

Origine migratoire et performance scolaire: décomposition des scores PISA 2000

Muriel Meunier

Ce papier présente l'analyse de la performance scolaire en fonction de l'origine migratoire des élèves à partir de l'échantillon national de PISA 2000 pour la Suisse. Dans une première étape, les fonctions de production éducationnelle sont estimées séparément pour les natifs et les immigrés. Dans une seconde étape, le différentiel de scores entre les natifs et les immigrés est décomposé à l'aide des modèles d'Oaxaca (1973) et Blinder (1973) et de Juhn, Murphy et Pierce (1993). Les résultats indiquent que 60% des différences de scores entre natifs et immigrés peuvent être expliquées par des différences de caractéristiques moyennes entre les deux groupes. Ils montrent également que les immigrés bénéficient davantage d'être dans un environnement où les enseignants sont qualifiés et que l'origine migratoire de l'élève est moins pénalisante lorsque ceux-ci sont des élèves performants.

JEL-Classification: I20, J16, C2

Keywords: fonction de production éducationnelle, décomposition, origine migratoire, PISA.

1. Introduction

La Suisse est l'un des pays d'Europe où la proportion d'étrangers est la plus élevée. En 2000, elle représentait environ 20% de la population dont un cinquième avait moins de 15 ans (OFS, 2005). En vingt ans, la proportion de jeunes issus de familles d'immigrés n'a cessé d'augmenter passant de 14% en 1980 à 21% en 2000 (OFS, 2002). Il n'est donc pas surprenant que la part des classes très hétérogènes, c'est-à-dire avec plus d'un tiers d'élèves d'origine ou de langue maternelle étrangère, ait fortement augmentée. Pourtant, il semblerait que le système éducatif suisse peine à faire face au défi représenté par les élèves immigrés et que les enfants étrangers n'aient pas les mêmes opportunités que les autres enfants dans le système éducatif. En dernière année de scolarité obligatoire, les données PISA 2000 pour la Suisse indiquent que "les jeunes de familles d'immigrés obtiennent en moyenne plus de 100 points de moins sur l'échelle de lecture de l'OCDE que les jeunes de familles autochtones"¹ (OFS et CDIP, 2002).

¹ Dans le rapport national de l'enquête PISA 2000 (OFS et CDIP, 2002), les familles autochtones correspondent aux familles dont les deux parents sont nés en Suisse, les familles biculturelles à celles dont le père ou la mère est né(e) à l'étranger et les familles d'immigrés à celles dont les parents sont nés à l'étranger. Précisons que les jeunes dont les parents ont émigrés par exemple d'Italie pour venir s'établir au Tessin ou de France pour se fixer en Suisse romande n'ont pas été inclus dans la catégorie des jeunes de familles d'immigrés.

L'objectif de cette étude sera de mettre en relation l'origine migratoire des élèves de l'échantillon national de PISA 2000 avec leurs performances scolaires en lecture. Quelle partie du différentiel de scores en lecture peut être attribuée à des différences de caractéristiques moyennes entre deux groupes migratoires considérés? Pour répondre à cette question, nous utilisons les techniques de décomposition d'OAXACA (1973) et BLINDER (1973) et de JUHN, MURPHY et PIERCE (1993). Il s'agira d'une part, de mettre en évidence les caractéristiques qui ont un impact significatif sur le différentiel de scores entre les natifs et les immigrés et d'autre part, de voir comment évolue le différentiel de scores le long de la distribution.

La première contribution à la littérature existante concerne l'impact de la langue parlée à la maison comme variable explicative des différences de scores entre natifs et immigrés. Les résultats indiquent que le fait de ne pas parler la langue du test à la maison n'a pas d'impact négatif plus important pour les natifs que pour les immigrés, qu'ils soient de deuxième ou première génération. La seconde contribution est relative à l'environnement éducationnel. Les immigrés ont tendance à fréquenter des écoles en moyenne plus grandes et dans lesquelles le niveau de qualification des enseignants est en moyenne inférieur. Les résultats indiquent que les immigrés bénéficient davantage d'avoir un enseignant qualifié que les natifs. Finalement, les résultats indiquent que l'origine migratoire de l'élève est moins pénalisante lorsque ceux-ci sont de bons élèves.

Le reste de l'article est organisé de la manière suivante. La prochaine section présente la méthodologie. La section 3 présente la base de données et la section 4, les résultats de l'analyse empirique. La dernière section conclut avec les suites possibles de recherche.

2. Méthodologie

Cette section présente la stratégie empirique utilisée pour estimer la performance scolaire et décomposer le différentiel de scores observé entre deux groupes distincts. Dans une première étape, les fonctions de production éducationnelle permettront d'identifier les caractéristiques ayant un impact sur les scores individuels. Dans une seconde étape, les méthodologies d'OAXACA (1973) et BLINDER (1973) et de JUHN, MURPHY et PIERCE (1993) permettront de décomposer le différentiel de scores.

Bien que les méthodes de décomposition ont initialement été utilisées dans le domaine de l'économie du travail avec des applications empiriques sur des données salariales, la littérature actuelle les utilise également dans des domaines de recherche tels que l'économie de l'éducation (COOK et EVANS, 2000; AMMERMÜLLER, 2008).

2.1 Fonctions de production éducationnelle

L'estimation de la fonction de production éducationnelle permet de mesurer l'impact des caractéristiques individuelles, familiales et institutionnelles sur

les acquis scolaires au niveau individuel. Dans sa forme générale, elle peut être modélisée de la manière suivante:

$$A_i = \beta Z_i + u_i \quad (1)$$

où A_i est une mesure de l'output scolaire du $i^{\text{ème}}$ élève ($i = 1, \dots, n$), β est le vecteur des coefficients, Z_i contient les variables relatives aux caractéristiques de l'élève, sa famille et son environnement scolaire et u_i est le terme d'erreur.

Si l'on souhaite tester l'hypothèse selon laquelle deux groupes démographiques distincts sont soumis à une évaluation différente de la part du système éducatif, il est alors nécessaire de réaliser des estimations séparées de l'équation (1) pour chacun des sous-groupes considérés:

$$A_i^n = \beta^n Z_i^n + u_i^n \quad (2)$$

$$A_i^m = \beta^m Z_i^m + u_i^m$$

Dans l'équation (2), les exposants permettent de faire la distinction entre les deux groupes d'individus. Lors de l'analyse empirique, les groupes considérés seront définis en fonction de l'origine migratoire des élèves. Les exposants n et m correspondant respectivement aux groupes des natifs et des immigrés.

2.2 Décomposition Oaxaca (1973) et Blinder (1973)

A partir des propriétés des estimations des moindres carrés ordinaires (MCO) et en reprenant les fonctions de production éducationnelle telles que définies ci-dessus, il est possible d'estimer la différence totale de scores moyens entre les deux groupes considérés telle que:

$$\overline{A^n} - \overline{A^m} = \hat{\beta}^n \overline{Z^n} - \hat{\beta}^m \overline{Z^m} \quad (3)$$

où $\overline{A^n}$ et $\overline{A^m}$ sont les scores moyens (moyenne géométrique) pour les natifs et les immigrés, $\overline{Z^n}$ et $\overline{Z^m}$ sont des vecteurs des valeurs moyennes des régresseurs pour les natifs et les immigrés respectivement et $\hat{\beta}^n$ et $\hat{\beta}^m$ sont les vecteurs correspondants des coefficients estimés.

La méthode développée par OAXACA (1973) et BLINDER (1973) permet de décomposer le différentiel total de scores en deux éléments: la composante des *caractéristiques* et la composante des *coefficients*². Le différentiel de scores entre natifs et immigrés peut alors être écrit de la manière suivante:

$$\overline{A^n} - \overline{A^m} = \underbrace{(\hat{\beta}^n - \hat{\beta}^m) \overline{Z^m}}_{C_m} + \hat{\beta}^n \underbrace{(\overline{Z^n} - \overline{Z^m})}_{E_m} \quad (4)$$

² Pour une décomposition alternative en trois parties, voir WINSBOROUGH et DICKINSON (1971).

Dans l'équation (4), le terme E_m représente la partie de la différence totale qui peut être expliquée par des niveaux différents de caractéristiques moyennes. Par exemple, étant donné que le niveau de capital humain de la mère a un impact important sur la performance scolaire des élèves, il peut s'agir d'un niveau de formation différent entre les mères du groupe des natifs et celles du groupe des immigrés. On peut aussi dire que le terme E_m représente les effets estimés des différences de caractéristiques individuelles c'est-à-dire la valeur de l'avantage en dotations possédés par le groupe des immigrés tel qu'évalué par l'équation des scores du groupe des natifs. Il s'agit d'une différence observable qui permet d'expliquer, sans pour autant justifier, que le score des immigrés est inférieur à celui des natifs.

La part du différentiel attribuable à la partie inexpliquée du différentiel de scores correspond au terme C_m . Ce terme illustre la partie du différentiel qui provient du fait que le système éducatif apprécie de manière différenciée des caractéristiques observées identiques que ce soit auprès du groupe des natifs ou de celui des immigrés. Ces effets estimés correspondent à la différence entre la manière dont l'équation des immigrés estimerait les caractéristiques du groupe des natifs et celle dont l'équation des scores des natifs l'estime en fait. Dans le cas présent, on fait l'hypothèse que la structure courante des scores des natifs s'appliquera à la fois aux natifs et aux immigrés sur un marché de l'éducation non discriminant ($W = 1$)³.

Si la structure des deux groupes était la même pour les variables considérées (genre, âge, etc.), tout écart de score ne pourrait provenir que d'un écart de *rendement* des caractéristiques⁴. A l'inverse, si les rendements étaient similaires, le différentiel de scores résulterait entièrement d'effets structurels, eux-mêmes pouvant éventuellement être la conséquence de diverses formes de ségrégation (accès à l'éducation, accès à certaines filières, etc.).

2.3 Décomposition Juhn, Murphy et Pierce (1993)

Un des avantages de la décomposition de JUHN, MURPHY et PIERCE (1993) est qu'elle ne considère pas seulement la moyenne pour la décomposition mais l'ensemble de la distribution⁵. Reprenons les fonctions de production éducationnelle des deux groupes migratoires d'élèves considérés:

$$A_i^n = \beta^n Z_i^n + u_i^n \quad (5)$$

$$A_i^m = \beta^m Z_i^m + u_i^m$$

³ Pour une décomposition alternative du différentiel de score, voir annexe A7.

⁴ Le jargon utilisé est emprunté au domaine de l'économie du travail.

⁵ Cette décomposition a été utilisée pour étudier des changements dans le temps (données longitudinales ou en coupe transversale répétée) mais a également été appliquée à des données en coupe transversale (BLAU et KAHN, 1992).

où A_i^n et A_i^m représentent le score en lecture pour les groupes n et m . Nous considérerons ici que u_i^n , le résidu du groupe n , regroupe deux composants: un centile individuel dans la distribution du résidu (θ_i^n) et la fonction de distribution des résidus de la fonction de production éducationnelle, $F^n(\cdot)$. La fonction de distribution cumulative des résidus peut alors être exprimée de la manière suivante:

$$u_i^n = F^{n(-1)}(\theta_i^n | Z_i^n) \quad (6)$$

où $F^{n(-1)}(\cdot | Z_i^n)$ est la fonction de distribution cumulative inverse du résidu.

Dans ce cadre conceptuel, les changements dans l'inégalité des deux groupes peut provenir de trois sources: des changements dans la distribution des caractéristiques individuelles (i.e. changements dans la distribution des Z), des changements dans les *prix des caractéristiques observables* (i.e. changements dans les β) ou des changements dans la distribution des résidus.

En utilisant les coefficients estimés à partir des équations (5), il est possible de définir la distribution actuelle des scores en lecture ainsi que deux distributions hypothétiques pour chacun des deux groupes considérés. Les distributions actuelles pour les natifs et les immigrants correspondent aux résultats avec des quantités qui varient, des prix qui varient et une distribution des résidus qui varie et peut être définie de la manière suivante:

$$A_i^n (\text{Actuel}) = \hat{\beta}^n Z_i^n + F^{n(-1)}(\theta_i^n | Z_i^n) \quad (7a)$$

$$A_i^m (\text{Actuel}) = \hat{\beta}^m Z_i^m + F^{m(-1)}(\theta_i^m | Z_i^m) \quad (7b)$$

Les résultats hypothétiques avec des quantités qui varient entre les groupes mais des prix fixés (coefficients) et une distribution des résidus fixées peuvent être définis tels que⁶:

$$A_i^n (\text{Hypothétique 1}) = \hat{\beta}^n Z_i^n + F^{n(-1)}(\theta_i^n | Z_i^n) \quad (8a)$$

$$A_i^m (\text{Hypothétique 1}) = \hat{\beta}^m Z_i^m + F^{m(-1)}(\theta_i^m | Z_i^m) \quad (8b)$$

Les résultats hypothétiques avec des quantités qui varient et des prix qui varient mais une distribution de résidus fixe peuvent être définis tels que:

$$A_i^n (\text{Hypothétique 2}) = \hat{\beta}^n Z_i^n + F^{n(-1)}(\theta_i^n | z_i^n) \quad (9a)$$

$$A_i^m (\text{Hypothétique 2}) = \hat{\beta}^m Z_i^m + F^{m(-1)}(\theta_i^m | z_i^m) \quad (9b)$$

⁶ La référence considérée est *natif*. Cela implique que les coefficients de l'estimation des natifs sont utilisés comme prix de référence et que les résidus de l'estimation des natifs sont utilisés pour déterminer la distribution des résidus de référence.

L'équation (8b) indique ce que serait la distribution des scores des immigrés si ces derniers utilisaient le processus de production éducationnel des natifs ainsi que les résidus correspondant à la distribution des résidus des natifs. L'équation (9b) indique ce que serait la distribution des scores des immigrés si la distribution des résidus était la même que celle des natifs.

Dénotons par une lettre majuscule, une statistique résumant la distribution d'une variable. Par exemple, A serait le moyenne ou la dispersion interquartile de la distribution de A_i . Le différentiel $A^n - A^m$ peut alors être décomposé tel que:

$$\begin{aligned} A^n - A^m &= [A^n(H1) - A^m(H1)] \\ &+ [(A^n(H2) - A^m(H2)) - (A^n(H1) - A^m(H1))] \\ &+ [(A^n(A) - A^m(A)) - (A^n(H2) - A^m(H2))] \end{aligned} \quad (10)$$

Le premier effet de la décomposition du différentiel de scores entre les natifs et les immigrés est l'*effet des caractéristiques* (Q), le second effet correspond à l'*effet des rendements* (P) et finalement, le troisième effet correspond à l'*effet des résidus* (U). La différence totale (T) correspond à la somme des ces trois effets telle que:

$$T = Q + P + U \quad (11)$$

3. Les données

Les données utilisées pour l'analyse empirique proviennent de l'échantillon national de l'enquête PISA (Programme International pour le Suivi des Acquis des élèves) réalisée en 2000 par l'OCDE (OCDE, 2002)⁷. L'objectif de cette enquête était de tester les compétences des élèves dans le but de comparer les acquis des jeunes (avec une priorité pour la lecture). La population de référence a été définie en fonction de l'année scolaire. Il s'agit de la 9^e année du système scolaire suisse qui correspond à la dernière année de la scolarité obligatoire⁸.

Lors de l'enquête PISA 2000, la question relative à la nationalité de l'élève n'a pas été explicitement posée dans le questionnaire. En revanche, deux questions permettent d'appréhender l'origine migratoire des élèves ayant participé aux tests (OCDE, 2002). La première concerne le lieu de naissance de l'élève et la seconde est relative au lieu de naissance des parents. La section suivante présente la classification des élèves selon leur origine migratoire. Les statistiques descriptives des variables qui seront utilisées lors de l'analyse empirique sont ensuite présentées.

⁷ Pour plus de détails sur l'échantillonnage et le déroulement de l'enquête, voir Essai 1.

⁸ La scolarité obligatoire dure 9 ans et comprend les degrés primaire et secondaire I. Pour plus de détails sur le système éducatif suisse, voir le site internet de la CDIP (http://www.ides.ch/umfrage2003/mainUmfrage_F.html).

3.1 L'origine migratoire des élèves

Dans le jargon juridique, la règle de droit intitulée droit du sol (ou *jus soli*) permet de définir le lien territorial qui existe entre un individu et son lieu de naissance. Elle implique que toute personne physique née sur le territoire national, indépendamment de la nationalité de ses parents, peut obtenir la nationalité. Alors que cette information est primordiale pour certains pays comme les Etats-Unis ou la Grande-Bretagne pour accorder la nationalité, la naissance sur le sol helvétique ne confère aucun droit particulier aux individus car le droit du sol n'y est pas appliqué.

Il est donc nécessaire de recourir au lieu de naissance de la génération précédente à savoir les parents. Il s'agit dans ce cas de la règle de droit intitulée droit du sang (ou *jus sanguinis*) qui renvoie au lien de filiation qui existe entre un individu et ses parents. Cette règle signifie que la nationalité est accordée aux individus nés de parents possédant eux-mêmes la nationalité. C'est le cas de pays comme la Suisse ou le Japon où l'accès à la nationalité y est ainsi considérablement limité et la présence étrangère particulièrement visible. Etant donné que l'enquête PISA ne permet pas de connaître la nationalité des parents des élèves, le lieu de naissance sera utilisé à la place.

Les élèves de l'échantillon national de PISA 2000 ont été scindés en trois groupes dépendamment de leur origine migratoire et en se basant sur la classification utilisée par l'OCDE dans ses rapports actuels⁹. Il s'agit des catégories natif, immigré de seconde génération et immigré de première génération. Les élèves ont été répartis de la manière suivante:

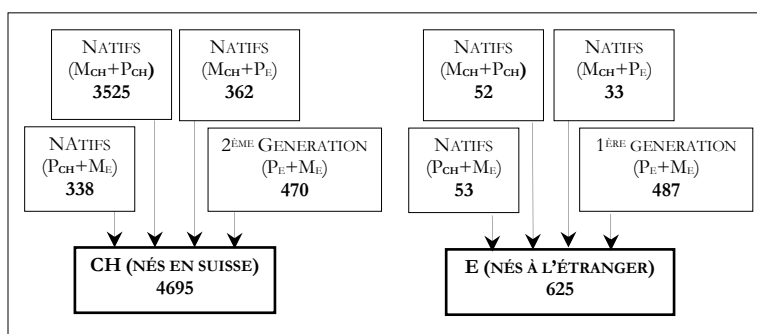
- *natif*: il s'agit des élèves nés dans le pays d'évaluation ou à l'étranger et dont au moins un parent est né dans le pays d'évaluation;
- *deuxième génération d'immigré*: il s'agit des élèves nés dans le pays d'évaluation mais dont les parents sont nés à l'étranger. Ces élèves ont suivi l'intégralité de leur scolarité dans le pays d'évaluation;
- *première génération d'immigré*: il s'agit des élèves nés en dehors du pays d'évaluation et dont les parents sont également nés à l'étranger.

Le graphique 1 illustre la répartition des élèves selon leur lieu de naissance et celui de leurs parents. Les natifs représentent 82% de l'échantillon utilisé pour les estimations (4363 élèves), les élèves de deuxième génération d'immigré 8.8% (470 élèves) et ceux de première génération d'immigré 9.2% (487 élèves). La majorité des natifs (81%) comprend les élèves nés en Suisse et dont les deux parents sont également nés en Suisse (3525 élèves). Les

⁹ L'OCDE (2005) utilise la classification suivante: nés en Suisse ou ayant au moins un parent d'origine suisse, première génération (élèves nés en Suisse dont les deux parents sont nés à l'étranger), nés à l'étranger (élèves et parents nés à l'étranger). Etant donné que l'OFS définit la première génération d'étrangers comme les personnes nées à l'étranger, qui ont immigré en Suisse et qui ne possèdent pas la nationalité suisse et la deuxième génération d'étrangers comme les personnes nées en Suisse, qui ne possèdent pas la nationalité suisse et dont les deux parents sont nés à l'étranger, nous avons procédé à un ajustement de la classification de l'OCDE.

élèves qui n'ont pas indiqué leur lieu de naissance ou celui de leurs deux parents dans les questionnaires ont été supprimés de l'échantillon.

Graphique 1: Répartition des 5320 élèves de PISA 2000 selon leur lieu de naissance et le lieu de naissance de leurs parents (M=mère et P=père)

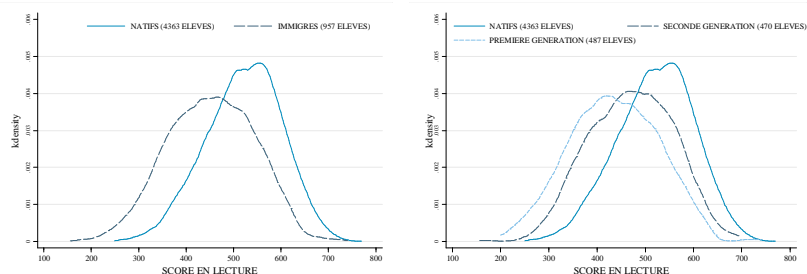


3.2 Les variables

La variable expliquée est le score en lecture obtenu par les élèves lors des tests (variable LECTURE). Le score moyen de l'échantillon total (5320 élèves) est de 509.0979 (tableaux A1 et A2 en annexe). Le différentiel de scores entre les natifs et les immigrés est de 69.0037 et les scores moyens sont significativement différents entre les natifs et les immigrés (respectivement 521.5108 et 452.5071). Exprimé en pour cent du score des immigrés, le différentiel est de 13.2% (gains en plus des natifs).

Lorsque l'on fait la distinction entre les natifs et les immigrés de deuxième et de première génération, le différentiel de scores est moins important dans le premier cas (50.6044) que dans le second (86.7607). Finalement, le différentiel le plus faible est celui entre les immigrés de deuxième et de première génération (36.1563). Les graphiques suivants présentent la densité des scores en lecture pour les natifs et les immigrés, avec une distinction entre les immigrés de deuxième et de première génération (graphique de droite).

Graphique 2: Densité des scores en lecture selon l'origine migratoire des élèves



Deux questions supplémentaires permettent d'évaluer le degré d'assimilation ou d'intégration des élèves dans le pays d'accueil. La durée de résidence de l'élève sur le territoire du test (variable DUREE exprimée en mois) ainsi que la langue parlée à la maison en comparaison avec la langue du test (variable LANGUE qui prend la valeur 1 si la langue parlée à la maison correspond à la langue du test). Il est intéressant de constater que la durée de résidence sur le territoire du test est très hétérogène dépendamment du groupe considéré¹⁰ : de 15 années en moyenne pour les natifs et les élèves de seconde génération à 9,5 années pour les élèves de première génération. Cela implique que ces derniers sont arrivés en Suisse vers l'âge de 6 ans (soit l'âge théorique de début du degré primaire) et qu'ils n'ont donc pas fait leur degré préscolaire dans le pays d'accueil (de 3 ans à 6 ans dépendamment des cantons). Concernant la langue du test, de grandes différences apparaissent : tandis que 94% des natifs parlent la langue du test à la maison, ils sont seulement 45% de la deuxième génération et plus que 19% de la première génération.

La littérature a mis en évidence que la productivité de l'élève dépend de ses caractéristiques individuelles telles que le genre et l'âge. Les proportions de filles et de garçons sont relativement constantes selon le groupe migratoire considéré (variable GENRE). Les filles représentent 50% du groupe des natifs et des élèves de deuxième génération et 46% des élèves de première génération. Par ailleurs, les élèves de première génération sont en moyenne 4 mois plus âgés que les élèves natifs ou de deuxième génération (variable AGE).

L'origine sociale est également un fort déterminant de la réussite scolaire des élèves. L'état civil, le nombre de frères et sœurs ainsi que la situation professionnelle des parents permettent d'évaluer la situation économique du ménage dans lequel les enfants vivent et donc indirectement des potentiels investissements effectués dans la cellule familiale qui peuvent augmenter la productivité des élèves. Des différences significatives apparaissent entre les différents groupes migratoires. On observe que 80% des élèves natifs ou de deuxième génération vivent avec leurs deux parents tandis que 86% des immigrés de première génération sont dans ce cas (variable NUCLEAIRE). Ces derniers sont donc moins nombreux à résider dans une structure familiale décomposée suite à une séparation ou un décès. C'est peut être ce qui explique le fait que les élèves immigrés de première génération sont également ceux qui, en moyenne, possèdent le plus grand nombre de frères et sœurs (variable NSIB): 1.88 contre 1.65 pour les natifs et 1.49 pour les immigrés de seconde génération.

La performance de l'élève dépend également de ses aptitudes innées, indépendantes de son environnement scolaire. Etant donné que ces aptitudes sont inobservables, elles seront mesurées par des variables proxy permettant d'appréhender leur capital génétique. Il s'agit par exemple du niveau de capital humain de chacun de ses parents, caractéristiques pour lesquelles on

¹⁰ La majorité des résidents étrangers vivent depuis longtemps en Suisse. Parmi ceux qui sont nés à l'étranger, 36% vivent depuis 15 ans ou plus de manière ininterrompue dans notre pays. Deux tiers des étrangers possèdent un permis d'établissement illimité. La durée de résidence et le statut des étrangers varient considérablement selon les nationalités (OFS, 2005).

observe également des différences significatives (variables EDUCM et EDUCF). Dans 63% des cas, les mères des natifs ont terminé des études post-obligatoires c'est-à-dire au-delà des 9 ans théoriques de scolarité obligatoire. Seules 36% des mères des élèves de deuxième génération et 27% de celles des élèves de première génération sont dans ce cas de figure. Concernant les pères, les différences sont moins importantes mais on observe tout de même que 67% des pères des natifs ont terminé des études post-obligatoires contre 39% pour les immigrés de deuxième génération et 41% pour ceux de première génération. Il en découle des différences importantes en terme de ressources économiques telles que mesurées par l'indice socio-économique du statut professionnel (variable ISEI¹¹). L'indice ISEI moyen est de 47 pour les natifs contre seulement 39 pour les immigrés de deuxième génération et 36 pour ceux de première génération.

Finalement, les caractéristiques de l'école permettent d'appréhender l'environnement scolaire des élèves qui peut influencer la productivité des élèves. Cet environnement est variable en fonction de l'origine migratoire des élèves. Au niveau de l'échantillon total, 69% des élèves fréquentent une école qui se situe dans les cantons germanophones, 21% dans les cantons francophones et 10% dans les cantons italophones (variables GERMANOPHONE, FRANCOPHONE et ITALOPHONE). Les immigrés de seconde et première génération sont en revanche bien plus nombreux dans les cantons italophones (respectivement 22% et 17%). Cette tendance se fait au détriment des cantons germanophones (respectivement 59% et 63%), les pourcentages dans les cantons francophones restant stables quelle que soit l'origine migratoire des élèves.

Il apparaît également que les immigrés ont tendance à se concentrer dans les grandes villes (variable CITY)¹². Tandis que les natifs sont seulement 6% à fréquenter une école qui se trouve dans une commune dont le nombre d'habitants est compris entre 100000 et un million environ, ils sont 17% pour les élèves de deuxième génération et 15% pour ceux de première génération. Les immigrés ont également tendance à fréquenter des écoles de taille plus grande en moyenne (variable SCHLSIZE). En effet, les élèves immigrés fréquentent des écoles dont le nombre total d'élèves inscrits dans l'établissement est d'environ 445 élèves en moyenne contre 402 pour les élèves natifs. En revanche il n'y a pas de différence significative entre les natifs et les immigrés en ce qui concerne la taille moyenne des classes mesurée par le ratio entre la taille de l'école et le nombre de professeurs (variable STRATIO) ainsi que le nombre d'heures d'encadrement moyen par année (variable TOTHR). Cela est peu surprenant dans la mesure où l'effectif des classes et l'encadrement sont définis légalement au niveau cantonal. Finalement, on observe que les immigrés fréquentent des écoles dans

¹¹ L'ISEI est dérivé des réponses des élèves concernant la profession des parents. Il s'agit d'une variable continue dont les valeurs sont comprises entre 16 et 90 dans le cas de PISA 2000. Pour plus d'information sur la méthodologie de cet indice, voir GANZEBOOM, DE GRAAF et TREIMAN (1992).

¹² La proportion d'étrangers varie beaucoup d'une région à l'autre notamment en fonction du degré d'urbanisation. Elle oscille au niveau cantonal entre 8% (Uri) et 38% (Genève). Elle est particulièrement élevée dans les grandes villes (OFS, 2005).

lesquelles la proportion de professeurs possédant un niveau ISCED 5 en pédagogie est de 49% contre 59% pour les natifs (variable PROPQUAL).

4. Analyse empirique

La stratégie empirique utilisée consiste en une analyse en deux temps. Dans un premier temps, les fonctions de production éducationnelle ont été estimées par moindres carrés ordinaires (tableau 1). Dans une seconde étape, les différentiels de scores ont été décomposés à l'aide des méthodes d'OAXACA et BLINDER (tableau 2) et de JUHN, MURPHY et PIERCE (tableau 3).

Cette analyse s'inscrit dans la lignée des recherches qui s'intéressent au lien entre l'origine migratoire des élèves et leur performance scolaire¹³. On peut citer l'étude d'AMMERMÜLLER (2008) qui utilise l'échantillon international de PISA 2000 pour expliquer les différences significatives de performances qui ont été observées entre des élèves Finlandais et Allemand¹⁴. L'auteur utilise également les techniques de décomposition d'OAXACA et BLINDER et de JUHN, MURPHY et PIERCE pour expliquer le différentiel de scores entre les élèves des deux pays¹⁵.

ENTORF et MINOIU (2005) utilisent également l'échantillon international de PISA 2000 afin de présenter une comparaison internationale des performances scolaires. Bien que les auteurs n'utilisent pas les techniques de décomposition, ils étudient l'importance du statut social et des antécédents migratoires sur la performance des élèves. Leurs résultats indiquent que les scores en lecture des élèves immigrés s'améliorent significativement quand la langue parlée à la maison est la langue nationale (par opposition aux langues étrangères). Ils trouvent également que l'impact du statut socio-économique des parents sur la performance scolaire de leurs enfants diffère fortement selon les pays.

COOK et EVANS (2000) quant à eux utilisent la méthode de décomposition d'OAXACA mais emploient les données NAEP (National Assessment of Educational Progress) en coupes transversales répétées afin de décomposer les différentiels de scores. Les auteurs essaient de déterminer ce qui pourrait causer la convergence des performances académiques entre les élèves noirs et blancs pendant la période 1970-1988. Leurs résultats indiquent qu'environ 25% de la convergence des scores est attribuable à des changements dans les caractéristiques de la famille et des écoles. Les auteurs trouvent également

¹³ Plusieurs études s'intéressent spécifiquement à l'éducation des immigrés aux Etats-Unis que ce soit au niveau de l'éducation (SMITH, 2006; COLLINS et MARGO, 2006), de la convergence des compétences entre noirs et blancs (NEAL, 2006) ou encore de la déségrégation pour les noirs (RIVKIN et WELCH, 2006).

¹⁴ On peut également citer le working paper de SCHNEEWEIS (2006) qui utilise les techniques de décompositions d'OAXACA et BLINDER. L'auteur utilise les données de cinq enquêtes internationales (dont PISA 2000 et 2003) afin d'étudier l'intégration éducationnelle des enfants avec des antécédents migratoires.

¹⁵ L'auteur trouve que les élèves allemands possèdent des caractéristiques en moyenne plus favorables (exceptés pour les déciles les plus faibles) mais avec des rendements plus faibles en terme de performance scolaire que les élèves finlandais.

que 75% de la convergence est attribuable à des changements à l'intérieur des écoles¹⁶.

Les résultats de l'estimation pour l'échantillon total sont reproduits à la colonne 1¹⁷. Les résultats indiquent que les variables utilisées sont toutes significatives à 1% (excepté ITALOPHONE et CITY, respectivement non significative et significative à 5%). Le niveau de qualification des enseignants de l'école est une variable qui a un impact conséquent sur le score en lecture. Lorsque la proportion de professeurs avec un niveau ISCED 5 en pédagogie augmente d'une unité, le score en lecture augmente de +14% (71 points).

La langue parlée à la maison, le genre et l'âge sont également des variables qui ont un impact significatif sur le score: +5.3% si la langue parlée à la maison correspond à celle du test (soit une augmentation de 27 points), +3.7% si l'élève est une fille (soit une augmentation de 19 points) et -3.9% si l'élève est âgé d'une année supplémentaire. La variable relative à l'origine migratoire des élèves indique que les natifs obtiendront, toutes choses égales par ailleurs, des scores en lecture supérieurs de 3.6% par rapport aux élèves immigrés. Par rapport aux élèves natifs, les immigrés de deuxième génération obtiennent des scores inférieurs de 3.1% et ceux de première génération des scores inférieurs de 5% (tableau A4).

Concernant les caractéristiques de la famille, le niveau de capital humain de la mère est une variable qui a un impact fort sur les scores en lecture: +4.3% si celle-ci a terminé des études post-obligatoires. Le niveau de capital humain du père (+2.7% si le père a terminé des études post-obligatoires), la structure familiale (+2.4% si l'élève vit dans une famille nucléaire), la taille de la fratrie (-1.3% par frère ou sœur supplémentaire) et le niveau de revenu de la famille (+1.7% si l'indice ISEI augmente de 10 points) sont des caractéristiques qui ont un moindre impact sur les performances en lecture.

Au niveau de l'environnement scolaire, l'élève qui fréquente une école dans un canton francophone obtiendra un score inférieur de 6.2% par rapport au fait de fréquenter une école située dans un canton germanophone. Si l'élève fréquente une école qui se situe dans une commune dont le nombre d'habitants est supérieur à 100'000, il obtiendra un score inférieur de 1.7%. La taille de l'école, la taille de la classe et le nombre d'heures d'encadrement par année ont en revanche peu d'impact sur les scores.

NATIFS VS. IMMIGRES

Les colonnes 2 et 3 présentent les résultats des estimations pour les groupes des natifs et des immigrés. Afin d'être en mesure de comparer les coefficients obtenus dans chacune des estimations, il est nécessaire de tester s'ils sont significativement différents (tableaux A5 et A6).

¹⁶ Précisons que les auteurs se penchent sur le fait que deux variables omises (le revenu de la famille et la structure de la famille) pourraient biaiser leurs résultats.

¹⁷ La fonction de production éducationnelle a également été estimée avec la méthodologie BOX-COX (1964). Des estimations par moindres carrés ordinaires avec des effets fixes par école et des effets aléatoires ont également été effectuées afin de vérifier la robustesse des résultats. Pour plus de détails voir MEUNIER (2006).

Tableau 1: Fonctions de production éducationnelle selon l'origine migratoire

Variables	(1) TOTAL		(2) NATIF		(3) IMMIGRE		(4) 2 ^{EME} GENERATION		(5) 1 ^{ERE} GENERATION	
	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
DUREE (MOIS)	0.1549	(3.62)***	-0.0248	(0.36)	0.2764	(5.07)***	0.1897	(0.84)	0.3183	(3.95)***
LANGUE	26.8771	(7.32)***	24.4204	(4.92)***	17.2838	(2.84)***	15.9832	(1.86)*	24.4656	(2.54)**
GENRE	18.6750	(9.72)***	18.7567	(9.01)***	17.0115	(3.60)***	19.9594	(2.89)***	17.1052	(2.51)**
AGE (MOIS)	-1.6471	(11.67)***	-1.6614	(9.97)***	-0.9379	(3.18)***	-1.0415	(2.25)**	-0.6333	(1.46)
NUCLEAIRE	12.4317	(4.99)***	14.6322	(5.44)***	5.5437	(0.87)	9.6506	(1.12)	2.2617	(0.23)
NSIB	-6.4326	(6.38)***	-4.4903	(4.16)***	-12.4462	(4.62)***	-7.6591	(1.86)*	-15.9028	(5.33)***
EDUCATION M	22.1214	(8.99)***	20.9349	(7.97)***	24.8795	(3.77)***	24.4436	(2.88)***	26.0896	(2.53)**
EDUCATION F	13.8877	(5.42)***	17.0862	(6.09)***	6.5032	(1.06)	12.7057	(1.48)	1.0199	(0.11)
ISEI	0.8435	(13.82)***	0.7874	(12.20)***	1.1244	(6.15)***	1.1282	(4.44)***	1.0810	(3.96)***
GERMANOPHONE	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.
FRANCOPHONE	-31.5402	(12.13)***	-37.4475	(13.56)***	-3.3633	(0.45)	-5.7986	(0.54)	-2.8172	(0.27)
ITALOPHONE	3.7868	(0.93)	-10.7987	(2.31)**	43.8911	(5.37)***	31.4450	(2.62)***	53.6955	(4.63)***
CITY	-8.8451	(2.26)**	-4.6099	(1.02)	-16.9874	(2.15)**	-18.6769	(1.63)	-15.8484	(1.45)
SCHLSIZE	0.0436	(10.30)***	0.0482	(10.20)***	0.0254	(2.83)***	0.0246	(1.70)*	0.0271	(2.48)**
STRATIO	-1.4978	(5.28)***	-1.6415	(5.55)***	-0.8550	(0.90)	-1.9302	(1.30)	-0.3809	(0.33)
TOTHR5	0.0381	(3.15)***	0.0322	(2.47)**	0.0719	(2.27)**	0.0577	(1.15)	0.0904	(2.29)**
PROPQUAL	71.3568	(20.31)***	64.7458	(16.93)***	90.6503	(10.38)***	81.1131	(6.38)***	97.0341	(7.81)***
CONSTANTE	614.6103	(20.58)***	679.3333	(20.32)***	420.024	(6.26)***	475.411	(5.18)***	338.9377	(3.43)***
NATIFS	18.5009	(4.97)***								
N	5320 (100%)		4363 (82%)		957 (18%)		470 (8.8%)		487 (9.2%)	
R ²	0.3486		0.2710		0.3729		0.3176		0.3857	

Source: PISA 2000. Notes: la variable dépendante est le score en lecture. Statistiques t robustes entre parenthèses. * significatif à 10%, ** significatif à 5%, *** significatif à 1%. Les données sont non pondérées.

Les résultats indiquent que le coefficient associé à la langue parlée à la maison est positif et significatif dans les deux estimations. Il est cependant intéressant de constater que les coefficients ne sont pas significativement différents entre les deux groupes considérés. Cela signifie que, toutes choses égales par ailleurs, un élève qui ne parle pas la langue du test à la maison obtiendra un score moindre en lecture quelle que soit son origine migratoire. Ce premier résultat est intéressant dans la mesure où la non maîtrise de la langue du pays d'accueil est généralement avancée pour expliquer les moindres performances des immigrés. On se serait donc raisonnablement attendu à ce que d'une part, le fait de ne pas parler la langue du test à la maison soit plus pénalisant pour les immigrés que pour les natifs (dans la mesure où les deux parents des immigrés ne sont pas nés dans le pays d'accueil) et d'autre part, à ce que cet effet soit significativement différent entre les deux groupes.

Deux autres variables, relatives à l'environnement scolaire de l'élève, attirent également l'attention. Premièrement, parce qu'elles sont significativement différentes à 1% entre les natifs et les immigrés. Deuxièmement, parce que leur impact sur les scores est significatif. Il s'agit de la région linguistique dans laquelle se situe l'école et du niveau de qualification des enseignants de l'école fréquentée. Par rapport à la région germanophone, les natifs obtiennent des scores inférieurs de 7.2% s'ils fréquentent une école située dans une région francophone. Les immigrés quant à eux, obtiennent des scores supérieurs de 9.7% s'ils fréquentent une école située dans une région italophone. Concernant le niveau de formation des enseignants, les bienfaits en terme de performance scolaire sont importants et très différents: +12.4% pour les natifs et +20% pour les immigrés.

La durée sur le territoire (non significative pour les natifs et +0.06% par mois supplémentaire pour les immigrés) et la taille de la fratrie (-0.9% pour les natifs et -2.7% pour les immigrés) sont également significativement différentes à 1%. Le niveau de revenu de la cellule familiale (+0.15% pour les natifs et +0.25% pour les immigrés) et la taille de l'école sont quant à elle significativement différentes à 10% et 5% respectivement.

NATIFS VS. IMMIGRES DE PREMIERE GENERATION

Il est possible que les résultats ci-dessus proviennent du fait que le groupe des immigrés comprend à la fois les immigrés de deuxième génération et ceux de première génération. Il est donc nécessaire de comparer les résultats de l'estimation du groupe des natifs avec ceux du groupe des immigrés de première génération seulement (colonne 5).

Les résultats indiquent que la langue parlée à la maison demeure une variable non significativement différente entre les deux groupes d'élèves. Par ailleurs, les variables significativement différentes à 1% sont identiques à celles présentées ci-dessus à savoir la durée sur le territoire (non significatif pour les natifs et +0.07% par mois supplémentaire pour les immigrés), la taille de la fratrie (-0.9% contre -3.7%) et la région linguistique de l'école (pour la région italophone, -2.1% contre +12.3%). L'âge ainsi que le niveau

de formation des enseignants (+12.4% contre +22.3%) sont significativement différents à 5%.

La comparaison des résultats de l'estimation des natifs vs. immigrés de deuxième génération (colonne 4) ainsi que celle des immigrés de deuxième génération vs. immigrés de première génération sont moins intéressantes. Dans le premier cas, seules les variables relatives à la région linguistique sont significativement différentes. Dans le second cas, aucune variable n'est significativement différente entre les deux groupes. La langue parlée à la maison demeure non significativement différente quels que soient les groupes considérés.

4.1 Résultats de la décomposition Oaxaca et Blinder

Le tableau 2 présente les résultats de la décomposition du différentiel de scores en lecture entre les natifs et les immigrés (colonne 1), les natifs et les immigrés de première génération (colonne 2), les natifs et les immigrés de deuxième génération (colonne 3) et finalement, les immigrés de deuxième génération et ceux de première génération (colonne 4). La méthode proposée par OAXACA et BLINDER permet de décomposer le différentiel de scores en deux composants. Le premier, C_m , correspond à la part inexpliquée du différentiel et le second, E_m , à la part expliquée.

Tout d'abord, on peut voir que le différentiel total de scores le plus important est entre les natifs et les immigrés de première génération (86.7607). Il est moindre entre les natifs et l'ensemble des immigrés (69.0037), entre les natifs et les immigrés de deuxième génération (50.6044) et entre les immigrés de deuxième et de première génération (36.1563). Les résultats indiquent par ailleurs que 61% des différences de scores observés entre natifs et immigrés sont attribuables à une dotation inférieure des immigrés¹⁸. Il s'agit de la part expliquée du différentiel de score et il en résulte donc, que la part non expliquée est de 39%. La part non expliquée est plus grande dans le cas de la décomposition entre natifs et immigrés de première génération (43%) et plus faible entre natifs et immigrés de deuxième génération (31%).

Il est également possible de déterminer les parts expliquée et inexpliquée pour certaines des variables considérées. Etant donné que les variables ne sont pas toutes significativement différentes, il semble plus pertinent de limiter l'analyse uniquement à celles qui ont été testées comme significativement différentes dans la partie précédente.

¹⁸ Il s'agit des résultats lorsque l'on pose comme hypothèse que la structure courante des scores des natifs s'applique à la fois aux natifs et aux immigrés sur un marché de l'éducation non discriminant (i.e. $W = 1$). Lorsque l'on pose l'hypothèse alternative que c'est la structure courante des scores des immigrés qui s'applique à la fois aux natifs et aux immigrés (i.e. $W = 0$), ce pourcentage est de 67%. Pour plus de détails, voir tableau A7.

Tableau 2: Décomposition OAXACA et BLINDER du différentiel de scores

DIFFERENCE TOTALE	(1) NATIF – IMMIGRE				(2) NATIF – 1 ^{ERE} GENERATION				(3) NATIF – 2 ^{EME} GENERATION				(4) 2 ^{EME} GENERATION – 1 ^{ERE} GENERATION			
	69.0037				86.7607				50.6044				36.1563			
	C _m		E _m		C _m		E _m		C _m		E _m		C _m		E _m	
	26.822	(%)	42.181	(%)	37.118	(%)	49.642	(%)	16.153	(%)	34.451	(%)	7.205	(%)	28.951	(%)
CONSTANTE	259.309				340.396				203.922				136.473			
DUREE (MOIS)	-44.394	97.98	-0.914	2.02	-38.829	95.66	-1.764	4.34	-39.228	99.92	-0.033	0.08	-14.559	52.40	13.227	47.60
LANGUE	2.260	12.76	15.446	87.24	-0.009	0.05	18.565	99.95	3.788	23.67	12.215	76.33	-1.602	27.83	4.156	72.17
GENRE	0.844	73.96	0.297	26.04	0.760	50.50	0.745	49.50	-0.612	78.64	-0.166	21.36	1.313	57.53	0.969	42.47
AGE (MOIS)	-137.20	97.58	3.396	2.42	-197.29	96.49	7.186	3.51	-116.08	99.54	-0.531	0.46	-78.324	94.18	4.837	5.82
NUCLEAIRE	7.569	93.24	-0.549	6.76	10.694	91.36	-1.012	8.64	3.985	98.31	-0.068	1.69	6.387	91.12	-0.622	8.88
NSIB	13.451	98.79	0.164	1.21	21.466	95.47	1.018	4.53	4.733	86.78	-0.721	13.22	15.506	83.94	2.966	16.06
EDUCATION M.	-1.245	16.08	6.498	83.92	-1.397	15.83	7.430	84.17	-1.269	18.66	5.532	81.34	-0.446	16.76	2.216	83.24
EDUCATION F.	4.269	48.63	4.508	51.37	6.664	60.71	4.313	39.29	1.715	26.69	4.711	73.31	4.847	94.25	-0.296	5.75
ISEI	-12.653	63.09	7.402	36.91	-10.614	55.53	8.501	44.47	-13.289	67.97	6.264	32.03	1.706	34.74	3.206	65.26
GERMANOPHONE	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.
FRANCOPHONE	-6.767	91.98	-0.590	8.02	-6.969	93.44	-0.489	6.56	-6.195	89.91	-0.695	10.09	-0.600	94.96	0.032	5.04
ITALOPHONE	-10.515	89.78	1.198	10.22	-10.859	92.04	0.940	7.96	-9.168	86.22	1.465	13.78	-3.746	71.01	1.530	28.99
CITY	1.966	81.62	0.443	18.38	1.662	80.91	0.392	19.09	2.394	82.86	0.495	17.14	-0.418	50.02	-0.418	49.98
SCHLSIZE	10.211	82.57	-2.155	17.43	9.402	81.84	-2.087	18.16	10.579	82.61	-2.226	17.39	-1.108	93.97	0.071	6.03
STRATIO	-9.116	91.06	-0.895	8.94	-14.847	96.20	-0.586	3.80	3.291	73.03	-1.215	26.97	-18.249	96.10	0.740	3.90
TOTHR	-39.047	99.89	-0.042	0.11	-57.282	99.95	-0.030	0.05	-25.133	99.79	-0.054	0.21	-32.168	99.87	0.042	0.13
PROPQUAL	-12.120	60.32	7.974	39.68	-15.831	70.82	6.522	29.18	-7.278	43.43	9.479	56.57	-7.806	67.82	-3.704	32.18
N	5320				4850				4833				957			

Source: PISA 2000. Notes: C_m correspond à la partie non expliquée du différentiel de scores tandis que E_m correspond à la partie expliquée du différentiel de scores. Il s'agit des résultats lorsque W=1 c'est-à-dire lorsque l'on pose l'hypothèse que la structure courante des scores des natifs s'applique à la fois aux natifs et aux immigrés sur un marché de l'éducation non discriminant. La variable dépendante est le score en lecture. Les données ne sont pas pondérées.

Si l'on commence par les variables relatives à l'élève ou sa famille, on observe que la durée sur le territoire et la taille de la fratrie sont des variables fortement *discriminantes* puisque seulement 1% à 2% des différences peuvent être attribués à des différences de durée moyenne ou de taille moyenne entre les deux groupes (natifs vs. immigrés). Tandis que la taille de la fratrie est favorable aux natifs, la durée est en revanche favorable aux immigrés. Cela signifie que le rendement estimé de la durée est davantage valorisé pour les immigrés que pour les natifs. Le revenu de la famille est en revanche une variable moins discriminante puisque 37% des différences peuvent être attribuées à des différences de revenu moyen entre les natifs et les immigrés.

Concernant les caractéristiques de l'école, la taille de la commune dans laquelle se situe l'école ainsi que la taille de l'école sont des variables discriminantes en faveur des natifs. Le niveau de qualification des enseignants de l'école est discriminant dans une moindre mesure puisque 40% des différences peuvent être expliquées par des différences de qualification moyenne entre les deux groupes.

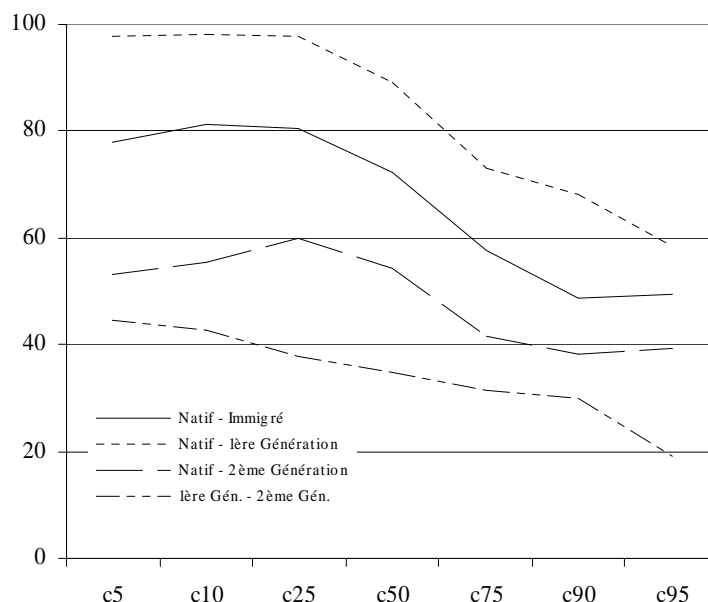
Bien que la langue parlée à la maison ne soit pas significativement différente entre le groupe des natifs et des immigrés, il est intéressant de constater qu'elle est peut être discriminante.

Lorsque l'on regarde la décomposition du différentiel de scores pour les autres groupes, on constate que les parts expliquée et inexpliquée restent stables quels que soient les groupes considérés. La seule variable dont les parts changent est PROPQUAL. La part expliquée de cette variable est plus importante dans le cas de la décomposition des natifs vs. immigrés de deuxième génération que dans celui de la décomposition des natifs vs. immigrés de première génération. Alors que 57% du différentiel de scores peut être attribué à des différences de qualification moyenne entre les natifs et les immigrés de deuxième génération, seul 30% du différentiel de scores peut être attribué à des différences de qualification moyenne entre les natifs et les immigrés de première génération.

4.2 Résultats de la décomposition Juhn, Murphy et Pierce

La décomposition de JUHN, MURPHY et PIERCE (1993) permet d'obtenir une décomposition du différentiel du scores pour l'ensemble de la distribution. Il est dès lors possible de voir comment le différentiel évolue le long de la distribution. Le différentiel total de scores en lecture entre les différents groupes migratoires est présenté au graphique 3 (tableau A9). Les différents centiles sont représentés sur l'axe des abscisses et le différentiel de scores sur l'axe des ordonnées. Alors que le différentiel de scores entre les natifs et les immigrés est de 77.8820 pour le 5^e centile soit le centile des élèves les moins performants, il est de seulement 49.5740 pour le 95^e centile soit le centile des élèves les plus performants en lecture.

Graphique 3: Distribution du différentiel total des scores en lecture



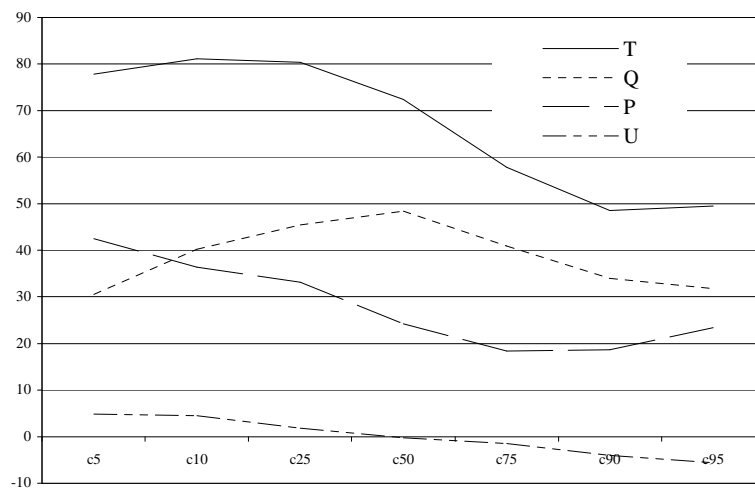
Il est intéressant de constater que le différentiel de scores entre les natifs et les immigrés de première génération est plus important au 5^e centile (97.6980) que le différentiel de scores entre les natifs et les immigrés de deuxième génération (53.2220). L'écart entre ces deux groupes est relativement important dans le 5^e centile (44.476) et beaucoup plus faible dans le 95^e centile (18.956). Cela peut signifier d'une part, que plus les élèves sont performants en lecture et moins l'origine migratoire de l'élève est importante et d'autre part, que plus les élèves sont performants en lecture et moins le fait que les immigrés soient de première ou deuxième génération est importante (étant donné que les lignes du graphique se rapprochent).

Le graphique 4 représente le différentiel total (T) entre les natifs et les immigrés. Il est composé du différentiel dans les quantités observables c'est-à-dire l'effet des caractéristiques (Q), du différentiel dans les prix observables c'est-à-dire l'effet des rendements (P) et du différentiel dans les quantités et prix non mesurables c'est-à-dire l'effet des résidus (U).

Il apparaît que l'effet des caractéristiques (Q) explique le différentiel de scores tout au long de la distribution étant donné qu'il est positif. Il augmente légèrement jusqu'au centile 50 puis décroît jusqu'à retrouver le niveau qu'il avait dans le 5^e centile. Cela signifie que les caractéristiques des immigrés se détériorent dans la première moitié de la distribution puis s'améliorent dans la seconde partie de la distribution. Cela signifie également que les inégalités de caractéristiques entre les natifs et les immigrés sont les plus grandes au niveau de la médiane et les plus faibles aux deux extrémités de la distribution.

L'effet des rendements (P) explique également le différentiel de scores tout au long de la distribution. Il est positif et décroît jusqu'au 75^e centile puis augmente légèrement entre le 90^e et 95^e centile. Cela signifie que les inégalités de rendements (en terme de performance scolaire en lecture) entre les natifs et les immigrés sont plus importantes pour les élèves faibles (5^e centile) que pour les élèves performants (95^e centile). Il est intéressant de voir que l'effet des rendements est supérieur à l'effet des caractéristiques jusqu'au 10^e centile puis devient inférieur.

Graphique 4: Répartition du différentiel total entre les natifs et les immigrés entre les différents composants (Q, P et U)



Finalement, l'effet des résidus (U) est très légèrement croissant entre le 5^e et le 10^e centile puis devient décroissant jusqu'au 95^e centile. Il devient même négatif à partir du 50^e centile. Cela signifie que les inégalités des facteurs inobservés sont de moins en moins importantes entre les natifs et les immigrés au fur et à mesure que l'on se déplace le long de la distribution. Cela signifie également qu'à partir du 50^e centile, les facteurs inobservés sont plus importants pour les immigrés que pour les natifs.

5. Conclusion

L'objet de cette étude était de mettre en relation l'origine migratoire des élèves de l'échantillon national de PISA 2000 avec leurs performances scolaires en lecture. Cette préoccupation semble légitime dans le cas de la Suisse dans la mesure où l'on observe que les scores des immigrés sont systématiquement inférieurs aux scores des natifs. L'objectif était d'une part, de mettre en évidence les caractéristiques qui ont un impact significatif sur le

différentiel de scores entre les natifs et les immigrés et d'autre part, de voir comment évolue le différentiel de scores le long de la distribution.

Les différences de performances entre les natifs et les immigrés sont souvent attribuées au fait que les immigrés ne parlent pas la langue du test à la maison. Dans une première étape, les fonctions de production éducationnelle ont été estimées séparément pour les deux groupes. Les résultats indiquent que le coefficient associé à la langue parlée à la maison est positif et significatif dans les deux estimations. Cela tend à confirmer que, toutes choses égales par ailleurs, un élève qui ne parle pas la langue du test à la maison obtiendra un score moindre en lecture. Il est cependant intéressant de constater que les coefficients ne sont pas significativement différents entre les deux groupes considérés. En effet, le fait de ne pas parler la langue du test à la maison n'a pas d'impact négatif plus important pour les natifs que pour les immigrés, qu'ils soient de deuxième ou première génération.

Les différences de performances entre les natifs et les immigrés sont également souvent attribuées au fait que les immigrés et les natifs ne font pas face au même environnement éducationnel. L'existence de ségrégation résidentielle qui tend à confiner les immigrés dans des quartiers dans lesquelles les écoles sont en moyenne de moins bonnes qualités peut être une explication. Les statistiques descriptives conduisent à confirmer cette hypothèse puisqu'il apparaît que les immigrés ont tendance à fréquenter des écoles en moyenne plus grandes et dans lesquelles le niveau de qualification des enseignants est en moyenne inférieur. Les résultats indiquent que le niveau de formation des enseignants a un impact sur les performances en lecture significativement différent entre les natifs (+12.4%) et les immigrés (+20%). La décomposition d'OAXACA et BLINDER du différentiel de scores indique par ailleurs que 60% des différences peuvent être expliquées par des différences de qualification moyenne entre les deux groupes. Il en découle que la part inexpliquée du différentiel est d'environ 40%. Par ailleurs, il apparaît que le rendement du niveau de formation des enseignants est différent entre les natifs et les immigrés, en l'occurrence en faveur des immigrés. Autrement dit, les immigrés bénéficient davantage d'avoir un enseignant qualifié que les natifs. Cela signifie que les performances inférieures des immigrés pourraient être attribuées au fait qu'en moyenne, ils font face à des enseignants moins formés (ségrégation éducationnelle).

Finalement, lorsque l'on s'intéresse à la décomposition du différentiel de scores pour l'ensemble de la distribution, il apparaît que l'origine migratoire de l'élève est moins pénalisante lorsque ceux-ci sont de bons élèves (i.e. en haut de la distribution). Les résultats indiquent également que plus les élèves sont performants en lecture et moins le fait que les immigrés soient de première ou deuxième génération est importante. La décomposition entre les différents effets tend à confirmer cette tendance puisque les inégalités de rendements par exemple sont plus importantes pour les élèves faibles que pour les élèves performants.

Les résultats de cette étude soulèvent la question de la ségrégation résidentielle et éducationnelle. Cette question est au cœur du débat public depuis au moins la fin des années 1990. En effet, des interventions politiques ont été déposées à diverses reprises à cette époque afin que soient créés des

classes séparées pour les enfants suisses et pour les enfants étrangers (CFR, 1999). Malheureusement, les données PISA ne permettaient pas d'aborder cette problématique dans le cadre de la présente analyse. Une telle étude nécessiterait l'utilisation de données longitudinales qui nous renseignent sur la répartition des élèves d'une même cohorte entre les différentes filières d'une année à l'autre. Il serait également nécessaire de disposer d'un identifiant de l'élève, de sa classe, du code postal de son école et finalement du code postal de son lieu de résidence. Il serait dès lors possible d'établir si la probabilité pour un élève d'accéder à certaines filières est uniquement conditionnelle à ses résultats scolaires antérieurs ou à d'autres caractéristiques individuelles (telles que l'origine migratoire par exemple). Dans l'état actuel des données à disposition, cette piste de recherche importante pour une meilleure compréhension des inégalités de performances scolaires entre les natifs et les immigrés doit malheureusement être remise à un projet de recherche ultérieur.

Annexes

Tableau A1: Définitions des variables

Variabiles	Définitions
LECTURE	Score en lecture (min = 155.898 ; max = 769.974)
ORIGINE MIGRATOIRE	
NATIFS	1 = si l'élève est un natif ; 0 = sinon
SNDGEN	1 = si l'élève est un immigré de seconde génération ; 0 = sinon
FSTGEN	1 = si l'élève est un immigré de première génération ; 0 = sinon
DUREE	Durée de vie en Suisse en mois (min = 12 ; max = 204)
LANGUE	1 = si la langue la plus souvent parlée à la maison est celle du test ; 0 = sinon
CARACTERISTIQUES INDIVIDUELLES	
GENRE	1 = si l'élève est une fille ; 0 = si l'élève est un garçon
AGE	Age de l'élève en mois (min = 142 ; max = 228)
ORIGINE SOCIALE	
NUCLEAIRE	1 = si nucléaire (les élèves ont déclaré vivre avec leur père et leur mère) ; 0 = sinon
NSIB	Nombre de frère(s) et sœur(s) (min = 0 ; max = 12)
EDUCM	1 = si la mère a terminé des études post-obligatoires ; 0 = sinon
EDUCF	1 = si le père a terminé des études post-obligatoires ; 0 = sinon
ISEI	Indice socio-économique du père ou de la mère
ENVIRONNEMENT SCOLAIRE	
GERMANOPHONE	1 = si l'école se situe dans un canton germanophone ; 0 = sinon
FRANCOPHONE	1 = si l'école se situe dans un canton francophone ; 0 = sinon
ITALOPHONE	1 = si l'école se situe dans un canton italophone ; 0 = sinon
CITY	1 = si le nombre d'habitants de la commune dans laquelle se situe l'école est entre 100'000 et un million environ ; 0 = sinon
SCHLSIZE	Nombre total d'élèves inscrits dans l'établissement au 31 mars 2000 (min = 14 ; max = 1942)
STRATIO	Ratio entre la taille de l'école et le nombre de professeurs (min = 1.89 ; max = 49.3)
TOTHR	Nombre d'heures d'encadrement par année (min = 579 ; max = 1267)
PROPQUAL	Proportion de professeurs avec un niveau ISCED 5 en pédagogie (min = 0 ; max = 1)

Source: PISA 2000.

Tableau A2: Statistiques descriptives

Variables	NATIF		DEUXIEME GENERATION		PREMIERE GÉNÉRATION		TOTAL	
	0.8201 (0.3841)		0.0883 (0.2838)		0.0915 (0.2884)		1.0000	
	Moy.	SD	Moy.	SD	Moy.	SD	Moy.	SD
LECTURE	521.5108[†]	79.8072	470.9064^{°*}	86.3864	434.7501[°]	91.3226	509.0979	86.0502
DUREE	184.219[†]	17.7161	182.885[*]	18.9237	113.15[°]	46.6645	177.5876	30.1090
LANGUE	0.9491[†]	0.2198	0.4489^{°*}	0.49797	0.1889[°]	0.3918	0.8353	0.3709
GENRE	0.4997	0.5001	0.50857	0.5005	0.4600	0.4989	0.4968	0.5000
AGE	187.564[†]	7.1400	187.245[*]	8.6511	191.889[°]	9.2534	187.9318	7.6037
NUCLEAIRE	0.7953[†]	0.4035	0.8000	0.4004	0.8645[°]	0.3426	0.8021	0.3985
NSIB	1.6541	1.0185	1.4936^{°*}	1.0484	1.8809[°]	1.2432	1.6607	1.0468
EDUCM	0.6259[†]	0.4839	0.3617[°]	0.48104	0.2710[°]	0.4450	0.5701	0.4951
EDUCF	0.6672[†]	0.4713	0.3915[°]	0.48864	0.4148[°]	0.4932	0.6197	0.4855
ISEI	46.9425[†]	17.1193	38.9872[°]	15.5598	36.1458[°]	15.0711	45.2513	17.2002
GERMANOPHONE	0.7043[†]	0.4564	0.58727[°]	0.4929	0.6304[°]	0.4832	0.6872	0.4637
FRANCOPHONE	0.2143	0.4104	0.19577	0.3972	0.2012	0.4013	0.2115	0.4084
ITALOPHONE	0.0814[†]	0.2734	0.21707[°]	0.4127	0.1684[°]	0.3746	0.1013	0.3018
CITY	0.0628[†]	0.2426	0.1703[°]	0.37627	0.1478[°]	0.3553	0.0801	0.2714
SCHLSIZE	402.098[†]	294.261	448.2617[°]	327.18	445.372	336.841	410.1378	301.8463
STRATIO	12.1359[†]	3.9945	11.3956[°]	2.7931	11.7789	3.6258	12.0378	3.8762
TOTHR5	982.562	85.1251	984.2298	84.2291	983.501	84.0475	982.7953	84.9340
PROPQUAL	0.5910[†]	0.3273	0.4446[°]	0.36611	0.4903[°]	0.3567	0.5689	0.3371
Observations	4363		470		487		5320	

Source: PISA 2000. Notes: les statistiques présentées sont celles de l'échantillon en lecture. Les données ne sont pas pondérées.

[†] signifie que le groupe des natifs est significativement différent à 1% du groupe des immigrés (2^{ème} et 1^{ère} génération confondues).

[°] signifie que le groupe des immigrés de 2^{ème} génération est significativement différent à 1% du groupe des natifs.

[°] signifie que le groupe des immigrés de 1^{ère} génération est significativement différent à 1% du groupe des natifs.

* signifie que le groupe des immigrés de 2^{ème} génération est significativement différent à 1% du groupe des immigrés de 1^{ère} génération.

Tableau A3: Pays de naissance des élèves et des parents

	NATIF			DEUXIEME GENERATION			PREMIERE GENERATION		
	ELEVE (%)	MERE (%)	PERE (%)	ELEVE (%)	MERE (%)	PERE (%)	ELEVE (%)	MERE (%)	PERE (%)
SUISSE	4225 (96.84)	3972 (91.04)	3968 (90.95)	470 (100)					
ALL. / AUTRICHE	16 (0.37)	105 (2.41)	84 (1.93)		18 (3.83)	13 (2.77)	19 (3.90)	16 (3.29)	16 (3.29)
FRANCE / BELGIQUE	15 (0.34)	58 (1.33)	44 (1.01)		31 (6.60)	25 (5.32)	8 (1.64)	8 (1.64)	9 (1.85)
ITALIE	9 (0.21)	50 (1.15)	143 (3.28)		156 (33.19)	175 (37.23)	39 (8.01)	41 (8.42)	44 (9.03)
ESPAGNE	1 (0.02)	6 (0.14)	17 (0.39)		46 (9.79)	43 (9.15)	19 (3.90)	18 (3.70)	20 (4.11)
PORTUGAL	1 (0.02)	6 (0.14)	8 (0.18)		25 (5.32)	18 (3.83)	61 (12.53)	61 (12.53)	61 (12.53)
EX-YOUGOSLAVIE	1 (0.02)	16 (0.37)	5 (0.11)		63 (13.40)	61 (12.98)	169 (34.70)	170 (34.91)	168 (34.50)
ALBANIE / KOSOVO	1 (0.02)	1 (0.02)	0 (0)		6 (1.28)	5 (1.06)	86 (17.66)	85 (17.45)	87 (17.86)
TURQUIE	2 (0.05)	3 (0.07)	4 (0.09)		48 (10.21)	55 (11.70)	33 (6.78)	33 (6.78)	33 (6.78)
AUTRES PAYS	92 (2.11)	146 (3.35)	90 (2.06)		77 (16.38)	75 (15.96)	53 (10.88)	55 (11.29)	49 (10.06)
TOTAL	4363 (100)	4363 (100)	4363 (100)	470 (100)	470 (100)	470 (100)	487 (100)	487 (100)	487 (100)

Source: PISA 2000. Notes: Les données ne sont pas pondérées. All. signifie Allemagne.

Tableau A4: Estimation MCO des fonctions de production éducationnelle

Variables	(1) TOTAL	
	Coeff.	t
DUREE (MOIS)	0.1112	(2.20)**
LANGUE	26.3235	(7.14)***
GENRE	18.5959	(9.68)***
AGE (MOIS)	-1.6038	(11.27)***
NUCLEAIRE	12.7114	(5.09)***
NSIB	-6.3936	(6.37)***
EDUCATION M	22.0021	(8.95)***
EDUCATION F	14.0450	(5.49)***
ISEI	0.8398	(13.75)***
GERMANOPHONE	REF.	REF.
FRANCOPHONE	-31.5071	(12.12)***
ITALOPHONE	3.7484	(0.92)
CITY	-8.9963	(2.29)*
SCHLSIZE	0.0436	(10.30)***
STRATIO	-1.4919	(5.25)***
TOTHR	0.0382	(3.15)***
PROPQUAL	71.4216	(20.33)***
CONSTANTE	633.3561	(21.11)***
NATIF	réf.	réf.
2 ^{EME} GENERATION	-15.5356	(3.80)***
1 ^{ERE} GENERATION	-25.4131	(4.59)***
N		5320
R ²		0.3490

Source: PISA 2000. Notes: * significatif à 10%, ** significatif à 5%, *** significatif à 1%. Statistiques t robustes entre parenthèses. La variable dépendante est le score en lecture. Les données ne sont pas pondérées.

Tableau A5: Comparaison des coefficients des différentes estimations

Variables	(1)		(2)		(3)		(4)	
	NATIF-IMMIGRE		NATIF-2 ^{EME} GENERATION		NATIF-1 ^{ERE} GENERATION		2 ^{EME} GEN-1 ^{ERE} GEN	
	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
DUREE (MOIS)	0.2764	(5.10)***	0.1897	(0.85)	0.3183	(4.00)***	0.3183	(3.95)***
DUREE*D	-0.3012	(3.41)***	-0.2145	(0.92)	-0.3432	(3.24)***	-0.1287	(0.54)
LANGUE	17.2838	(2.85)***	15.9832	(1.89)*	24.4656	(2.57)**	24.4656	(2.53)**
LANGUE*D	7.1367	(0.91)	8.4373	(0.86)	-0.0451	(0.00)	-8.4824	(0.66)
GENRE	17.0115	(3.62)***	19.9594	(2.94)***	17.1052	(2.55)**	17.1052	(2.51)**
GENRE*D	1.7452	(0.34)	-1.2027	(0.17)	1.6515	(0.24)	2.8542	(0.29)
AGE (MOIS)	-0.9379	(3.19)***	-1.0415	(2.28)**	-0.6333	(1.48)	-0.6333	(1.46)
AGE*D	-0.7236	(2.14)**	-0.6200	(1.28)	-1.0281	(2.24)**	-0.4082	(0.64)
NUCLEAIRE	5.5437	(0.88)	9.6506	(1.13)	2.2617	(0.23)	2.2617	(0.23)
NUCLEAIRE*D	9.0885	(1.33)	4.9816	(0.56)	12.3705	(1.21)	7.3888	(0.56)
NSIB	-12.4462	(4.65)***	-7.6591	(1.89)*	-15.9028	(5.41)***	-15.9028	(5.33)***
NSIB*D	7.9558	(2.76)***	3.1687	(0.75)	11.4124	(3.64)***	8.2437	(1.62)
EDUCATION M	24.8795	(3.80)***	24.4436	(2.93)***	26.0896	(2.56)**	26.0896	(2.53)**
EDUCAT.M*D	-3.9446	(0.56)	-3.5087	(0.40)	-5.1547	(0.49)	-1.6460	(0.12)
EDUCATION F	6.5032	(1.07)	12.7057	(1.50)	1.0199	(0.11)	1.0199	(0.11)
EDUCAT.F*D	10.5830	(1.58)	4.3805	(0.49)	16.0663	(1.71)*	11.6859	(0.93)
ISEI	1.1244	(6.19)***	1.1282	(4.50)***	1.0810	(4.02)***	1.0810	(3.96)***
ISEI*D	-0.3370	(1.75)*	-0.3409	(1.32)	-0.2937	(1.06)	0.0472	(0.13)
GERMANOPHONE	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.
FRANCOPHONE	-3.3633	(0.46)	-5.7986	(0.55)	-2.8172	(0.27)	-2.8172	(0.27)
FRANCOPH.*D	-34.0843	(4.33)***	-31.6489	(2.89)***	-34.6304	(3.25)***	-2.9815	(0.20)
ITALOPHONE	43.8911	(5.40)***	31.4450	(2.66)***	53.6955	(4.69)***	53.6955	(4.63)***
ITALOPH.*D	-54.6898	(5.83)***	-42.2437	(3.32)***	-64.4942	(5.22)***	-22.2505	(1.33)
CITY	-16.9874	(2.17)**	-18.6769	(1.66)*	-15.8484	(1.47)	-15.8484	(1.45)
CITY*D	12.3774	(1.37)	14.0670	(1.16)	11.2384	(0.96)	-2.8286	(0.18)
SCHLSIZE	0.0254	(2.84)***	0.0246	(1.73)*	0.0271	(2.52)**	0.0271	(2.48)**
SCHLSIZE*D	0.0229	(2.26)**	0.0236	(1.57)	0.0211	(1.80)*	-0.0025	(0.14)
STRATIO	-0.8550	(0.91)	-1.9302	(1.32)	-0.3809	(0.34)	-0.3809	(0.33)
STRATIO*D	-0.7865	(0.80)	0.2888	(0.19)	-1.2605	(1.08)	-1.5493	(0.83)
TOTHS	0.0719	(2.28)**	0.0577	(1.17)	0.0904	(2.33)**	0.0904	(2.29)**
TOTHS*D	-0.0397	(1.17)	-0.0255	(0.50)	-0.0582	(1.42)	-0.0327	(0.51)
PROPQUAL	90.6503	(10.44)***	81.1131	(6.48)***	97.0341	(7.93)***	97.0341	(7.81)***
PROPQUAL*D	-25.9044	(2.73)***	-16.3673	(1.25)	-32.2883	(2.52)**	-15.9210	(0.90)
NATIF	259.3089	(3.48)***	203.922	(2.11)**	340.396	(3.30)***		
SCDGEN							136.4732	(1.01)
CONSTANTE	420.0244	(6.30)***	475.411	(5.25)***	338.938	(3.48)***	338.9377	(3.43)***
N	5320		4833		4850		957	
R ²	0.3605		0.3005		0.3526		0.3798	

Source: PISA 2000.

Notes: * significatif à 10%, ** significatif à 5%, *** significatif à 1%. Statistiques t robustes entre parenthèses. La variable dépendante est le score en lecture. Les données sont non pondérées. D équivaut à la variable binaire natif (1 = natif, 0 sinon) pour les colonnes (1), (2) et (3) et à la variable binaire scdgen (1= immigré de deuxième génération et 0 sinon) pour la colonne (4). La colonne (1) comprend les 4363 élèves natifs et les 957 élèves immigrés, la colonne (2) les 4363 élèves natifs et les 470 élèves immigrés de deuxième génération, la colonne (3) les 4363 élèves natifs et les 487 élèves immigrés de première génération et la colonne (4) les 470 élèves immigrés de deuxième génération et les 487 élèves immigrés de première génération.

Les coefficients de la colonne (1) permettent de faire la distinction entre les coefficients de la régression des natifs et ceux de la régression des immigrés. Pour cela, on teste l'hypothèse nulle $H_0 : B_{\text{natif}} = B_{\text{immigré}}$ où B_{natif} est le coefficient de la régression pour les natifs et $B_{\text{immigré}}$ est le coefficient de la régression pour les immigrés. Par exemple, le terme **genre*D** teste l'hypothèse nulle $H_0 : \text{genre}_{\text{natif}} = \text{genre}_{\text{immigré}}$. Etant donné que la valeur t est non significative, cela signifie que le coefficient de la régression $\text{genre}_{\text{natif}}$ n'est pas significativement différent de $\text{genre}_{\text{immigré}}$.

Tableau A6: Comparaison des coefficients des différentes estimations

Variables	(1)		(2)		(3)		(4)	
	IMMIGRE-NATIF		2 ^{EME} GENERATION-NATIF		1 ^{ERE} GENERATION-NATIF		1 ^{ERE} GEN-2 ^{EME} GEN	
	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
DUREE (MOIS)	-0.0248	(0.36)	-0.0248	(0.36)	-0.0248	(0.36)	0.1897	(0.84)
DUREE*D	0.3012	(3.41)***	0.2145	(0.92)	0.3432	(3.24)***	0.1287	(0.54)
LANGUE	24.4204	(4.91)***	24.4204	(4.91)***	24.4204	(4.91)***	15.9832	(1.86)*
LANGUE*D	-7.1367	(0.91)	-8.4373	(0.86)	0.0451	(0.00)	8.4824	(0.66)
GENRE	18.7567	(9.00)***	18.7567	(8.99)***	18.7567	(9.00)***	19.9594	(2.90)***
GENRE*D	-1.7452	(0.34)	1.2027	(0.17)	-1.6515	(0.24)	-2.8542	(0.29)
AGE (MOIS)	-1.6614	(9.96)***	-1.6614	(9.95)***	-1.6614	(9.95)***	-1.0415	(2.25)**
AGE*D	0.7236	(2.14)**	0.6200	(1.28)	1.0281	(2.24)**	0.4082	(0.64)
NUCLEAIRE	14.6322	(5.43)***	14.6322	(5.43)***	14.6322	(5.43)***	9.6506	(1.12)
NUCLEAIRE*D	-9.0885	(1.33)	-4.9816	(0.56)	-12.3705	(1.21)	-7.3888	(0.56)
NSIB	-4.4903	(4.16)***	-4.4903	(4.15)***	-4.4903	(4.15)***	-7.6591	(1.86)*
NSIB*D	-7.9558	(2.76)***	-3.1687	(0.75)	-11.4124	(3.64)***	-8.2437	(1.62)
EDUCATION M	20.9349	(7.96)***	20.9349	(7.96)***	20.9349	(7.96)***	24.4436	(2.88)***
EDUCAT.M*D	3.9446	(0.56)	3.5087	(0.40)	5.1547	(0.49)	1.6460	(0.12)
EDUCATION F	17.0862	(6.08)***	17.0862	(6.08)***	17.0862	(6.08)***	12.7057	(1.48)
EDUCAT.F*D	-10.5830	(1.58)	-4.3805	(0.49)	-16.0663	(1.71)*	-11.6859	(0.93)
ISEI	0.7874	(12.19)***	0.7874	(12.18)***	0.7874	(12.18)***	1.1282	(4.44)***
ISEI*D	0.3370	(1.75)*	0.3409	(1.32)	0.2937	(1.06)	-0.0472	(0.13)
GERMANOPHONE	REF.	REF.	REF.	REF.	REF.	REF.	REF.	REF.
FRANCOPHONE	-37.4475	(13.54)***	-37.4475	(13.54)***	-37.4475	(13.54)***	-5.7986	(0.54)
FRANCOPH.*D	34.0843	(4.33)***	31.6489	(2.89)***	34.6304	(3.25)***	2.9815	(0.20)
ITALOPHONE	-10.7987	(2.31)**	-10.7987	(2.31)**	-10.7987	(2.31)**	31.4450	(2.62)***
ITALOPH.*D	54.6898	(5.83)***	42.2437	(3.32)***	64.4942	(5.22)***	22.2505	(1.33)
CITY	-4.6099	(1.02)	-4.6099	(1.02)	-4.6099	(1.02)	-18.6769	(1.63)
CITY*D	-12.3774	(1.37)	-14.0670	(1.16)	-11.2384	(0.96)	2.8286	(0.18)
SCHLSIZE	0.0482	(10.19)***	0.0482	(10.18)***	0.0482	(10.18)***	0.0246	(1.70)*
SCHLSIZE*D	-0.0229	(2.26)**	-0.0236	(1.57)	-0.0211	(1.80)*	0.0025	(0.14)
STRATIO	-1.6415	(5.54)***	-1.6415	(5.54)***	-1.6415	(5.54)***	-1.9302	(1.30)
STRATIO*D	0.7865	(0.80)	-0.2888	(0.19)	1.2605	(1.08)	1.5493	(0.83)
TOTHR	0.0322	(2.46)**	0.0322	(2.46)**	0.0322	(2.46)**	0.0577	(1.15)
TOTHR*D	0.0397	(1.17)	0.0255	(0.50)	0.0582	(1.42)	0.0327	(0.51)
PROPQUAL	64.7458	(16.91)***	64.7458	(16.90)***	64.7458	(16.90)***	81.1131	(6.38)***
PROPQUAL*D	25.9044	(2.73)***	16.3673	(1.25)	32.2883	(2.52)**	15.9210	(0.90)
IMMIGRE	-259.3089	(3.48)***	-203.922	(2.11)**	-340.396	(3.30)***		
FSTGEN							-136.4732	(1.01)
CONSTANTE	679.3333	(20.29)***	679.333	(20.28)***	679.333	(20.28)***	475.4109	(5.18)***
N	5320		4833		4850		957	
R ²	0.3605		0.3005		0.3526		0.3798	

Source: PISA 2000.

Notes: * significatif à 10%, ** significatif à 5%, *** significatif à 1%. Statistiques t robustes entre parenthèses. La variable dépendante est le score en lecture. Les données sont non pondérées. D équivaut à la variable binaire immig (1 = immigré, 0 sinon) pour les colonnes (1), (2) et (3) et à la variable binaire fstgen (1= immigré de première génération et 0 sinon) pour la colonne (4). La colonne (1) comprend les 4363 élèves natifs et les 957 élèves immigrés, la colonne (2) les 4363 élèves natifs et les 470 élèves immigrés de deuxième génération, la colonne (3) les 4363 élèves natifs et les 487 élèves immigrés de première génération et la colonne (4) les 470 élèves immigrés de deuxième génération et les 487 élèves immigrés de première génération.

Les coefficients de la colonne (1) permettent de faire la distinction entre les coefficients de la régression des natifs et ceux de la régression des immigrés. Pour cela, on teste l'hypothèse nulle $H_0 : B_{natif} = B_{immigré}$ où B_{natif} est le coefficient de la régression pour les natifs et $B_{immigré}$ est le coefficient de la régression pour les immigrés. Par exemple, le terme $genre*D$ teste l'hypothèse nulle $H_0 : genre_{natif} = genre_{immigré}$. Etant donné que la valeur t est non significative, cela signifie que le coefficient de la régression $genre_{natif}$ n'est pas significativement différent de $genre_{immigré}$.

Tableau A7: Décomposition OAXACA et BLINDER
(répartition du différentiel de scores selon la référence considérée)

	Décomposition Two Fold			
	W=1 OAXACA (1973)	W=0 BLINDER (1973)	W=0.5 REIMERS (1983)	W=omega NEUMARK (1988)
NATIFS – IMMIGRES (69.00368)				
E (expliqué)	42.18168	49.73008	45.95588	60.08262
C (non expliqué)	26.822	19.2736	23.0478	8.921063
NATIFS – 2^{EME} GEN. (50.6044)				
E (expliqué)	34.45124	32.24523	33.34824	40.28324
C (non expliqué)	16.15316	18.35917	17.25617	10.32116
NATIFS – 1^{ERE} GEN. (86.76069)				
E (expliqué)	49.64227	74.29556	61.96892	77.35594
C (non expliqué)	37.11841	12.46513	24.79177	9.404744
2^{EME} GEN. – 1^{ERE} GEN. (36.15628)				
E (expliqué)	28.95131	41.89125	35.42128	36.79668
C (non expliqué)	7.20497	-5.734966	0.7350023	-0.6403941

Source: PISA 2000. Notes: Les résultats ont été estimés à partir de l'échantillon des 4363 natifs, 470 immigrés de deuxième et 487 immigrés de première génération. Le différentiel total de score est tel que: $A^n - A^m = 69.00368$.

Une décomposition alternative du différentiel de scores peut être obtenu en posant comme hypothèse que la structure courante des scores des immigrés s'appliquera à la fois aux natifs et aux immigrés sur un marché de l'éducation non discriminant ($W = 0$). Le différentiel de scores entre natifs et immigrés peut alors être écrit de la manière suivante:

$$\overline{A^n} - \overline{A^m} = \underbrace{(\hat{\beta}^n - \hat{\beta}^m) \overline{Z^n}}_{C_n} + \hat{\beta}^m \underbrace{(\overline{Z^n} - \overline{Z^m})}_{E_n}$$

Dans cette équation, le terme C_n représente les effets estimés de la discrimination. Ils correspondent à la différence entre la manière dont l'équation des natifs estimerait les caractéristiques du groupe des immigrés et celle dont l'équation des scores des immigrés l'estime en fait. Le terme E_n représente les effets estimés des différences de caractéristiques individuelles c'est-à-dire la valeur de l'avantage en dotations possédés par le groupe des natifs tel qu'évalué par l'équation des scores du groupe des immigrés. La part du différentiel attribuable à la discrimination correspond à C_n .

Tableau A8: Décomposition OAXACA et BLINDER - natifs vs. immigrants

Variables	$\hat{\beta}^n$	$\hat{\beta}^m$	\bar{Z}^n	\bar{Z}^m	E	$U + C$
DUREE	-0.0248198	0.2763623	184.2095	147.3981	-0.913651586	-44.39366929
LANGUE	24.42044	17.28376	0.9491176	0.3166144	15.44600645	2.259575656
GENRE	18.7567	17.01148	0.4996562	0.4838036	0.297342462	0.844343719
AGE	-1.66145	-0.9378516	187.5641	189.6082	3.396169945	-137.2001901
NUCLEAIRE	14.6322	5.543723	0.7953243	0.8328109	-0.548511429	7.56898271
NSIB	-4.490334	-12.44617	1.654137	1.6907	0.164180082	13.45093193
EDUCM	20.93488	24.87949	0.6259455	0.3155695	6.497684315	-1.244798605
EDUCF	17.0862	6.503241	0.6672015	0.4033438	4.508325434	4.268570898
ISEI	0.7873513	1.1244	46.94247	37.54127	7.402047042	-12.65323625
GERMANOPHONE	Réf.	Réf.	0.7043319	0.6091954	Réf.	Réf.
FRANCOPHONE	-37.44753	-3.363268	0.2143021	0.1985371	-0.59036031	-6.766990533
ITALOPHONE	-10.79869	43.89112	0.081366	0.1922675	1.197590919	-10.51507304
CITY	-4.609946	-16.98739	0.0628008	0.1588297	0.442688043	1.965905717
SCHLSIZE	0.0482266	0.0253721	402.0981	446.791	-2.155386611	10.21118491
STRATIO	-1.641462	-0.8549935	12.13592	11.59064	-0.895056399	-9.115673255
TOTHR	0.0321696	0.0718569	982.562	983.8589	-0.041720754	-39.04670332
PROPQUAL	64.74581	90.65025	0.5910429	0.4678788	7.974359417	-12.1201383
CONSTANTE	679.3333	420.0244				259.3089
N	4363	957	4363	957		
R ²	0.2710	0.3729				
Total		Différence totale = 69.00368			42.1817 (61.13%)	26.822 (38.87%)

Source: PISA 2000.

Notes: * significatif à 10%; ** significatif à 5%; *** significatif à 1%. Statistiques t robustes entre parenthèses.

$$E = \hat{\beta}^n (\bar{Z}^n - \bar{Z}^m) \text{ et } U = \bar{Z}^m (\hat{\beta}^n - \hat{\beta}^m) \text{ et } U = \bar{Z}^m (\hat{\beta}^n - \hat{\beta}^m). W = 1.$$

Tableau A9: Décomposition JUHN, MURPHY et PIERCE

	NATIF - IMMIGRE				NATIF - 1 ^{ERE} GENERATION				NATIF - 2 ^{EME} GENERATION				1 ^{ERE} GENERATION - 2 ^{EME} GENERATION			
	T	Q	P	U	T	Q	P	U	T	Q	P	U	T	Q	P	U
moyenne	69.0037	41.9083	26.8220	0.2734	86.7607	49.1207	37.1184	.521546	50.6044	33.9165	16.1532	0.5348	36.1563	28.5973	7.205	0.3540
p5	77.8820	30.5050	42.5254	4.8516	97.6980	34.5664	57.9586	5.1730	53.2220	22.7053	23.816	6.7007	44.4760	26.7194	18.6773	-0.9207
p10	81.112	40.2195	36.3988	4.4937	98.114	43.1622	53.4658	1.4861	55.334	32.5110	20.4805	2.3425	42.78	30.1585	12.6317	-0.0101
p25	80.4060	45.4620	33.1336	1.8104	97.8	53.7389	42.1771	1.8840	59.986	37.6634	20.1495	2.1731	37.814	29.3896	9.2040	-0.7797
médiane	72.4000	48.4206	24.2495	-2.701	89.0740	57.3773	30.2437	1.4530	54.2000	39.9297	13.0648	1.2055	34.8740	28.4234	5.16083	1.2898
p75	57.8200	40.9151	18.4018	-1.4969	73.028	50.0174	21.0544	1.9562	41.496	33.2336	9.9663	-1.7039	31.5320	28.3902	0.7143	2.4276
p90	48.5880	34.0269	18.6302	-4.0691	67.986	39.1807	33.612	-4.8147	38.208	28.9286	11.6981	-2.4187	29.778	25.3048	6.1467	-1.6736
p95	49.5740	31.7945	23.3704	-5.5909	58.404	33.0351	36.7726	-11.4037	39.448	28.9333	16.9411	-6.4264	18.956	16.512	7.9783	-5.5343
d90-10	-32.5240	-6.1926	-17.7686	-8.5628	-30.128	-3.9815	-19.8458	-6.3008	-17.126	-3.5824	-8.7824	-4.7612	-13.002	-4.8536	-6.4849	-1.6634
d75-25	-22.5860	-4.5469	-14.732	-3.3073	-24.772	-3.7215	-21.1227	0.0721	-18.490	-4.4298	-10.1832	-3.877	-6.282	-0.9995	-8.4898	3.2073
d90-50	-23.812	-14.3937	-5.6193	-3.79895	-21.0881	-18.1967	3.3763	-6.2677	-15.992	-11.001	-1.3667	-3.6242	-5.096	-3.1185	0.9859	-2.9634
d50-10	-8.712	8.2011	-12.149	-4.7639	-9.0399	14.2151	-23.222	-0.0330	-1.134	7.4187	-7.4157	-1.137	-7.906	-1.7351	-7.4708	1.3
N		5320				4850				4833				957		

Source: PISA 2000. Notes: la référence est natif. L'échantillon comprend 4363 natifs et 957 immigrants (de deuxième et première génération). Les données sont non pondérées. T représente la différence totale, Q les quantités observables, P les prix observables et U les quantités et prix inobservables. La contribution aux différences est égale à Q+P+U.

Bibliographie

- AMMERMÜLLER, A. (2008), "PISA: What Makes the Difference?", *Empirical Economics*, 33 (2), pp.263-287.
- BLAU, F. and L. KAHN (1992), "The Gender Earnings Gap: Learning from International Comparisons", *The American Economic Review*, 82 (2), pp. 533-538.
- BLINDER, A. (1973), "Wage Discrimination: Reduced Form and Structural Estimates", *The Journal of Human Resources*, 8 (4), pp. 436-455.
- BOX, G. E. P. and D. R. COX (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society*, 26 (2), pp. 211-252.
- COMMISSION FEDERALE CONTRE LE RACISME (1999). "Des classes séparées? Dossier sur les demandes politiques de ségrégation des enfants parlant une langue étrangère à l'école", EKR/CFR.
- COLLINS, W. J. and R. A. MARGO (2006), "Historical Perspectives on Racial Differences in Schooling in the United States", *Handbook of The Economics of Education*, Volume 1, pp. 107-150.
- COOK, M. D. and W. N. EVANS (2000), "Families or Schools? Explaining the Convergence in White and Black Academic Performance", *Journal of Labor Economics*, 18 (4), pp. 729-754.
- ENTORF, H. and N. MINOIU (2005), "What a Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration", *German Economic Review*, 6 (3), pp. 355-376.
- GANZEBOOM, H. B. G., P. M. DE GRAAF and D. J. TREIMAN (1992), "A Standard International Socio-Economic Index of Occupational Status", *Social Science Research*, 21, pp. 1-56.
- JUHN, C., K. MURPHY and B. PIERCE (1993), "Wage Inequality and the Rise in Returns to Skill", *The Journal of Political Economy*, 101 (3), pp. 410-442.
- MEUNIER, M. (2006), "Fonctions de production éducationnelle: le cas de la Suisse", *Swiss Journal of Economics and Statistics*, 142 (4), pp. 579-615.
- NEAL, D. (2006), "Why Has Black-White Skill Convergence Stopped?", *Handbook of The Economics of Education*, Volume 1, pp. 511-573.
- OAXACA, R. (1973), "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 14 (3), pp. 693-709.
- OFFICE FEDERAL DE LA STATISTIQUE et CONFERENCE SUISSE DES DIRECTEURS CANTONAUX DE L'INSTRUCTION PUBLIQUE (2002), *Préparés pour la vie ? Les compétences des jeunes - Rapport national de l'enquête PISA 2000*, Neuchâtel.

- OFFICE FEDERAL DE LA STATISTIQUE et CONFERENCE SUISSE DES DIRECTEURS CANTONAUX DE L'INSTRUCTION PUBLIQUE (2005), *PISA 2003: Compétences pour l'avenir, Deuxième rapport national*, Neuchâtel/Berne.
- OFFICE FEDERAL DE LA STATISTIQUE (2005), *Annuaire statistique de la Suisse 2005*.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2002), *Programme for International Student Assessment (PISA) – Manuel for the PISA 2000 Database*, OECD Publishing.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2002), *PISA 2000 Technical Report*, OECD Publishing.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (2006), *Where immigrant students succeed – A comparative review of performance and engagement in PISA 2003*, Paris: OECD.
- RIVKIN, S. AND F. WELCH (2006), “Has School Desegregation Improved Academic and Economic Outcomes for Blacks?”, *Handbook of The Economics of Education*, Volume 2, pp. 1019-1047.
- SCHNEEWEIS, N. (2006), “How Should We Organize Schooling to Further Children with Migration Background?”, Department of Economics, Working Paper n°0620, Johannes Kepler University of Linz.
- SMITH, J. P. (2006), “Immigrants and Their Schooling”, *Handbook of The Economics of Education*, Volume 1, pp. 155-187.
- WINSBOROUGH, H. H. and P. DICKINSON (1971), “Components of Negro-White Income Differences”, *Proceedings of the Social Statistics Section: 6-8*.