématiques. En 2011, ils sont 68,7 % à prendre des cours de soutien en mathématiques. En 2008, 3,7 % des élèves prennent 1h de cours de soutien, soit 40 % des élèves qui prennent des cours.

Annexe C Interprétation d'un modèle Logit

Définition

Il existe des modélisations statistiques très intuitives et facilement interprétables pour représenter des phénomènes définis en continu (comme par exemple la relation espérance de vie et niveau de richesse d'un pays). Cependant, de nombreux événements sont de nature binaire et ne peuvent être modélisés à partir de ces outils. Le soutien scolaire est une variable aléatoire à deux issues (prendre des cours ou non) et doit donc être représenté à partir d'outils adaptés comme le modèle Logit.

Un modèle Logit estime la probabilité de réalisation d'un événement binaire en fonction d'un ensemble de paramètres dont on mesure l'influence. De façon intuitive, on suppose que la réalisation de l'événement (prendre ou non des cours) dépend directement d'une variable latente, ou cachée, mais dont la distribution de probabilité suit une loi connue. On mesure alors l'impact des variables explicatives sur cette variable latente, ce qui impacte la probabilité de suivre des cours.

Analyse des résultats

L'estimation du modèle Logit s'obtient par la méthode du maximum de vraisemblance. Les coefficients estimés représentent alors l'impact du paramètre sur la variable latente. Le lien direct avec le soutien privé n'est alors pas établi. Pour interpréter directement l'impact du paramètre sur la variable dépendante binaire, on peut présenter d'autres résultats et notamment les rapports de côtes (ou Odd ratios) ou les effets marginaux.

Odds ratios ou rapports de côtes

Les Odds ratios sont un indicateur de la probabilité relative de prendre des cours plutôt que de ne pas en prendre associée aux modalités d'une variable explicative. Ils sont à interpréter *toutes choses égales par ailleurs*, c'est-à-dire en considérant les autres caractéristiques incluses dans le modèle comme données. Par exemple, si on cherche à mesurer l'effet de la taille de la commune sur la probabilité de prendre des cours de soutien, on peut utiliser l'Odd ratio qui sera le rapport entre la probabilité de suivre des cours plutôt que de ne pas en suivre pour un élève vivant à Paris (par exemple) sur la probabilité de suivre des cours plutôt que de ne pas en suivre pour un élève vivant dans une commune de taille moyenne (ici, 50 à 100 000 habitants). L'interprétation de l'Odd ratio est assez intuitive et lisible. S'il est supérieur à 1, le paramètre augmente la probabilité de prendre des cours privés. S'il est inférieur à 1, il la diminue. La distribution logistique a l'avantage de fournir directement les Odds ratios attachés au paramètre en calculant l'exponentiel du coefficient. Comme pour toute estimation, on estime les erreurs standards et on teste la significativité des coefficients.

Annexe D Le biais dans les modèles à valeur ajoutée

Le recours au soutien scolaire n'est pas indépendant du niveau à l'entrée en 6^e, il est même en grande partie déterminé par ce dernier. Aussi, inclure cette caractéristique dans un modèle visant à estimer l'effet des cours de soutien sur les performances produit des résultats biaisés. Pour le montrer, nous reprenons le cadre conceptuel de l'article de thèse de [Bouguen \(2015\)](#)⁵⁴ en l'adaptant au contexte de cet article.

Supposons que les performances d'un élève puissent être représentées par une équation à quatre composantes additives :

$$Y_{i1} = \beta T_i + \mu_i + \tau_{i1} + \varepsilon_{i1} \quad (8)$$

où T_i indique si l'élève suit des cours de soutien, μ_i est son niveau de performance naturel fixe et τ_{it} sa capacité à progresser. Enfin, ε_{it} est l'erreur de mesure, indépendante de toute caractéristique observable ou non. β est donc l'effet moyen théorique que nous cherchons à identifier dans ce papier. Comme l'accès au soutien scolaire n'est pas aléatoirement assigné, il y a de fortes chances pour que la capacité de l'élève à progresser soit corrélée à T et aussi à son niveau de performance fixe μ_i . Décomposons alors cette capacité à progresser de la façon suivante :

$$\tau_{i1} = \rho \mu_i + \gamma_{i1} \quad (9)$$

où ρ indique la corrélation entre le niveau de performance fixe et la capacité à progresser et γ_{it} est la capacité à progresser de l'élève qui ne dépend pas de lui. Typiquement, γ_{it} va être affecté par les cours privés de soutien mais aussi l'aide de ses parents, les conditions dans lesquelles il étudie, ... En insérant la seconde équation dans la première on obtient immédiatement :

$$Y_{i1} = \beta T_i + (1 + \rho) \mu_i + \gamma_{i1} + \varepsilon_{i1} \quad (10)$$

On définit alors de façon similaire le niveau de performance à l'entrée en 6^e :

$$Y_{i0} = \mu_i + \varepsilon_{i0} \quad (11)$$

où τ_{i0} est normalisé à 0 pour simplifier. Il vient alors immédiatement :

$$\begin{aligned} Y_{i1} &= \beta T_i + (1 + \rho) \mu_i + \gamma_{i1} + \varepsilon_{i1} \\ &= \beta T_i + (1 + \rho) Y_{i0} + \gamma_{i1} + \varepsilon_{i1} - (1 + \rho) \varepsilon_{i0} \end{aligned} \quad (12)$$

Cette dernière équation est très compliquée à estimer car les performances initiales Y_{i0} sont corrélées avec les chocs de la période 0 ε_{i0} qui sont inobservables. Cette corrélation avec les résidus induit un biais qui atténue l'estimation de $(1 + \rho)$ ainsi que celle du paramètre d'intérêt β .

On peut toutefois estimer l'équation (12) en contraignant certains paramètres pour permettre l'identification : ainsi, un modèle à valeur ajoutée simple qui estime l'équation :

$$\begin{aligned} Y_{i1} &= \beta T_i + Y_{i0} + \gamma_{i1} + \varepsilon_{i1} - \varepsilon_{i0} \\ \iff Y_{i1} - Y_{i0} &= \beta T_i + \gamma_{i1} + \varepsilon_{i1} - \varepsilon_{i0} \end{aligned} \quad (13)$$

54. Nous remercions Marc Gurgand de nous en avoir fait part

consiste à contraindre $\rho = 0$. Dans ces conditions, et si la progression indépendante de l'élève n'est pas corrélée avec le soutien scolaire (i.e. $E(\gamma_{i1}|T) = 0$), alors un modèle de régression standard (OLS) estime correctement le paramètre β . Contraindre $\rho = 0$ revient à supposer que la progression des élèves ne dépend pas de leur niveau initial. Autrement dit, sans intervention extérieure, les bons élèves et les moins bons élèves progressent au même rythme selon une trajectoire parallèle.

On peut également vouloir estimer l'équation (12) en relaxant la contrainte sur ρ en estimant le modèle de régression suivant :

$$Y_{i1} = \beta T + \underbrace{\lambda}_{1+\rho} Y_{i0} + \gamma_{i1} + \underbrace{\epsilon_{i1}}_{\epsilon_{i1}-\epsilon_{i0}} \quad (14)$$

Mais comme les performances initiales Y_{i0} sont corrélées avec les chocs de la période 0 ϵ_{i0} (i.e. $E(Y_{i0}|\epsilon_{i0}) \neq 0$), l'estimation de λ sera inférieure au vrai paramètre.

En utilisant la formule du biais de variable omise⁵⁵, où ϵ_{i0} est considérée comme la variable omise, Bouguen (2015) montre que l'estimation de β par ce modèle : $\hat{\beta}$ a pour espérance :

$$E(\hat{\beta}) = \beta + \lambda \frac{r(Y_{i0}, T) \times V_{\epsilon_0} S_{\epsilon_0}}{1 - r^2(Y_{i0}, T) \times S_{T_i}} \quad (15)$$

Où $r(Y_{i0}, T)$ est la corrélation entre les performances initiales et le soutien scolaire, V_{ϵ_0} la variance de l'erreur de mesure initiale et S_{ϵ_0} son écart type, et enfin, S_{T_i} l'écart type du groupe de soutien.

Si on fait l'hypothèse raisonnable que $1 + \rho \geq 0$ ⁵⁶, alors le signe du biais est déterminé par le signe de la corrélation $r(Y_{i0}, T)$, qui est négative dans notre cas. Dès lors, on sait que l'évaluation des modèles à valeur ajoutée fournissent une estimation conservatrice sous-estimant le véritable effet des cours privés. Comme nous ne voulons pas exagérer l'estimation de l'effet des cours de soutien, nous optons pour ce dernier modèle et gardons à l'esprit le fait que nos estimations sous-estiment le véritable impact.

55. Voir (Angrist et Pischke, 2009, p.44) ou (Cameron et Trivedi, 2005, sec. 4.7.4).

56. En théorie. Nos estimations, que nous savons inférieures au coefficient théorique, sont toutes positives ce qui valide cette hypothèse.

Annexe E Distribution des instruments

Figure A.2 – Histogramme des variables instrumentales en 2008

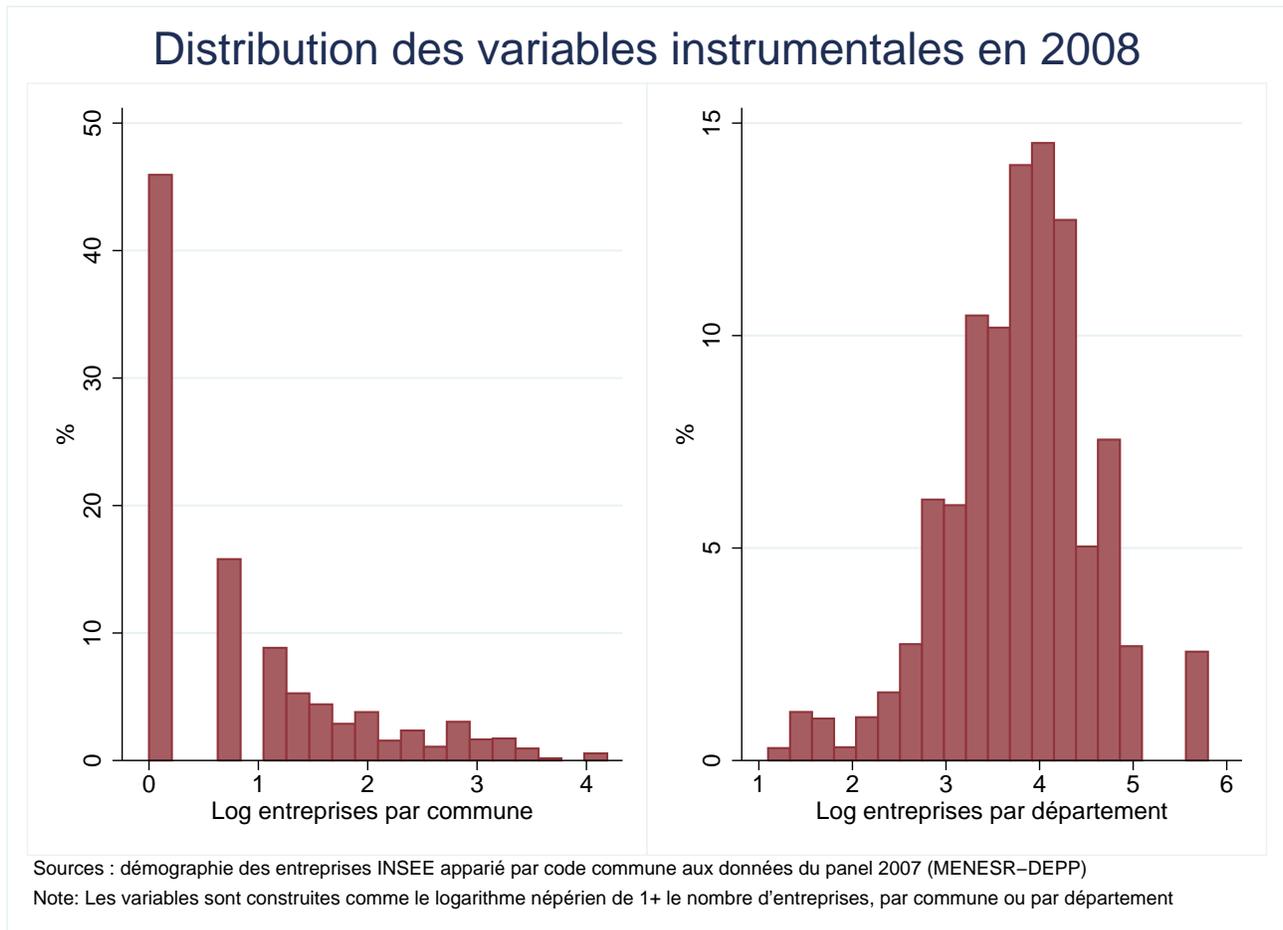


Figure A.3 – Entreprises privées dans les communes de France fournissant du soutien scolaire privé en 2008

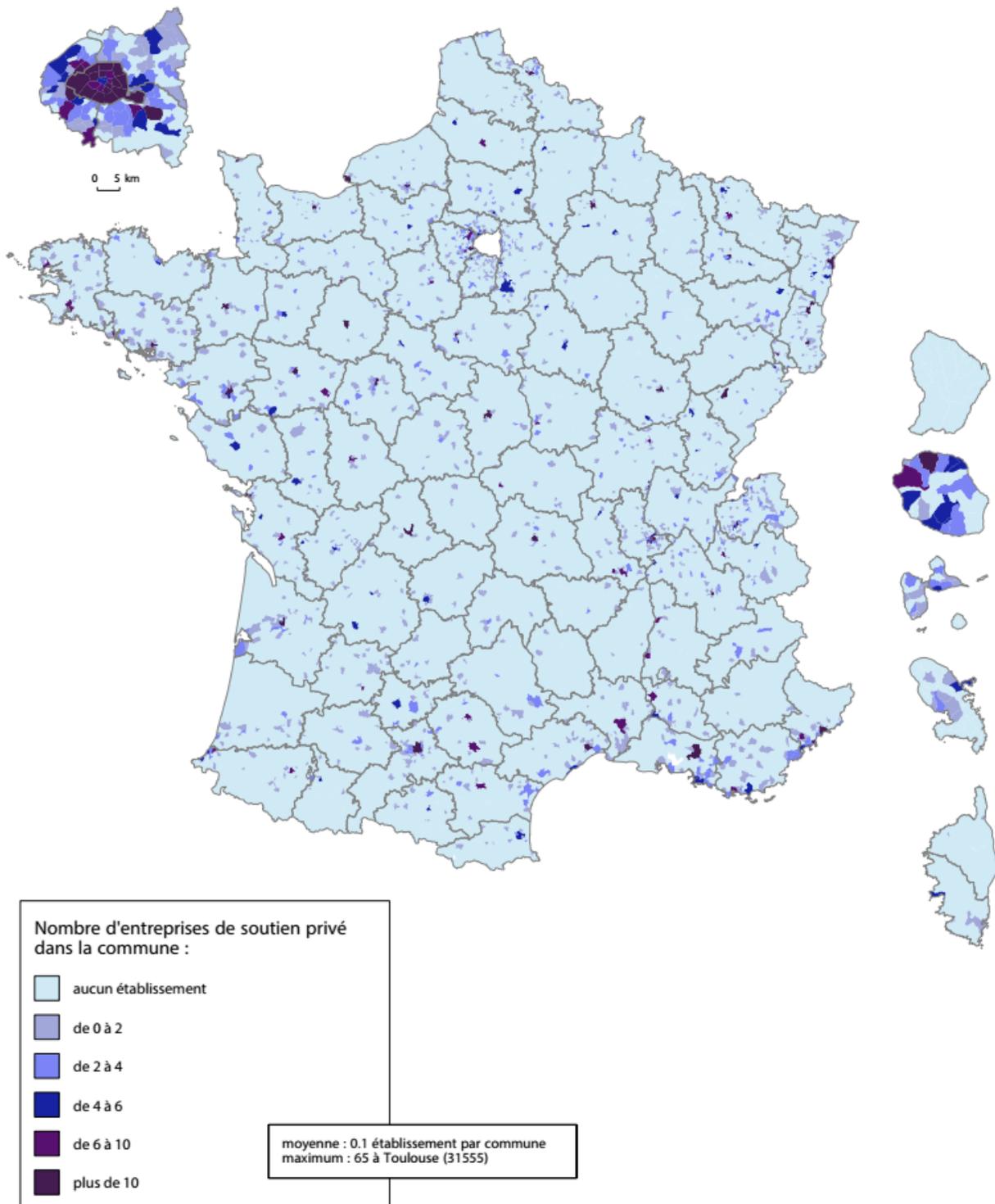
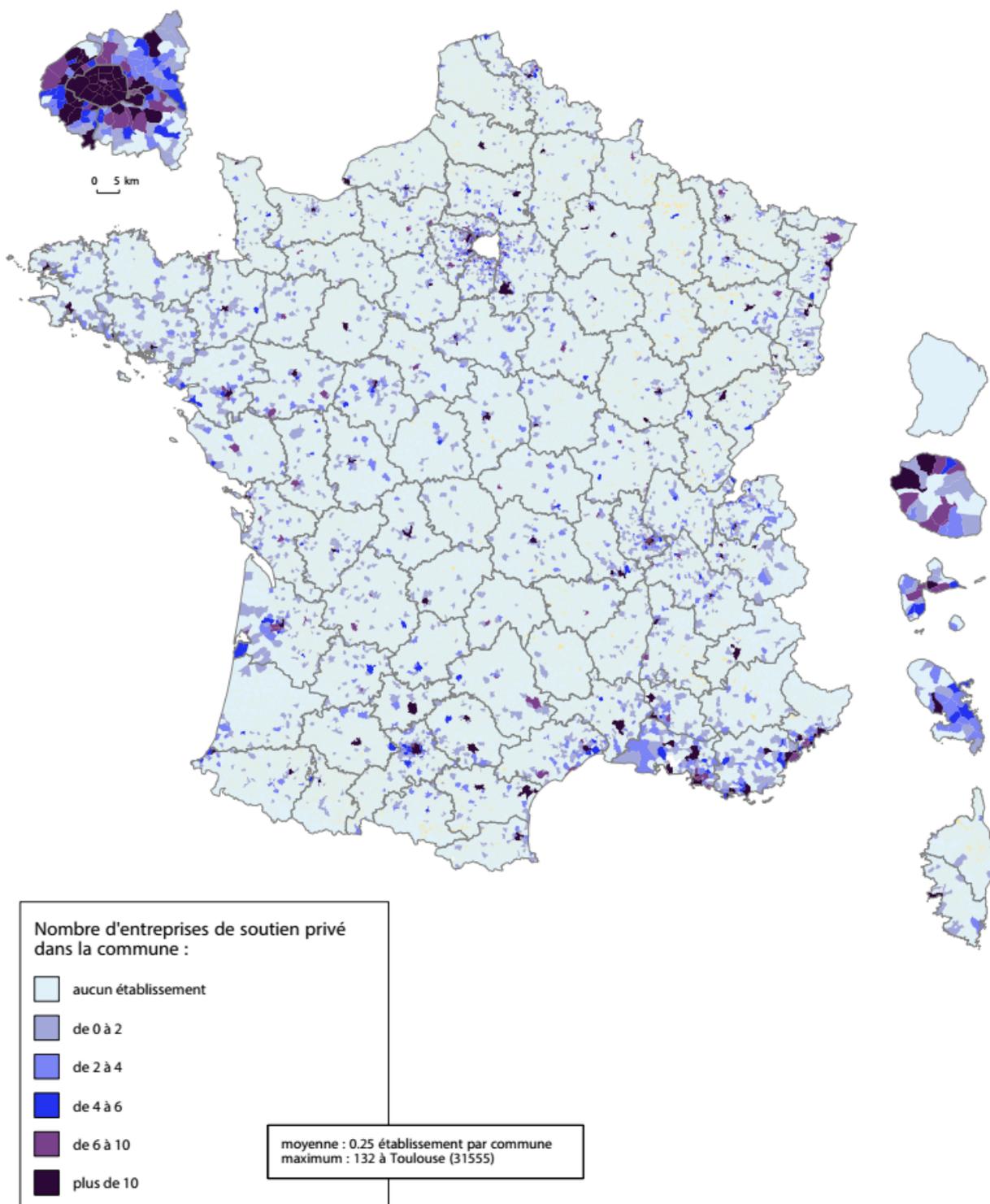


Figure A.4 – Entreprises privées dans les communes de France fournissant du soutien scolaire privé en 2011



Annexe F Estimation de la première équation des modèles à variables instrumentales en 2008

Le tableau ci-dessous présente les variantes des équations de sélection des modèles 2SLS du tableau 4. Les variables choisies sont toutes significatives (au moins une des modalités). Les instruments sont significativement différents de 0, de signe et de magnitude attendus. Puisqu'il s'agit d'une variable en logarithme, les coefficients s'interprètent en semi-élasticité. Ainsi, une augmentation d'1 % du nombre d'entreprise de soutien privé dans la commune augmente la probabilité de prendre des cours de soutien pour un élève de 6^e de 0.97 à 1.5 point de pourcentage, toutes choses égales par ailleurs. Avec une probabilité des modalités de références de 9.0 %⁵⁷, l'effet obtenu n'est pas négligeable. Le modèle est globalement significatif comme l'indique la F-stat globale. Le test de significativité jointe des instruments est systématiquement différent de 0 et supérieur à 10 pour toutes les spécifications.

Tableau A.2 – Équation de sélection (*first stage*) des modèles 2SLS en 2008

		(OLS)		(OLS)		(OLS)	
		$\hat{\beta}$	(\hat{SE})	$\hat{\beta}$	(\hat{SE})	$\hat{\beta}$	(\hat{SE})
Instruments	Par commune	0.0126***	(0.0023)	0.0097***	(0.0024)	0.0153***	(0.0037)
	ln(entreprises + 1)	0.0143***	(0.0030)	0.0115***	(0.0029)	0.0096***	(0.0032)
Caractéristiques de la ville :	Rurale	-0.0020	(0.0086)	-0.0018	(0.0084)	-0.0203**	(0.0099)
	ref : Commune de 50 000						
	≤ 5000 hab.	-0.0108	(0.0082)	-0.0078	(0.0080)	-0.0186**	(0.0085)
	à 100 000 hab.						
	5 à 10000 hab.	0.0044	(0.0088)	0.0057	(0.0087)	-0.0004	(0.0087)
	10 à 20 000 hab	-0.0016	(0.0089)	-0.0004	(0.0086)	-0.0028	(0.0087)
	50 à 100 000hab	-0.0010	(0.0091)	0.0015	(0.0089)	0.0040	(0.0089)
	100 à 200 000 hab	0.0067	(0.0105)	0.0095	(0.0105)	0.0142	(0.0105)
200 000 à 2 million hab	0.0045	(0.0078)	0.0059	(0.0076)	0.0125	(0.0077)	
	Paris	0.0424***	(0.0091)	0.0351***	(0.0091)	0.0347***	(0.0092)
Eval. 6^e	Français	-0.0283***	(0.0026)	-0.0276***	(0.0029)	-0.0266***	(0.0029)
	Français ²	0.0003	(0.0019)	0.0002	(0.0019)	0.0003	(0.0020)
	Maths	-0.0302***	(0.0026)	-0.0353***	(0.0029)	-0.0354***	(0.0029)
	Maths ²	-0.0044**	(0.0019)	-0.0044**	(0.0019)	-0.0033*	(0.0019)

Suite page suivante

57. Obtenue à partir de la commande "margins, grand" de Stata.

Suite de la page précédente

Sexe	Fille	-0.0084**	(0.0039)	-0.0093**	(0.0039)
Éducation de la mère ref. Bac	Inconnue	0.0211**	(0.0099)	0.0216**	(0.0101)
	Au plus primaire	-0.0270***	(0.0079)	-0.0266***	(0.0080)
	collège	0.0047	(0.0071)	0.0028	(0.0071)
	CAP BEP	0.0077	(0.0057)	0.0068	(0.0057)
	Sup. court	-0.0058	(0.0058)	-0.0055	(0.0058)
	Sup. long	0.0055	(0.0072)	0.0046	(0.0072)
Éducation du père ref. Bac	Inconnue	-0.0038	(0.0085)	-0.0024	(0.0086)
	Au plus primaire	-0.0058	(0.0089)	-0.0091	(0.0090)
	Collège	-0.0131	(0.0081)	-0.0151*	(0.0082)
	CAP BEP	-0.0128**	(0.0064)	-0.0137**	(0.0064)
	Sup. court	-0.0168**	(0.0071)	-0.0173**	(0.0072)
	Sup. long	0.0008	(0.0083)	-0.0026	(0.0085)
Origine géographique des parents ⁵⁸ Ref : deux parents nés en France	Famille mixte dont 1 parent français	0.0171***	(0.0061)	0.0193***	(0.0062)
	Famille mixte étrangère	0.0757***	(0.0190)	0.0658***	(0.0201)
	Europe	0.0039	(0.0174)	0.0047	(0.0176)
	Moyen Orient	0.0393*	(0.0205)	0.0535**	(0.0213)
	Maghreb	0.0653***	(0.0123)	0.0758***	(0.0129)
	Reste Afrique	0.0067	(0.0173)	0.0230	(0.0186)
	Amérique	0.0521	(0.0422)	0.0494	(0.0423)
Asie	0.0766***	(0.0261)	0.0693***	(0.0269)	
Statut boursier	Non boursier	0.0214***	(0.0053)	0.0213***	(0.0053)
Redoublement	A déjà redoublé	0.0152**	(0.0071)	0.0139*	(0.0072)
PCS chef de famille Ref : employé(e)	Agriculteur	0.0038	(0.0110)	0.0013	(0.0111)
	Commerçant, artisans, etc.	0.0564***	(0.0081)	0.0529***	(0.0082)
	Cadres et prof. libérales	0.0191***	(0.0074)	0.0168**	(0.0076)

Suite page suivante

Suite de la page précédente

	Prof. intermédiaires		-0.0010	(0.0066)	-0.0011	(0.0067)
	Ouvriers		-0.0124**	(0.0059)	-0.0122**	(0.0061)
	Chômeurs et inactifs		-0.0435**	(0.0195)	-0.0505**	(0.0201)
	Instituteurs, professeurs et scientifiques		0.0001	(0.0117)	-0.0039	(0.0117)
Taille de la fratrie ref : enfant unique	1 frère/sœur		-0.0211***	(0.0076)	-0.0207***	(0.0077)
	2 frères et sœurs		-0.0319***	(0.0077)	-0.0307***	(0.0078)
	3 frères et sœurs		-0.0368***	(0.0089)	-0.0360***	(0.0090)
	4 frères et sœurs		-0.0280**	(0.0109)	-0.0311***	(0.0110)
	5 ou plus		-0.0572***	(0.0114)	-0.0548***	(0.0116)
Structure familiale ref : 2 parents	Alternée ou 1 parent		0.0088	(0.0064)	0.0090	(0.0065)
	Recomposée		-0.0054	(0.0071)	-0.0053	(0.0072)
	Autres		-0.0096	(0.0187)	-0.0063	(0.0191)
Fréquence de l'aide ref : jamais ou presque	De temps en temps		-0.0113*	(0.0059)	-0.0128**	(0.0060)
	Régulièrement		0.0057	(0.0058)	0.0037	(0.0058)
Log revenu de la commune	Revenu total				-0.0108***	(0.0034)
	Revenu moyen				-0.0245	(0.0242)
	Constante	0.0285**	(0.0124)	0.0475***	(0.0174)	0.4661* (0.2495)
	F-stat Globale	69.894		21.388		19.102
	R^2 ajusté	.0397		.052		.050
	F-test instruments ⁵⁹	19.273		12.5374		9.604
	Observations	25661		25176		24057

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses ajustées par établissement (cluster).

59. Obtenus sur le modèle des mathématiques. Les 1st stages étant les mêmes pour tous les modèles, le choix de la variable dépendante est sans importance

Annexe G Estimation des différences de performances en 2011

Tableau A.3 – Différence de performances en 2011 entre élèves suivant ou non des cours, avec ou sans variables de contrôle, performances passées et soutien passé.

Variable	Méthode	Variables en niveau		Modèle à valeur ajoutée		
		Tout le monde	Hors soutien en 2008	Tout le monde	Hors soutien en 2008	
Lexique 6		$E(Y T = 1) - E(Y T = 0)$	-0,1581*** (0,027)	-0,0976*** (0,0296)		
	Inverse probability weighting RA		-0,0625* (0,035)	-0,0269 (0,038)	-0,0006 (0,03)	0,0114 (0,032)
	Propensity score matching		-0,0985* (0,057)	-0,0442 (0,058)	-0,0505 (0,041)	-0,0551 (0,047)
Phrases lacunaires		$E(Y T = 1) - E(Y T = 0)$	-0,1542*** (0,022)	-0,1139*** (0,024)		
	Inverse probability weighting RA		-0,1182*** (0,027)	-0,1181*** (0,029)	-0,0743*** (0,023)	-0,0728*** (0,025)
	Propensity score matching		-0,1535*** (0,044)	-0,194*** (0,05)	-0,1255*** (0,038)	-0,104*** (0,039)
Maths		$E(Y T = 1) - E(Y T = 0)$	-0,3614*** (0,025)	-0,3186*** (0,027)		
	Inverse probability weighting RA		-0,2438*** (0,03)	-0,2204*** (0,033)	-0,096*** (0,025)	-0,0864*** (0,027)
	Propensity score matching		-0,3054*** (0,051)	-0,3006*** (0,049)	-0,1491*** (0,044)	-0,1232** (0,052)
Compréhension		$E(Y T = 1) - E(Y T = 0)$	-0,1554*** (0,02)	-0,1042*** (0,021)		

Suite page suivante

Suite de la page précédente

	Inverse probability weighting RA	-0,0685*** (0,024)	-0,0432* (0,026)	-0,0464** (0,023)	-0,0312 (0,024)
	Propensity score matching	-0,152*** (0,034)	-0,112*** (0,038)	-0,0582* (0,033)	-0,0689* (0,036)
Raisonnement	$E(Y T = 1) - E(Y T = 0)$	-0,2779*** (0,022)	-0,2479*** (0,024)		
	Inverse probability weighting RA	-0,2057*** (0,026)	-0,1771*** (0,028)	-0,1275*** (0,023)	-0,1114*** (0,025)
	Propensity score matching	-0,2234*** (0,039)	-0,2321*** (0,043)	-0,1598*** (0,036)	-0,1518*** (0,037)

Lecture : Le groupe d'élèves qui prend des cours de soutien en 2011 a des performances en lexique plus faibles de 0,16 écart-type que le groupe qui n'en prend pas.

En retirant les élèves qui prenaient des cours en 2008, la différence est de 10 % d'un écart type.

Toutefois, lorsqu'on retire les différences dues aux caractéristiques observables de l'élève et de sa famille, l'écart de performances n'est plus significativement différent de 0.

Les colonnes de droite contrôlent également par le niveau d'acquis en 2008 dans la dimension.

Annexe H Estimations par régressions quantiles en 2011

Figure A.5 – Estimations par régressions quantiles en 2011

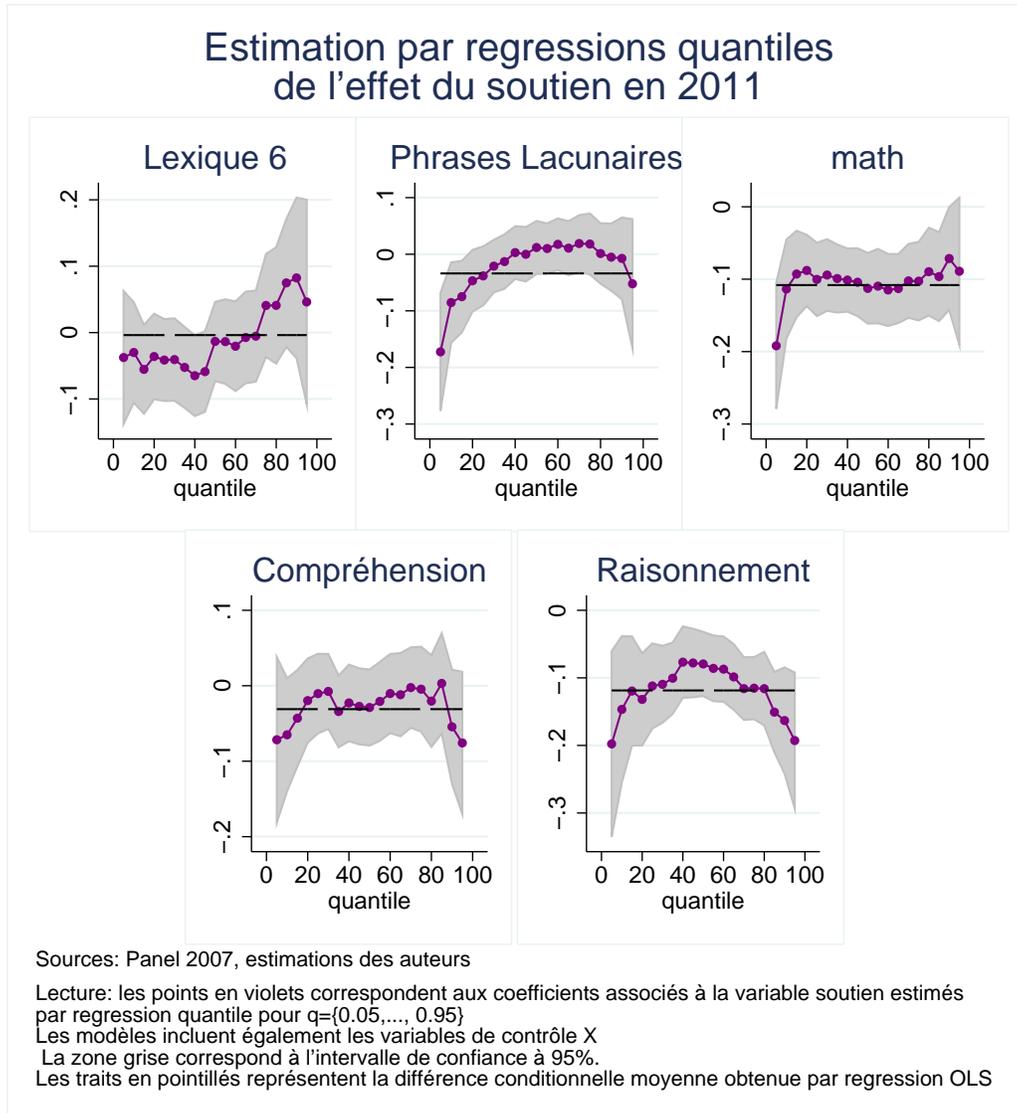
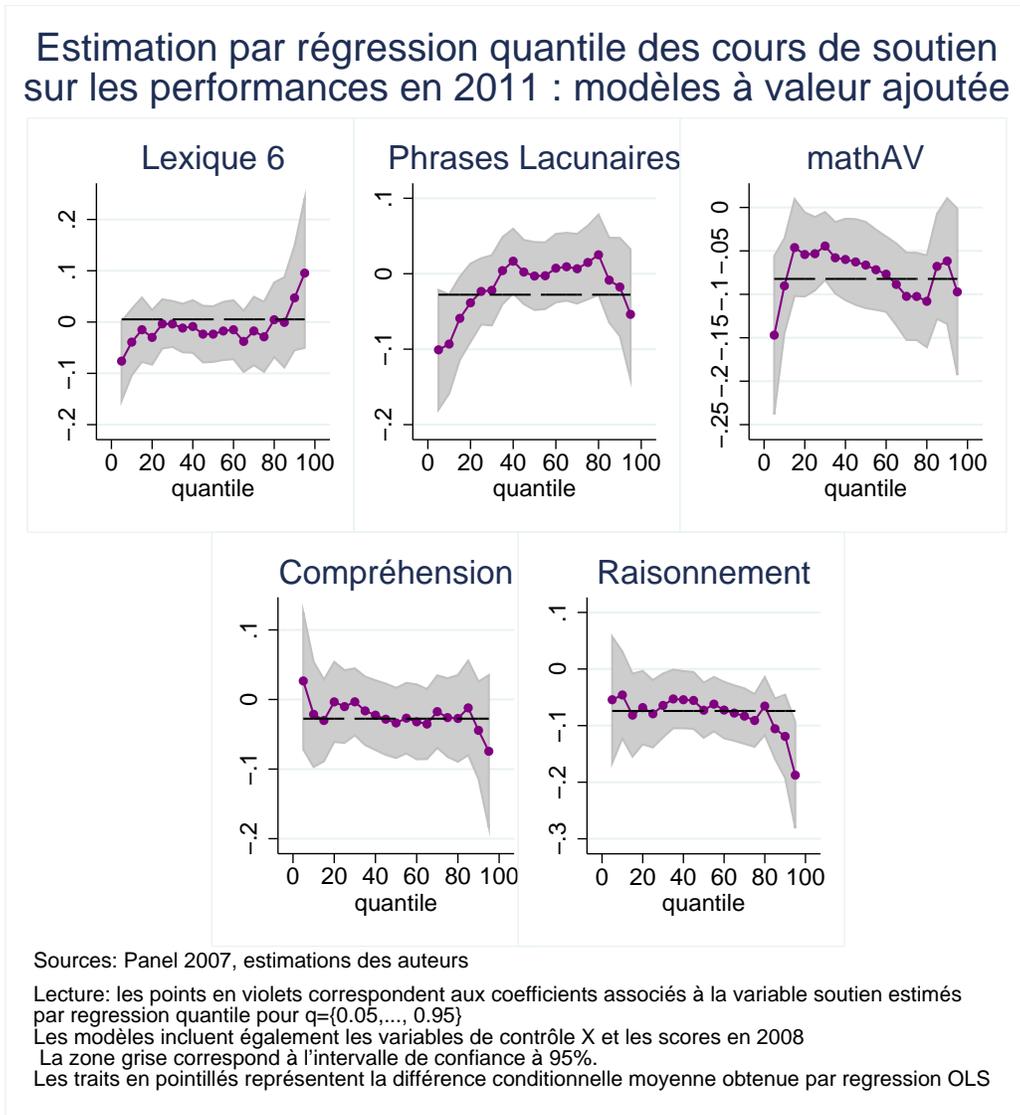


Figure A.6 – Estimations par régressions quantiles en 2011 en tenant compte du niveau de 2008



Annexe I Équation de sélection (*first stage*) en 2011, modèle à valeur ajoutée

Tableau A.4 – Équation de sélection (*first stage*) des modèles 2SLS en 2011

		(OLS)		(OLS)		(OLS)	
		$\hat{\beta}$	(\hat{SE})	$\hat{\beta}$	(\hat{SE})	$\hat{\beta}$	(\hat{SE})
Instruments	Par commune	0.0190***	(0.0027)	0.0156***	(0.0029)	0.0087*	(0.0050)
ln(entreprises + 1)	Par département	0.0272***	(0.0037)	0.0247***	(0.0041)	0.0136***	(0.0043)
Caractéristiques de la commune :	Rurale	-0.0189*	(0.0108)	-0.0198*	(0.0118)	-0.0125	(0.0138)
ref : Commune de 50 000	≤ 5000 hab.	-0.0243**	(0.0104)	-0.0303***	(0.0113)	-0.0285**	(0.0119)
à 100 000 hab.	5 à 10000 hab.	-0.0027	(0.0111)	-0.0008	(0.0122)	-0.0005	(0.0123)
	10 à 20 000 hab	-0.0151	(0.0113)	-0.0230*	(0.0123)	-0.0246**	(0.0122)
	50 à 100 000hab	-0.0161	(0.0116)	-0.0262**	(0.0123)	-0.0269**	(0.0124)
	100 à 200 000 hab	0.0338**	(0.0132)	0.0287**	(0.0142)	0.0254*	(0.0143)
	200 000 à 2 million hab	0.0077	(0.0103)	0.0038	(0.0112)	0.0030	(0.0114)
	Paris	0.0487***	(0.0120)	0.0435***	(0.0134)	0.0171	(0.0141)
Eval. 6^e	Français	0.0017	(0.0038)	-0.0080*	(0.0045)	-0.0069	(0.0045)
	Français ²	-0.0155***	(0.0024)	-0.0114***	(0.0027)	-0.0110***	(0.0027)
	Maths	-0.0224***	(0.0049)	-0.0347***	(0.0055)	-0.0353***	(0.0055)
	Maths ²	-0.0176***	(0.0025)	-0.0158***	(0.0028)	-0.0149***	(0.0028)
Log math⁶⁰		-0.0159***	(0.0041)	-0.0200***	(0.0046)	-0.0203***	(0.0047)
Sexe	Fille			-0.0061	(0.0059)	-0.0071	(0.0059)
	Inconnue			-0.0229*	(0.0136)	-0.0176	(0.0138)
	Au plus primaire			-0.0409***	(0.0118)	-0.0356***	(0.0120)
Éducation de la mère	Collège			-0.0226**	(0.0104)	-0.0188*	(0.0105)
ref. Bac	CAP BEP			-0.0163*	(0.0085)	-0.0155*	(0.0086)
	Sup. court			0.0109	(0.0096)	0.0094	(0.0097)

Suite page suivante

Suite de la page précédente

	Sup. long	-0.0008	(0.0112)	-0.0053	(0.0114)
Éducation du père ref. Bac	Inconnue	0.0023	(0.0124)	0.0003	(0.0125)
	Au plus primaire	-0.0150	(0.0132)	-0.0139	(0.0134)
	Collège	-0.0064	(0.0119)	-0.0093	(0.0119)
	CAP BEP	-0.0071	(0.0096)	-0.0086	(0.0097)
	Sup. court	-0.0082	(0.0118)	-0.0067	(0.0119)
	Sup. long	0.0077	(0.0132)	0.0086	(0.0134)
	Origine géographique des parents Ref : deux parents nés en France	Famille mixte dont 1 parent français	0.0170*	(0.0096)	0.0227**
Famille mixte étrangère		0.0555**	(0.0280)	0.0528*	(0.0295)
Europe		-0.0233	(0.0253)	-0.0211	(0.0253)
Moyen-orient		-0.0198	(0.0229)	-0.0242	(0.0228)
Maghreb		0.0307*	(0.0174)	0.0395**	(0.0182)
Reste Afrique		-0.0110	(0.0268)	0.0170	(0.0290)
Amérique		0.1060	(0.0660)	0.0849	(0.0642)
Asie		-0.0117	(0.0307)	-0.0308	(0.0290)
Statut boursier	Non boursier	0.0188**	(0.0074)	0.0143*	(0.0075)
Redoublement	rdb	-0.0394***	(0.0084)	-0.0399***	(0.0086)
PCS chef de famille Ref : employé(e)	Agriculteur	0.0191	(0.0169)	0.0204	(0.0169)
	Commerçant, artisan, etc.	0.0819***	(0.0121)	0.0792***	(0.0122)
	Cadres et prof. libérales	0.0537***	(0.0121)	0.0456***	(0.0122)
	Prof intermédiaires	0.0146	(0.0101)	0.0141	(0.0102)
	Ouvriers	-0.0135	(0.0087)	-0.0118	(0.0088)
	Chômeurs et inactifs	-0.0120	(0.0326)	0.0025	(0.0352)
	Instituteurs, professeurs et scientifiques	-0.0076	(0.0180)	-0.0127	(0.0180)
Taille de la fratrie ref : enfant unique	Un(e) frère/sœur	-0.0279	(0.0172)	-0.0305*	(0.0174)
	2 frères et sœurs	-0.0538***	(0.0172)	-0.0574***	(0.0174)
	3 frères et sœurs	-0.0822***	(0.0179)	-0.0838***	(0.0181)

Suite page suivante

Suite de la page précédente

	4 frères et sœurs			-0.0666***	(0.0196)	-0.0721***	(0.0197)
	5 ou plus			-0.0839***	(0.0198)	-0.0857***	(0.0200)
Structure familiale ref : 2 parents	Alternée ou 1 parent			0.0016	(0.0084)	-0.0010	(0.0085)
	Recomposée			0.0030	(0.0108)	0.0009	(0.0108)
	Autre			0.0097	(0.0210)	0.0140	(0.0213)
Fréquence de l'aide ref : jamais ou presque	De temps en temps			0.0024	(0.0061)	0.0013	(0.0061)
	Régulièrement			0.0616***	(0.0073)	0.0597***	(0.0074)
Log revenu de la commune	Revenu total					0.0072	(0.0051)
	Revenu moyen					0.0786***	(0.0145)
	Constante	0.0181	(0.0156)	0.0515*	(0.0291)	-0.8068***	(0.1656)
	F-stat	54.930		17.922		15.950	
	R ² ajusté	.0306362		.0659064		.0631958	
	F-test instruments	57.3341		35.6784		5.85485	
	Observations	24173		18859		18046	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Erreurs standards robustes entre parenthèses ajustées par établissement (cluster).

Sources : Panel d'élèves entrés en 6^e en 2007, Estimations par OLS sur données non pondérées.

Annexe J Estimations par variables instrumentales en 2011

Tableau A.5 – Estimation de l'effet des cours particuliers sur les performances des élèves en 2011 par variable instrumentale et Heckit

	OLS	2SLS (1)	2SLS (2)	2SLS (3)	Probit-2SLS (1)	Pr.-2SLS (2)	Pr.-2SLS (3)	Heckit (1)	Heckit (2)	Heckit (3)
lexique 6	-0.065** (0.027)	1.428*** (0.335)	0.744** (0.379)	-1.250 (1.037)	0.795*** (0.284)	0.061 (0.234)	-0.597** (0.294)	1.072*** (0.122)	0.161 (0.594)	-0.199* (0.111)
Phrases lacunaires	-0.068*** (0.021)	0.201 (0.233)	-0.442 (0.279)	-1.454* (0.821)	-0.055 (0.212)	-0.166 (0.178)	-0.256 (0.219)	0.022 (0.062)	-0.066 (0.098)	-0.087 (0.139)
Maths	-0.167*** (0.021)	0.231 (0.254)	-0.258 (0.298)	-0.825 (0.780)	0.061 (0.222)	-0.120 (0.185)	-0.198 (0.227)	0.086 (0.085)	0.038 (0.110)	0.042 (0.124)
Compréhension	-0.069*** (0.021)	-0.270 (0.238)	-0.748** (0.308)	-2.262** (1.022)	-0.355* (0.214)	-0.364* (0.186)	-0.389 (0.241)	-0.127 (0.149)	-0.166 (0.228)	-0.126 (0.256)
Raisonnement	-0.163*** (0.023)	-1.065*** (0.289)	-1.177*** (0.358)	-2.089** (1.003)	-1.071*** (0.262)	-0.545** (0.212)	-0.253 (0.244)	-1.267*** (0.074)	-0.268 (0.177)	-0.189** (0.088)
contrôles	Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui
revenu commune	non	Non	Non	Oui	Non	Non	Oui	Non	Non	Oui
Observation	14234	18387	14814	14234	18387	14814	14234	18387	14814	14234

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Erreurs standards robustes entre parenthèses ajustées par établissement (cluster).

Tous les modèles utilisent comme instrument le nombre d'entreprises enregistrées dans la démographie des entreprises (INSEE) en 2011. Ces bases ont été appariées au Panel 2007 sur la base du code commune de l'établissement scolaire où étudie l'élève l'année 2010-2011. Les spécifications contiennent donc le nombre d'entreprises de soutien scolaire dans la commune et dans le département exprimé en logarithme.

Les différentes colonnes sont identiques aux modèles estimés en 2008.

60. Log du revenu total imposable des ménages et log du revenu moyen par ménage

Tableau A.6 – Estimation de l'effet des cours particuliers sur les performances des élèves en 2011 par variable instrumentale et Heckit

		OLS	2SLS (1)	2SLS (2)	2SLS (3)	Pr-2SLS (1)	Pr-2SLS (2)	Pr-2SLS (3)	Heckit (1)	Heckit (2)	Heckit (3)
lexique 6	Soutien	-0.061*** (0.023)	0.506* (0.273)	0.184 (0.327)	-0.777 (0.882)	0.017 (0.237)	-0.194 (0.209)	-0.453* (0.259)	0.508 (0.563)	-0.112 (0.111)	-0.224** (0.090)
	Score 2008	0.676*** (0.012)	0.722*** (0.011)	0.675*** (0.012)	0.675*** (0.012)	0.727*** (0.011)	0.675*** (0.012)	0.675*** (0.012)	0.722*** (0.012)	0.675*** (0.012)	0.675*** (0.012)
Phrases lac.	Soutien	-0.060*** (0.019)	-0.126 (0.214)	-0.597** (0.263)	-1.444* (0.774)	-0.342* (0.197)	-0.311* (0.169)	-0.338 (0.207)	-0.064 (0.079)	-0.139 (0.105)	-0.170 (0.144)
	Score 2008	0.421*** (0.010)	0.469*** (0.010)	0.417*** (0.011)	0.415*** (0.012)	0.470*** (0.010)	0.418*** (0.011)	0.420*** (0.011)	0.469*** (0.010)	0.419*** (0.011)	0.420*** (0.011)
Mathématiques	Soutien	-0.138*** (0.019)	-0.094 (0.222)	-0.435 (0.272)	-0.924 (0.735)	-0.389* (0.203)	-0.341** (0.172)	-0.392* (0.216)	-0.217 (1.301)	-0.253 (0.494)	-0.379 (0.309)
	Score 2008	0.512*** (0.013)	0.590*** (0.013)	0.512*** (0.014)	0.497*** (0.020)	0.586*** (0.013)	0.514*** (0.014)	0.507*** (0.014)	0.588*** (0.022)	0.516*** (0.016)	0.508*** (0.015)
Compréhension	Soutien	-0.063*** (0.021)	-0.380* (0.229)	-0.769*** (0.298)	-2.115** (0.972)	-0.520** (0.211)	-0.374** (0.180)	-0.337 (0.226)	-0.320 (0.435)	-0.210 (0.249)	-0.102 (0.268)
	Score 2008	0.268*** (0.010)	0.289*** (0.009)	0.263*** (0.010)	0.260*** (0.012)	0.288*** (0.009)	0.264*** (0.009)	0.267*** (0.010)	0.289*** (0.009)	0.265*** (0.009)	0.268*** (0.010)
Raisonnement	Soutien	-0.109*** (0.022)	-0.935*** (0.273)	-1.072*** (0.339)	-1.822* (0.932)	-1.228*** (0.266)	-0.640*** (0.207)	-0.395* (0.237)	-0.239** (0.094)	-0.213** (0.086)	-0.149** (0.067)
	Score 2008	0.391*** (0.012)	0.400*** (0.012)	0.372*** (0.014)	0.355*** (0.024)	0.395*** (0.012)	0.381*** (0.012)	0.385*** (0.012)	0.413*** (0.011)	0.390*** (0.012)	0.390*** (0.012)
contrôles		Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui
revenus		Oui	Non	Non	Oui	Non	Non	Oui	Non	Non	Oui
Obs.		14234	18387	14814	14234	18387	14814	14234	18387	14814	14234

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Erreurs standards robustes entre parenthèses ajustées par établissement (cluster).

Tous les modèles utilisent comme instrument le nombre d'entreprises enregistrées dans la démographie des entreprises (INSEE) en 2011. Ces bases ont été appariées au Panel 2007 sur la base du code commune de l'établissement scolaire où étudie l'élève durant l'année 2010-2011. Les spécifications contiennent donc le nombre d'entreprises de soutien scolaire dans la commune et dans le département exprimé en logarithme. Les spécifications sont identiques aux modèles des tableaux précédents mais contrôlent et reportent le niveau de performances de l'élève en 2008 dans la dimension mesurée.

Bibliographie

- Angrist, J. et J. Pischke (2009). *Mostly Harmless Econometrics : An Empiricist's Companion*. Princeton University Press.
- Banerjee, A. V., S. Cole, E. Duflo, et L. Linden (2007, 08). Remedying Education : Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics* 122(3), 1235–1264.
- Ben Ali, L. et R. Vourc'h (2015). Évolution des acquis cognitifs au collège au regard de l'environnement des élèves. constat et mis en perspective longitudinale. In E. et Formations (Ed.), *Évaluation des acquis : principes, méthodologie, résultats*, pp. 86–87. MENESR-DEPP.
- Berkowitz, D., M. Caner, et Y. Fang (2008). Are "nearly exogenous instruments" reliable? *Economics Letters* 101, 20–23.
- Bouguen, A. (2015, Mars). *Adjusting content to individual student needs : Further evidences from a teacher training program*. Phd thesis, Paris School of Economics.
- Bray, M. (2000). The shadow education system : Private tutoring and its implications for planners. fundamentals of educational planning no. 61. paris : Unesco international institute for educational planning. *International Review of Education* 46(6), 641–644.
- Bray, M. (2011). The challenge of shadow education : Private tutoring and its implications for policy makers in the european union. Technical report, European Commission.
- Bray, M. (2013). Shadow education : comparative perspectives on the expansion and implications of private supplementary tutoring. *Procedia - Social and Behavioral Sciences* (77), 412–420.
- Bray, M. et P. Kwok (2003, Décembre). Demand for private supplementary tutoring : conceptual considerations, and socio-economic patterns in Hong Kong. *Economics of Education Review* 22(6), 611–620.
- Bray, M. et C. Lykins (2012). Shadow education - Private supplementary tutoring and its implications for policy makers in Asia. Technical Report 9, Asian Development Bank.
- Bray, M., S. Zhan, C. Lykins, D. Wang, et O. Kwo (2014). Differentiated demand for private supplementary tutoring : Patterns and implications in Hong Kong secondary education. *Economics of Education Review* 38(C), 24–37.
- Brinbaum, Y. (2002). *Educational Investments by Immigrant Families in France : Expectations and Disillusions*. Theses, Université René Descartes - Paris V.

- Brinbaum, Y. et A. Kieffer (2005). D'une génération à l'autre, les aspirations éducatives des familles immigrées. ambition et persévérance. *Education et Formations* (72), 53–75.
- Brinbaum, Y. et A. Kieffer (2009). Les scolarités des enfants d'immigrés de la sixième au baccalauréat : différenciation et polarisation des parcours. *Population* 64(3). www.cairn.info/revue-population-2009-3-page-561.htm.
- Buchmann, C., D. J. Condrón, et V. J. Roscigno (2010, Décembre). Shadow education, American style : test preparation, the SAT and college enrollment. *Social Forces* 89(2), 435–462.
- Caille, J.-P. (2010). Les cours particuliers en première année de collège : un entrant en sixième sur dix bénéficie de soutien scolaire payant. *Education et formations* (79).
- Caille, J.-P. (2014). Les transformations des trajectoires au collège : des parcours homogènes mais encore très liés au passé scolaire et à l'origine sociale. *Educations & formations*.
- Cameron, A. et P. Trivedi (2005). *Microeconometrics : Methods and Applications*. Cambridge University Press.
- Cattell, R. B. (1971). *Abilities : Their structure, growth, and action*. Houghton Mifflin.
- Centre d'analyse stratégique (2013, Janvier). La note d'analyse - Quelle organisation pour le soutien scolaire? (315).
- Cerulli, G. (2015). *Econometric Evaluation of Socio-Economic Programs : Theory and Applications*. Advanced Studies in Theoretical and Applied Econometrics. Springer Berlin Heidelberg.
- Chartier, P. (2012). Evaluer les capacités de raisonnement. Eurotests Éditions. <http://www.eurotests.fr/media/rcc-5XU882YP.pdf>.
- Cnesco (2015, Janvier). Le redoublement en France et dans le monde. In *Lutter contre la difficulté scolaire : le redoublement et ses alternatives*, Volume 1 à 3 et synthèse. Conseil national d'évaluation du système scolaire (CNESEO).
- Dang, H.-A. (2007, Décembre). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review* 26(6), 683–698.
- Dang, H.-A. et F. H. Rogers (2008, February). How to interpret the growing phenomenon of private tutoring : human capital deepening, inequality increasing, or waste of resources? Policy Research Working Paper Series 4530, The World Bank.
- Davies, S. (2004). School choice by default? Understanding the demand for private tutoring in Canada. *American Journal of Education* 110(3), 233–255.
- De Castro, B. V. et A. B. De Guzman (2010). Push and pull factors affecting Filipino students' shadow education (SE) participation. *KEDI Journal Of Educational* 7(1), 43–66.

- Firestone, W., R. Schorr, et L. Monfils (2004). *The Ambiguity of Teaching to the Test : Standards, Assessment, and Educational Reform*. Taylor & Francis.
- Gauci, D. et S. Wetz (2009). The private lessons phenomenon in a form five girls' mathematics classroom. Bed (hons) dissertation, University of Malta.
- Glasman, D. (2010). Le rôle préventif des dispositifs d'aide aux élèves en dehors de l'école. *Informations Sociales* 161.
- Glasman, D. et L. Besson (2004, Décembre). Le travail des élèves pour l'école en dehors de l'école. *Rapport établi à la demande du Haut conseil de l'évaluation de l'école* (15).
- Haag, L. et M. Van Kessel (2000). Learning out of school : Homework and tutoring - a research study in luxembourg. Technical report, Nürnberg : institut für Psychologie.
- Heim, A., C. Steinmetz, et A. Tricot (2015, Juin). *Faut-il encore redoubler?* Réseau Canopé. ISBN : 978-2-240-03698-8.
- Hof, S. (2014, Janvier). Does private tutoring work ? The effectiveness of private tutoring : a non parametric bounds analysis. Working paper 96.
- INRP (2006, Décembre). Le soutien scolaire entre éducation populaire et industrie de service. *Lettre d'information de la VST* (23).
- Kim, H. et D. Lalancette (2013). Literature review on the value-added measurement in higher education. Technical report, OECD.
- Liu, J. (2012). Does cram schooling matter ? Who goes to cram schools ? *International Journal of Educational Development* 32(1), 46–52.
- Loyalka, P. K. et A. Zakharov (2014). Does shadow education help students prepare for college ? Working papers, National Research University Higher School of Economics (HSE).
- Manski, C. F. (1990, Mai). Nonparametric Bounds on Treatment Effects. *American Economic Review* 80(2), 319–23.
- Manski, C. F. (1997, Novembre). Monotone Treatment Response. *Econometrica* 65(6), 1311–1334.
- Manski, C. F. et J. V. Pepper (2000, Juillet). Monotone Instrumental Variables, with an Application to the Returns to Schooling. *Econometrica* 68(4), 997–1012.
- Melot, L. (2005). Le marché du soutien scolaire.
- Melot, L. (2007). Le marché du soutien scolaire.
- Ministère de l'Éducation nationale (2006). L'accompagnement à la scolarité : pour une politique coordonnée, équitable et adossée aux technologies de l'information et de la communication.

- OCDE (2011). Environnement pédagogique et organisation scolaire. In OCDE (Ed.), *Regards sur l'éducation*, Chapter D.
- OCDE (2015, Juin). Les devoirs entretiennent-ils les inégalités en matière d'éducation? PISA in Focus 46. <http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-46-%28fr%29-rev-june%202015.pdf>.
- Ono, H. (2007). Does examination hell pay off? a cost-benefit analysis of "Ronin" and college education in Japan. *Economics of Education Review* 26(3), 271–284.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5), 688–701.
- Smyth, E. (2008). The more the better?, intensity of involvement in private tuition and examination performance. *Educational research and evaluation* 14(4).
- Smyth, E. (2009). Buying your way to college? private tuition and the transition to higher education in Ireland. *Oxford review of education* 35(1).
- Stock, J. H., J. H. Wright, et M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4), 518–529.
- Tansel, A. et F. Bircan (2006, Juin). Demand for education in Turkey : A tobit analysis of private tutoring expenditures. *Economics of Education Review* 25(3), 303–313.
- Thongphat, N. (2012). A survey of Thai student performance in mathematics and English : evaluating the effect of supplementary tutoring. *Procedia Economics and Finance* 2, 353–362.
- Treiman, D. J. (2009). *Quantitative Data Analysis : Doing Social Research to Test Ideas*.
- Vrignaud, P. (2015). L'évolution de l'équité au collège de la fin du xxème au début du xxième siècle : Etude des trajectoires au collège à partir des données des panels 1995 et 2007 de la depp. In *Inégalités scolaires d'origine sociale et ethnoculturelle : une possible amplification ?* Conseil national d'évaluation du système scolaire (CNESEO).
- Wooldridge, J. (2007). Inverse probability weight estimation for general missing data problem. *Journal of econometrics* 141.
- Zhang, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance? - A case study from Jinan, China. *Economics of Education Review* 32, 1–28.
- Zimmer, R., L. Hamilton, et R. Christina (2010). After school tutoring in the context of no child left behind : Effectiveness of two programs in the pittsburgh public schools. *Economics of Education Review*.

Abstract (Fr)

Le soutien privé a pu être activement contrôlé, voire interdit dans certains pays (e.g. Cambodge, Île Maurice, et même en Corée du Sud où le phénomène est néanmoins très populaire (Bray, 2000)), en réponse aux nombreuses critiques qui l'accusent d'accentuer les inégalités sociales et de détériorer le système d'éducation public sans nécessairement parvenir à améliorer les performances scolaires des élèves qui y ont recourt. Le soutien privé peut prendre des formes très différentes, être dispensé individuellement ou en groupe, par des intervenants de compétences et de qualifications variées, avec des méthodes très hétérogènes, se focalisant sur des sujets ou des axes de travail différents, . . . rendant la définition du soutien privé et la comparaison des élèves extrêmement complexes. Ce rapport est basé sur une analyse économétrique de données de panel des élèves qui entrent au collège en France en 2007. Afin d'examiner un possible effet causal du soutien privé sur les performances des élèves et leur orientation, ce rapport utilise une stratégie par variable instrumentale qui fait appel à l'accessibilité à du soutien privé dans la commune et qui permet d'identifier l'effet réel sur les résultats scolaires des élèves. Les résultats montrent que le soutien privé peut avoir un effet positif significatif sur les résultats en 6^e, particulièrement sur les résultats en mathématiques, en compréhension de texte et en culture générale. Cependant, les performances des élèves en fin de collège, à l'exception d'un effet négatif et significatif sur le raisonnement, ne semblent plus affectées par le soutien privé. Comme nous n'avons pas toutes les informations sur les raisons qui poussent les élèves à prendre des cours privés, et pas d'information sur la prise de soutien privé entre la 6^e et la 3^e, la validité de ces analyses peut apparaître faible. Néanmoins, le soutien privé semble augmenter la probabilité pour les élèves de s'orienter en 2^{nde} générale et technologique au lycée même si cette relation ne peut pas être interprétée comme une relation causale.

Abstract (En)

Private tutoring has been actively controlled and even made illegal in some countries (e.g. Cambodia, Mauritius, even South Korea where it is nevertheless very popular (Bray, 2000)) in response to numerous critiques of its possible role in aggravating educational inequalities and deteriorating the public school system without necessarily improving pupils' performance. Private tutoring can take various forms, be delivered to individuals or in groups, by people with varying skill levels and qualifications, using heterogeneous methods, focused on very different subjects, . . . making it very difficult to define the concept and compare pupils. This report is based on an econometric analysis of panel data on secondary school pupils who entered the 6th grade ("6^e" in France) in 2007. In order to examine the possible causal effect of private tutoring on pupils' educational performance and trajectory, the report uses an instrumental variable approach which capitalizes on information regarding the availability of private tutoring at the level of the "commune" in order to identify the effect of private tutoring on pupil performance. Results show that private tutoring may have a positive, significant effect in the 6th grade, particularly on performance in mathematics, reading comprehension and general culture. However, pupil performance at the end of middle school, with the exception of a negative and significant effect on logical reasoning, seems not to be affected by private tutoring. As we have neither a complete set of variables surrounding pupils' motivation to take private tutoring nor data for the years in between 6th and 9th grade, the validity of these results may be weak. Nevertheless, private tutoring seems to increase pupils' likelihood of entering the general and technological track in high school although these results are probably not perfectly causal.