

COMMENT MESURER LA SÉGRÉGATION DANS LE SYSTÈME ÉDUCATIF ?

Une étude de la composition sociale des collèges français

**Pauline Givord et
Marine Guillerm**

Insee, direction de la méthodologie et la coordination statistique et internationale

Olivier Monso

MENESR-DEPP, unité des méthodes et synthèses statistiques

Fabrice Murat

MENESR-DEPP, bureau des études sur les établissements et de l'éducation prioritaire

La mesure de la ségrégation consiste à quantifier un état de séparation de personnes appartenant à des groupes (sociaux, ethniques, etc.) différents, sur un territoire donné. Son opposé correspond à la mesure, plus positive, de la mixité sociale. Mesurer la ségrégation, en particulier celle qui existe entre les établissements scolaires, est un enjeu fort pour l'analyse et le pilotage du système éducatif. Cet article rappelle les principes généraux des indicateurs permettant de mesurer la ségrégation, et discute les propriétés des principaux indices habituellement utilisés. Nous nous appuyons sur l'un de ces indicateurs, l'indice d'entropie normalisé, pour proposer des éléments de diagnostic sur la ségrégation sociale entre les collèges français, à partir des données issues du système d'information Scolarité. Ces éléments permettent, dans un premier temps, de caractériser les territoires selon un degré plus ou moins fort de ségrégation entre les collèges : nous mettons en évidence une ségrégation plus forte dans les académies et dans les départements urbains. L'indice d'entropie normalisé, de par sa propriété de décomposabilité, permet de mettre en évidence des mécanismes de ségrégation différents selon les groupes sociaux et d'évaluer l'importance du secteur d'enseignement dans la ségrégation.

« R enforcer les actions contre les déterminismes sociaux et territoriaux » constitue l'une des « onze mesures pour une grande mobilisation de l'École pour les valeurs de la République » annoncées le 22 janvier 2015 par la ministre de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Cette huitième mesure promeut notamment « une politique active de mixité pour agir sur la composition des collèges ».

La « mixité » dont il est question est la « mixité sociale », qui désigne le degré avec lequel les collèges scolarisent, en leur sein, des élèves de milieux sociaux différents. L'exact opposé de cette notion est celle de « ségrégation sociale », qui désigne au contraire la tendance, pour les collèges, à accueillir des élèves d'un même groupe social. Faire progresser la mixité sociale, c'est donc favoriser une plus grande diversité des élèves accueillis dans chaque collège, en termes d'origine sociale¹.

La direction de l'évaluation, de la prospective et de la performance (DEPP) contribue à ce chantier de deux façons. Tout d'abord, elle aide à constituer un état des lieux sur la mixité sociale entre collèges : le niveau plus ou moins élevé qu'elle atteint en divers endroits du territoire, son évolution dans le temps, les différences entre collèges publics et privés. Ensuite, elle propose aux acteurs de terrain (conseils généraux, rectorats, notamment) des outils leur permettant d'affiner ce diagnostic sur leur territoire, en leur fournissant une aide à la décision, notamment en vue des réformes annoncées des secteurs de recrutement des collèges.

Cette interrogation sur la mixité sociale rejoint et renouvelle les travaux de la DEPP sur la diversité des collèges. Cette mesure de la diversité a déjà été approchée par différents outils et notamment par des typologies de collèges publics [THAUREL-RICHARD, 2005 ; THOMAS, 2005]. Ces typologies n'étaient pas à proprement parler des outils de mesure de la ségrégation sociale. En effet, les variables mobilisées comprenaient non seulement l'origine sociale des parents, mais aussi des variables évoquant d'autres formes de ségrégation (comme les résultats aux évaluations de sixième, la part d'étrangers, etc.) et pour certaines, relatives au fonctionnement des établissements (offre scolaire, taux d'accès de la sixième à la troisième). Elles visaient notamment à fournir aux académies des outils permettant de prendre en compte la diversité de leurs collèges, à des fins de pilotage et d'analyse. Une de ces typologies a, par exemple, été mobilisée dans une étude sur les demandes de dérogation parmi les collèges parisiens [GIRARD et GILLOTTE, 2005].

Plus récemment, cette analyse s'est recentrée sur l'hétérogénéité des collèges suivant l'origine sociale des élèves. L'examen de la catégorie socioprofessionnelle (par la suite nommée « catégorie sociale » pour simplifier) a été renouvelé : à la réflexion, récurrente, sur les modalités de recueil de ces variables [JASPAR, 2007], se sont ajoutées de nouvelles approches sur son utilisation. Les regroupements de catégories sociales traditionnellement utilisés par la DEPP (très favorisés, favorisés, moyens, défavorisés), ont été réinterrogés, pour répondre à certaines critiques dont ils faisaient l'objet, notamment sur les contours de la

1. La mixité sociale peut être distinguée d'autres formes de mixité, par exemple selon le sexe, ou encore de la mixité « scolaire », où c'est la cohabitation d'élèves de niveaux scolaires différents qui est recherchée. Une recherche récente [FACK, GRENET, BENHENDA, 2014 ; FACK et GRENET, ce numéro, p. 77] a souligné l'importance de distinguer la mixité sociale et la mixité scolaire, en montrant comment une politique éducative (en l'occurrence une modification des règles d'affectation des élèves dans les lycées parisiens) pouvait avoir des effets différenciés sur ces deux phénomènes.

catégorie « origine défavorisée² ». LE DONNÉ et ROCHER [2010] ont ainsi proposé un « indice de synergie socio-scolaire » pour placer les différentes professions sur une échelle quantitative. Cet indice vient de faire l'objet d'une refonte sous la forme d'un « indice de position sociale » [ROCHER, 2016]. La première version citée a déjà été utilisée pour étudier les écarts entre établissements [MURAT et THAUREL-RICHARD, 2013 ; MENESR-DEPP, 2014].

Quelles que soient les modalités d'utilisation de la catégorie sociale, l'examen des outils de mesure de la mixité sociale est une étape indispensable et fait l'objet de la présente étude, fruit d'une collaboration entre la DEPP et l'Insee. La mesure de la mixité se fait en principe au travers d'indicateurs dits « de ségrégation », qui caractérisent une absence de mixité. De nombreux indicateurs de ségrégation sont disponibles. De façon générale, ils visent à quantifier un état de séparation de personnes appartenant à des groupes (sociaux, ethniques, etc.) différents, sur un territoire donné. Tous n'ont pas les mêmes propriétés et ne sont pas adaptés au cadre de la ségrégation sociale dans le système éducatif. Outre certaines propriétés mathématiques, ces indicateurs doivent rester lisibles par les acteurs de terrain, qui peuvent ainsi se les approprier et les utiliser pour comprendre leur territoire. On cherche également à construire des indicateurs robustes pour des comparaisons dans le temps, entre les établissements et entre les académies, et décomposables selon des regroupements tels que le secteur de scolarisation pour quantifier son rôle dans la ségrégation sociale.

Cet article propose donc, dans une première partie, une revue des principaux indicateurs permettant de mesurer la ségrégation, en discutant de leur intérêt pour l'analyse et le pilotage du système éducatif. Cette introduction méthodologique permet d'identifier les propriétés importantes de ces indices de ségrégation, qui peuvent être calculés en particulier pour des catégories sociales, et de justifier le choix de l'indice d'entropie pour la suite de l'analyse. Celui-ci présente en effet le double intérêt d'avoir une version « multigroupe », c'est-à-dire de pouvoir appréhender la diversité sociale de manière plus fine qu'une opposition binaire entre deux groupes, et d'être simplement décomposable, par exemple selon le secteur de scolarisation (public ou privé).

Dans une seconde partie, l'indice d'entropie est utilisé pour mesurer le niveau de ségrégation dans les collèges français, à partir des données de la base Scolarité. Cette étude permet ainsi, dans un premier temps, de fournir un tableau du degré de mixité sociale au niveau des territoires (ici les académies et les départements), qui peut être mise en regard de la composition sociale de ces territoires. Des divergences importantes sur le niveau de ségrégation, pour des territoires dont les compositions sociales sont proches, sont des éléments qui peuvent appeler à une analyse plus précise des mécanismes à l'œuvre. Le caractère plutôt rural ou urbain par exemple peut jouer. Au-delà de ce constat, il faut aussi s'interroger sur l'évolution de cette ségrégation et des facteurs qui peuvent l'expliquer. Cette dernière partie de l'article propose donc une analyse de l'évolution du niveau de ségrégation des collèges français entre 2003 et 2015, en s'intéressant en particulier aux contributions respectives des secteurs privé et public dans cette évolution. Le cas particulier de Paris permet d'illustrer les mécanismes de ségrégation à l'œuvre entre les différentes catégories sociales.

2. La catégorie « origine défavorisée » de la DEPP comprend actuellement les parents ouvriers, les chômeurs et les inactifs n'ayant jamais travaillé ainsi que les retraités employés et ouvriers. Cette catégorie est elle-même très hétérogène, les enfants d'ouvriers qualifiés étant en général dans une situation plus favorable que les enfants des autres catégories. Inversement, les employés de services directs aux particuliers pourraient être considérés comme « défavorisés ».

MESURER LA SÉGRÉGATION, QUESTIONS ET MÉTHODES

Les indices de ségrégation, principes généraux

Si l'intérêt pour les phénomènes de ségrégation urbaine ou scolaire est plutôt récent en France, il s'agit d'un sujet d'étude déjà ancien dans la littérature sociologique anglo-saxonne. De très nombreux indices et mesures ont été proposés [pour une revue récente des indices de ségrégation entre établissements scolaires, voir par exemple FRANKEL et VOLIJ, 2011], reflet de la difficulté à appréhender par une mesure unique des phénomènes par nature complexes. De fait, il n'existe pas d'indice idéal et le choix de privilégier l'un ou l'autre résultera en général d'un compromis pondérant les différentes propriétés considérées comme « indispensables » en fonction de la problématique étudiée.

De manière concrète, ces indices partent tous du postulat qu'on peut distinguer différents groupes parmi les élèves (typiquement, selon l'origine sociale divisée en catégories) scolarisés dans différentes « unités », comme les collèges, les classes, etc. Le principe général des indices de ségrégation est de comparer la distribution des différents groupes dans chacune des unités avec la distribution de ces groupes dans la population au niveau global (par exemple ici, celle observée au niveau du département ou de l'académie). Cette distribution, qui dépend de la taille de chacun des groupes, sert de référence : elle serait celle qu'on s'attendrait à observer dans tous les établissements si les élèves se répartissaient de façon parfaitement homogène dans les différentes unités, c'est-à-dire s'il n'y avait aucun phénomène de ségrégation. Ce principe est le même que celui qui prévaut pour l'analyse des inégalités de revenu, et on ne s'étonnera donc pas de retrouver des similitudes entre certains des indicateurs de ségrégation et ceux utilisés pour étudier les inégalités de revenus.

Derrière ce postulat de départ, les choix pratiques ne vont pas toujours de soi, d'autant qu'ils peuvent avoir des conséquences sur le diagnostic que l'on pourra porter sur les phénomènes de ségrégation. La première question est celle du découpage, plus ou moins fin, des catégories d'élèves que l'on souhaite retenir. Les premiers indices de ségrégation proposés dans la littérature sociologique s'appuient sur un découpage en deux groupes. Cela s'explique en partie par le fait que ces indicateurs avaient d'abord été développés pour étudier l'ampleur de la ségrégation raciale aux États-Unis et il s'agissait donc essentiellement de distinguer « noirs » et « blancs ». Cette distinction binaire ne rend évidemment pas bien compte de la diversité de la société et des indices « multigroupes » ont depuis été développés. De fait, une segmentation en deux groupes va rarement de soi. Lorsqu'on utilise une catégorisation reposant sur l'origine sociale par exemple, on peut penser à distinguer les élèves issus de familles défavorisées *versus* tous les autres élèves. Il peut aussi y avoir un sens à distinguer plutôt ceux issus de milieux privilégiés, par comparaison avec l'ensemble des autres élèves. Ce choix n'est pas anodin, car mettre l'accent sur l'une ou l'autre des catégories peut conduire à des diagnostics différents, et il est donc nécessaire de croiser les points de vue. Utiliser un indice multigroupe, à l'inverse, permet une description plus fine et complète des groupes sociaux, sans particulariser une catégorie par rapport aux autres. Il s'agit de juger si chaque élève se trouve à l'école en présence de milieux sociaux diversifiés, lui ouvrant un large choix de possibles, sans privilégier un milieu social. En revanche, résumer par un chiffre unique une situation complexe est forcément réducteur. Des évolutions conséquentes pour une catégorie peuvent être masquées par un indicateur global. Si la concentration de cette catégorie sociale

dans les établissements est en soi un enjeu (par exemple si l'on pense que cela peut avoir des conséquences sur la réussite des élèves), cela réduit la pertinence du diagnostic global. Il est donc nécessaire de confronter les résultats obtenus par des choix différents et d'articuler les constats obtenus selon différents regroupements.

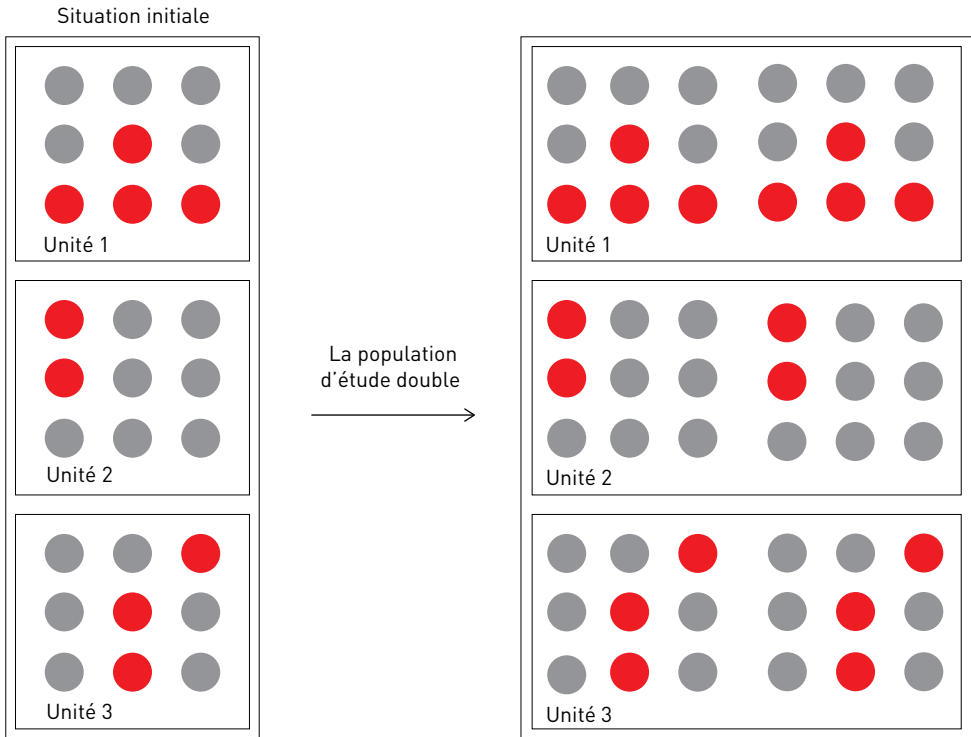
L'unité d'observation est une autre question importante. Elle peut être la classe, le collège, le département, etc. Si les choix sont *a priori* plus simples que dans le cas des catégories d'élèves, ils ne sont pas neutres non plus. En particulier, calculer des indicateurs de ségrégation lorsque la taille des unités est faible peut conduire à une image faussée des phénomènes à l'œuvre. Même en l'absence de toute ségrégation, il est plus probable d'observer des accumulations « locales » de certains groupes : contrairement à l'intuition, le hasard ne produit que rarement une répartition parfaite et homogène ! Pour prendre un exemple concret, si les classes étaient constituées de manière totalement aléatoire au sein des établissements, rien n'assurerait que les proportions de filles et de garçons, ou de redoublants soient identiques dans l'ensemble des classes [pour une discussion, voir LY, MAURIN, RIEGERT, 2014]. À titre d'illustration, on peut considérer le cas caricatural où chaque classe ne serait constituée que de deux élèves. Si on compte au total autant de filles que de garçons, on s'attend à observer en moyenne une situation dans laquelle seulement la moitié des unités sont constituées d'une fille et d'un garçon (un quart de deux filles et le dernier quart de deux garçons). Par ailleurs, une ségrégation nulle ne peut généralement pas être atteinte avec des unités de petite taille, car son calcul repose sur des nombres entiers : à titre d'exemple, avec des unités comptant dix individus, il est impossible d'observer une répartition totalement homogène d'une population représentant sur l'ensemble un tiers des effectifs. On a donc une augmentation mécanique des indices de ségrégation lorsque la taille des unités diminue. Des corrections ont été proposées pour résoudre ce problème en cas d'unités de petite taille [dans le cas de l'indice de dissimilarité, voir par exemple CARRINGTON et TROSKE, 1997, ou RATHÉLOT, 2012]. Dans cette étude, nous nous concentrons sur la mesure de la ségrégation au niveau des collèges, dont la taille est suffisamment élevée pour que ce problème soit négligeable.

Les principales propriétés souhaitables d'un indice de ségrégation

On peut définir les propriétés « minimales » qu'on attend d'un indice de ségrégation. Une propriété souhaitable est de dépendre de manière monotone du niveau de regroupement adopté pour définir les unités [« *school division property* » dans FRANKEL et VOLIJ, 2011] : si on divise une unité en plusieurs sous-unités (par exemple des écoles en classes), l'indice de ségrégation calculé en utilisant cette nouvelle subdivision ne peut diminuer. Il reste stable dans le cas particulier où chaque sous-unité a la même distribution des différents groupes que l'unité dont elle est issue, et augmente sinon.

D'autres propriétés importantes, en particulier lorsque l'on souhaite mener des comparaisons dans le temps ou entre académies, tiennent à la sensibilité de l'indice aux caractéristiques de la population étudiée (par exemple l'ensemble des élèves d'une académie). Une première propriété, minimale, est que l'indice de ségrégation retenu ne soit pas modifié lorsque la taille de la population globale varie, tant que la distribution des groupes sociaux reste la même au sein des unités (*invariance d'échelle*) ↘ **Figure 1**. Une autre propriété est liée à la composition de cette population globale étudiée. Comme décrit plus haut, le principe d'un indice de ségrégation est de comparer la répartition des différents groupes, telle qu'elle est observée dans les unités, avec la répartition de ces groupes sur l'ensemble de la population,

↘ **Figure 1** Illustration de la propriété d'invariance d'échelle

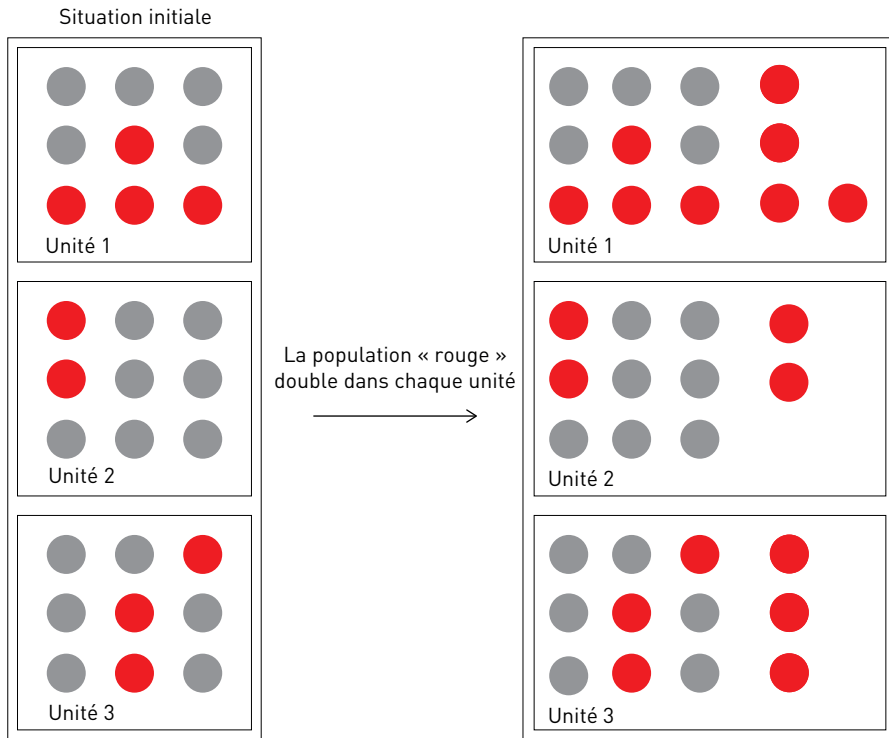


Lecture : la taille de la population totale double, mais sans changer les proportions d'individus « rouge » et « gris » dans chaque unité (et au niveau global).

prise comme référence. On souhaite donc que la valeur de l'indice ne varie pas en fonction de la distribution « de référence » des groupes (celle calculée sur l'ensemble de la population) : on parle d'*invariance à la composition*. Dans le cas contraire, le risque existe de considérer à tort que le degré de ségrégation est, par exemple, plus élevé dans une académie que dans une autre, simplement parce que les tailles respectives des groupes sont différentes. La définition de l'invariance à la composition repose sur la manière de définir la façon dont les effectifs doivent se répartir entre les différentes unités quand la distribution « de référence » évolue pour avoir une ségrégation constante.

La manière formelle la plus courante de définir cette propriété est de dire que lorsque la taille d'un groupe donné au sein de la population augmente, mais que la répartition entre les unités reste la même (le nombre des individus de ce groupe est multiplié par la même constante dans chaque unité), alors l'indice de ségrégation n'est pas modifié. Nous illustrons cette propriété par un exemple ↘ **Figure 2**. Le nombre d'individus « rouge » double dans l'ensemble des unités (donc sans modifier leur répartition entre ces unités). Cette définition est intuitive, mais, comme l'illustre aussi la **figure 2**, elle est également arbitraire : typiquement, dans cet « exercice » formel, la taille de chaque unité augmente de façon non proportionnelle. Une autre manière de définir l'invariance à la composition suppose que l'indice soit invariant

Figure 2 Illustration de la propriété d'invariance à la composition



Lecture : dans chaque unité, le nombre d'individus « rouge » double. Ceci ne modifie pas la distribution des individus « rouge » [ni celle des individus « gris »] à travers les différentes unités.

lorsque la proportion (et non le nombre) des individus d'un groupe dans chaque unité est multipliée par une constante (la taille de chaque unité restant constante).

Enfin, une propriété importante est que l'indice soit décomposable. Comme discuté plus haut, on peut souhaiter, selon l'angle choisi, utiliser un découpage plus ou moins fin des groupes sociaux. Cependant, les indices de ségrégation sont en général sensibles au choix effectué dans la segmentation de ces groupes et on risque donc, avec le même indice, d'aboutir à des diagnostics différents selon la catégorisation choisie, plus ou moins détaillée. Pouvoir décomposer un indice global dans ces différentes composantes correspondant à des segmentations plus fines est donc utile, car cela permet d'articuler les messages obtenus en isolant une catégorie particulière sur laquelle on souhaite mettre l'accent (par exemple les élèves issus de milieux défavorisés) et ceux obtenus en considérant une segmentation plus générale.

Dans le même ordre d'idée, on peut souhaiter décomposer l'indice selon un regroupement des unités, et mesurer ce qui relève de la ségrégation entre et à l'intérieur de ces groupes d'unités. Comme illustré dans cet article et l'article compagnon dans ce même numéro [GIVORD, GUILLERM *et alii*, p. 53], on peut, par exemple, évaluer l'importance de la ségrégation entre les collèges du public et du privé, et la ségrégation au sein de ces deux secteurs.

Les indices de ségrégation usuels, intérêts et limites

Il existe de très nombreux indices permettant de mesurer la ségrégation. Il ne s'agit pas ici d'en faire une revue exhaustive mais de présenter les principaux, en insistant sur les facteurs qui peuvent conduire à privilégier l'un ou l'autre de ces indices en fonction de l'usage qui en est attendu. Afin d'illustrer pratiquement ces indicateurs, nous utilisons une segmentation des élèves des collèges en quatre groupes définis par leur origine sociale : très favorisée (notée a), favorisée (b), moyenne (c) et défavorisée³ (d). Formellement, on note P la distribution des différents groupes sociaux dans la population de référence $P = (q_a, q_b, q_c, q_d)$ où q_g correspond à la proportion du groupe g dans cette population de référence (par exemple, au niveau de l'académie ou du département). Dans la suite, les indicateurs sont présentés en supposant qu'on s'intéresse à la mesure de la ségrégation entre les collèges d'une académie. Les formules s'adaptent bien sûr à d'autres configurations (les collèges d'un département, les secteurs d'enseignement sur l'ensemble de la France, etc.).

L'un des indices les plus classiques pour mesurer la ségrégation est l'indice de dissimilarité proposé par DUNCAN et DUNCAN [1955]. Dans sa version la plus simple, avec deux groupes, cet indice est défini par :

$$D = \frac{1}{2} \sum_{k=1}^K |t_a^k - t_{\bar{a}}^k| \tag{1}$$

Il compare dans chaque collège k la part d'élèves du groupe a qui sont scolarisés dans ce collège (t_a^k) et la part d'élèves d'un autre groupe, souvent le groupe complémentaire noté \bar{a} , qui sont scolarisés dans ce même collège ($t_{\bar{a}}^k$), l'idée étant que si les deux parts s'égalisent dans chaque unité, les individus du groupe a (et ceux du groupe \bar{a}) sont uniformément répartis et la ségrégation est nulle.

Cet indicateur a l'avantage d'être très directement interprétable : il correspond au pourcentage d'élèves du groupe a dans l'académie qui doivent changer de collège pour atteindre une ségrégation nulle, c'est-à-dire une situation dans laquelle chaque collège accueille la même proportion d'élèves de milieu a . Dans l'exemple présenté dans la **figure 1** (situation initiale à gauche), si la distribution des individus « rouge » sert de référence, les unités devraient compter respectivement 4/9, 2/9 et 3/9 d'individus « gris ». Pour cela, trois individus « gris » de l'unité 2 doivent déménager dans l'unité 1. L'indice de dissimilarité vaut donc 3/18 \approx 0,17⁴. Notons que cet indice peut aussi se réécrire comme :

$$D = \frac{1}{2} \times \frac{\sum_{k=1}^K (\pi_k \times |q_a^k - q_{\bar{a}}^k|)}{q_a \times (1 - q_a)} \tag{2}$$

3. Dans les travaux de la DEPP, les professions dites « très favorisées » comprennent les cadres et professions intellectuelles supérieures, professions libérales et chefs d'entreprise de dix salariés ou plus, ainsi que les enseignants. Les catégories « favorisées » correspondent aux professions intermédiaires hors enseignants, ainsi qu'aux retraités des cadres et des professions intermédiaires. Les catégories « moyennes » incluent les agriculteurs, les artisans, les commerçants (et les retraités correspondants) ainsi que les employés. Enfin, les catégories « défavorisées » comprennent les ouvriers, les chômeurs et inactifs n'ayant jamais travaillé, ainsi que les retraités employés et ouvriers.

4. Ce calcul aboutit au même résultat si la distribution de référence choisie est celle des « gris ». Dans ce cas, c'est 1,5 individu « rouge » de l'unité 1 qui doit déménager dans l'unité 2, soit un indice de dissimilarité de 1,5/9 \approx 0,17.

où π_k correspond à la proportion d'élèves de l'académie scolarisés dans le collège k , q_a est la proportion d'élèves du groupe a dans l'académie et q_a^k cette même proportion dans le collège k .

On peut le généraliser à plusieurs groupes, en suivant par exemple MORGAN [1975] :

$$D = \frac{1}{2I_{Simpson}} \sum_{k=1}^K \pi_k [|q_a^k - q_a| + |q_b^k - q_b| + |q_c^k - q_c| + |q_d^k - q_d|] \quad (3)$$

où $p_k = (q_a^k, q_b^k, q_c^k, q_d^k)$ correspond à la distribution des groupes dans le collège k et $I_{Simpson}$ est l'indice d'interaction de Simpson. Ce dernier correspond à la probabilité que deux élèves tirés aléatoirement ne soient pas du même milieu social :

$$I_{Simpson} = q_a(1 - q_a) + q_b(1 - q_b) + q_c(1 - q_c) + q_d(1 - q_d) \quad (4)$$

Un autre indice classique est l'indice d'exposition normalisé qui mesure la probabilité de contact d'un groupe (noté a ici) avec des individus d'autres groupes sociaux. Il s'écrit :

$$S = \frac{1}{q_a(1 - q_a)} \sum_{k=1}^K \pi_k (q_a^k - q_a)^2 \quad (5)$$

Comme souligné par LY, MAURIN et RIEGERT [2014], il peut s'interpréter comme la part de la variance de la variable d'appartenance au groupe a expliquée par les unités géographiques (ou part de la variance inter). Cette approche est particulièrement pertinente lorsqu'on cherche à définir la probabilité de contact d'un élève aux élèves venant d'un milieu favorisé (ou aux meilleurs élèves scolairement parlant), en supposant que cette interaction est *a priori* bénéfique. L'indice d'exposition normalisé correspond alors à la façon dont cette exposition varie selon que l'élève est lui-même issu ou non d'un milieu favorisé. Notons qu'il existe également des versions multigroupes de cet indice, qui consistent en une moyenne pondérée des différents indices binaires obtenus en traitant chaque catégorie séparément au regard de toutes les autres. Néanmoins, l'intérêt de cette version multigroupe est peu évident, dans la mesure où l'on perd ce qui fait le premier intérêt de cet indice, sa simplicité d'interprétation et le lien avec un indicateur connu, la variance.

De par leur relative simplicité, ces indicateurs ont souvent été utilisés dans les travaux sur la ségrégation, mais ils ne respectent pas toutes les propriétés souhaitables (voir section précédente) : l'indice de dissimilarité respecte le principe d'invariance à la composition dans sa version binaire, mais pas dans sa version multigroupe ; de plus, il n'est pas décomposable. L'indice d'exposition normalisé est décomposable dans sa version binaire, mais ne respecte pas le principe d'invariance à la composition et sa généralisation à l'étude de plusieurs groupes prête à discussion.

Nous privilégierons donc dans la suite l'indice d'entropie, qui possède de meilleures propriétés, ce qui explique, malgré une certaine complexité, le regain d'intérêt récent dont il fait l'objet dans la littérature sur la ségrégation scolaire. Il a par exemple été utilisé par FACK, GRENET et BENHENDA [2014] dans une étude sur la ségrégation dans les lycées en Île-de-France. Cet indice, dérivé de celui proposé par THEIL dans le cadre de travaux sur les inégalités de revenu, est inspiré des travaux de SHANNON sur la théorie de l'information ↪ [Annexe 1](#) p. 44.

Plus précisément, l'entropie correspondant à la distribution des élèves au sein de l'académie est par définition :

$$h(P) = q_a \times \ln\left(\frac{1}{q_a}\right) + q_b \times \ln\left(\frac{1}{q_b}\right) + q_c \times \ln\left(\frac{1}{q_c}\right) + q_d \times \ln\left(\frac{1}{q_d}\right) \quad (6)$$

On peut simplement vérifier que si elle est toujours positive, elle vaut zéro lorsqu'un seul groupe est représenté, et est maximale lorsque les groupes sont en égale proportion. L'entropie $h(P)$ peut donc être considérée comme une mesure synthétique de la diversité sociale de la population des élèves de l'académie.

Le principe de l'indice d'entropie normalisé H (l'adjectif « normalisé », dont nous verrons l'importance plus loin, ne sera pas systématiquement repris dans la suite de cet article, conformément à l'usage dans la littérature) est alors de comparer cette valeur avec les valeurs des entropies observées au niveau de chaque collège. Plus précisément, il correspond à :

$$H = \sum_{k=1}^K \pi_k \frac{h(P) - h(p_k)}{h(P)} \quad (7)$$

avec $h(p_k) = q_a^k \times \ln\left(\frac{1}{q_a^k}\right) + q_b^k \times \ln\left(\frac{1}{q_b^k}\right) + q_c^k \times \ln\left(\frac{1}{q_c^k}\right) + q_d^k \times \ln\left(\frac{1}{q_d^k}\right)$ (8)

L'indice d'entropie correspond donc à une moyenne pondérée (par la taille de chaque collège) des écarts entre la diversité sociale des collèges et celle de la population de référence, normalisés par cette dernière. Il est compris entre 0 (aucune ségrégation, la distribution des groupes sociaux dans chaque collège correspond à celle de l'académie) et 1 (ségrégation maximale, qu'on observerait si les groupes n'étaient jamais mélangés dans les collèges). La **figure 3** présente une illustration dans un cas très simple avec deux groupes uniquement (le groupe « rouge » rassemblant un tiers de l'effectif total), répartis dans trois unités de taille identique. La diversité de la population globale, soit l'ensemble des élèves des trois unités, vaut dans chacune des situations présentées $h(P) = 1 / 3 \ln(3) + 2 / 3 \ln(3 / 2) \approx 0,64$. Dans la première situation, toutes les unités reproduisent cette proportion : l'indice d'entropie est nul. Dans la dernière, tous les points rouges sont concentrés dans une seule unité : les entropies de chaque unité sont nulles et l'indice total vaut 1. Dans la situation 2, intermédiaire, deux unités s'éloignent légèrement de cette distribution de référence et l'indice d'entropie vaut :

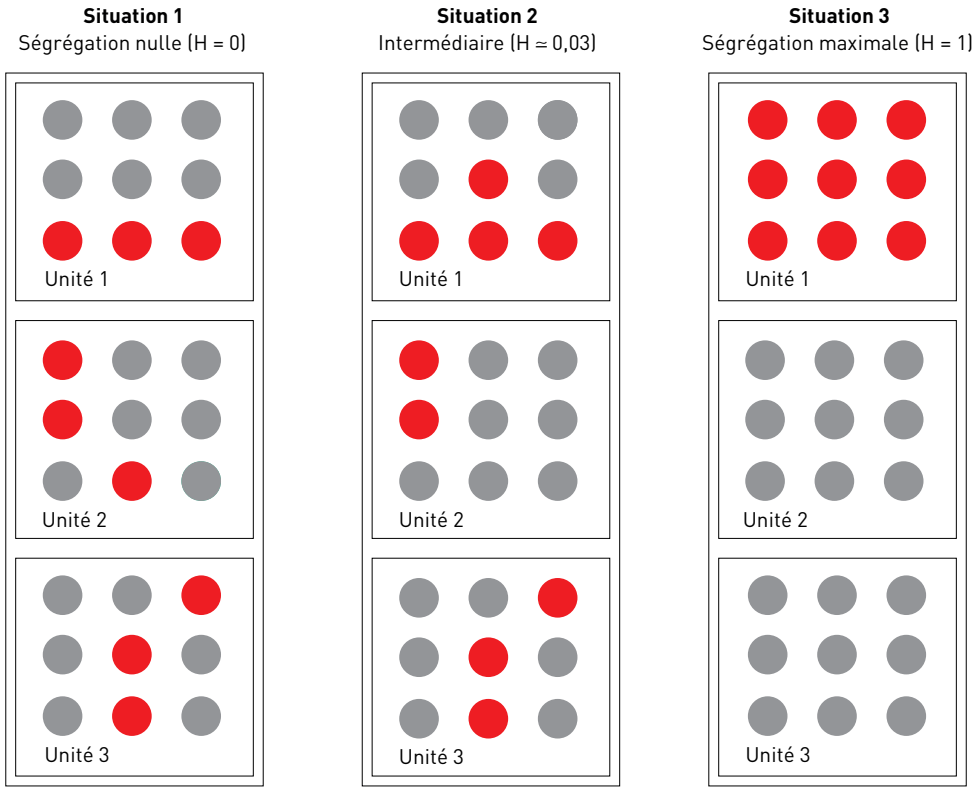
$$H = \frac{1(0,64 - 0,69) + (0,64 - 0,53) + (0,64 - 0,64)}{3 \times 0,64} \approx 0,03.$$

L'indice d'entropie multigroupe est décomposable de sorte à mettre en évidence des mécanismes de ségrégation spécifiques à une catégorie, en évaluant l'importance du niveau de ségrégation de ce groupe dans l'indice global de ségrégation. On peut en effet montrer que l'indice d'entropie multigroupe s'écrit comme une somme pondérée :

$$H = \theta_{a/(b,c,d)} \times H_{a/(b,c,d)} + \theta_{(b,c,d)}^{(-a)} \times H_{(b,c,d)}^{(-a)} \quad (9)$$

où $H_{a/(b,c,d)}$ correspond à l'indice de ségrégation binaire où on considère la catégorie d'élèves issus du groupe a , versus tous ceux qui n'y appartiennent pas, tandis que $H_{(b,c,d)}^{(-a)}$ correspond à l'indice de ségrégation multigroupe où on ne retient que les catégories b , c et d (les élèves du groupe a sont exclus du calcul).

Figure 3 Exemples de différentes situations et des indices d'entropie correspondants



Les termes $\theta_{a/(b,c,d)}$ et $\theta_{(b,c,d)}^{(-a)}$ correspondent à des termes de pondérations, respectivement :

$$\theta_{a/(b,c,d)} = \frac{h(p_{a/(b,c,d)})}{h(P)} \quad \text{et} \quad \theta_{(b,c,d)}^{(-a)} = \pi_{(b,c,d)} \times \frac{h(p_{b,c,d}^{(-a)})}{h(P)} \quad (10)$$

où $h(P)$ est l'entropie calculée comme ci-dessus sur les quatre catégories sociales, $h(p_{a/(b,c,d)})$ est l'entropie définie en considérant deux catégories, les élèves de milieu défavorisé et tous les autres, $h(p_{b,c,d}^{(-a)})$ est l'entropie calculée sur trois catégories sociales parmi les quatre de départ (les élèves de milieu défavorisé étant exclus du calcul), enfin $\pi_{(b,c,d)}$ est la part des élèves qui ne sont pas de milieu défavorisé au niveau global.

L'indice d'entropie est également décomposable selon des regroupements effectués pour les unités. Il peut, par exemple, être utilisé pour préciser le rôle de chacun des secteurs de scolarisation dans la ségrégation. On peut montrer par exemple que :

$$H = H_{PU/PR} + \theta_{PU} \times H_{PU} + \theta_{PR} \times H_{PR} \quad (11)$$

Cette décomposition fait apparaître trois composantes. Les deux composantes H_{PU} et H_{PR} rendent compte respectivement de la ségrégation entre collèges publics et entre collèges privés. On parle de ségrégation intra. La composante $H_{PU/PR}$, dite ségrégation inter, mesure la ségrégation entre les deux secteurs de scolarisation. Elle est ainsi d'autant plus élevée que les distributions des catégories sociales sont différentes d'un secteur à l'autre. Il s'agit donc de la part de l'indice d'entropie mesuré au niveau général qui est imputable à la segmentation sociale entre les collèges publics (pris dans leur ensemble) et les collèges privés. Concrètement, H_{PU} correspond à un indice d'entropie calculé sur les seuls collèges publics, soit

$$H_{PU} = \sum_{k, \text{collège public}} \pi_k^{PU} \frac{h(P_{PU}) - h(p_k)}{h(P_{PU})} \quad (12)$$

avec π_k^{PU} la proportion d'élèves du public qui sont scolarisés dans le collège k et P_{PU} la distribution des groupes sociaux dans le secteur public. H_{PR} est calculé de manière analogue sur les seuls collèges privés. $H_{PU/PR}$ est obtenu en considérant seulement deux unités dans l'équation (7) (au lieu des k collèges) — le secteur public et le secteur privé —, et en comparant la distribution des groupes sociaux dans ces deux secteurs avec la distribution au niveau global, soit

$$H_{PU/PR} = \pi_{PU} \frac{h(P) - h(P_{PU})}{h(P)} + \pi_{PR} \frac{h(P) - h(P_{PR})}{h(P)} \quad (13)$$

où π_{PU} est la proportion d'élèves scolarisés dans le public et π_{PR} la proportion d'élèves scolarisés dans le privé.

Les termes de pondérations θ_{PU} et θ_{PR} correspondent à :

$$\theta_{PU} = \pi_{PU} \times \frac{h(p_{PU})}{h(P)} \quad \text{et} \quad \theta_{PR} = \pi_{PR} \times \frac{h(p_{PR})}{h(P)} \quad (14)$$

où $h(p_{PU})$ est l'entropie globale calculée en se limitant aux élèves scolarisés dans les collèges publics et $h(p_{PR})$ son équivalent pour les collèges privés.

Cependant, il faut noter que ces décompositions ne sont pas parfaites. Par exemple, dans la comparaison du public et du privé, les pondérations utilisées ne reflètent pas uniquement les poids de chacune des deux populations (élèves scolarisés dans le public ou dans le privé). Le terme $H_{PU/PR}$ ne quantifie pas exactement de combien serait réduit l'indice s'il n'existait plus d'écart de composition sociale entre le public et le privé. Si cette égalisation se produisait, les modifications de composition qu'elles impliqueraient modifieraient également les termes $h(p_{PR})$ et $h(p_{PU})$ qui interviennent dans les pondérations des deux autres termes. Pour cette raison, on peut préférer utiliser l'indice d'information mutuelle M [voir FRANKEL et VOLJU, 2011] qui est simplement l'indice d'entropie *non normalisé* :

$$M = h(P) - \sum_{k=1}^K \pi_k h(p_k) \quad (15)$$

Cet indice vérifie la propriété de décomposition *forte*, c'est-à-dire que si l'on considère des groupements des unités, l'indice au niveau global se décompose comme la somme pondérée des indices intra à ces groupements et de l'indice inter. Dans le cas public/privé, on a :

$$M = M_{PU/PR} + \pi_{PU} \times M_{PU} + \pi_{PR} \times M_{PR} \quad (16)$$

Lorsqu'on souhaite singulariser un groupe par rapport aux autres dans la segmentation, on a également :

$$M = M_{a/(b,c,d)} + \pi_{(b,c,d)} \times M_{(b,c,d)}^{(-a)} \quad (17)$$

où $M_{a/(b,c,d)}$ correspond à l'indice d'information mutuelle calculé en ne considérant qu'une classification simple (élèves du groupe a en comparaison avec l'ensemble des autres élèves), $M_{(b,c,d)}^{(-a)}$ cet indice calculé en considérant simplement la population des trois autres groupes, dont la proportion dans la population est $\pi_{(b,c,d)}$.

Cependant, cet indice n'est pas compris entre 0 et 1. L'absence de normalisation peut de fait le rendre plus sensible que l'indice d'entropie normalisé à des modifications de composition (la diversité mesurée au niveau global), ce qui peut rendre les comparaisons plus difficiles.

Même si l'indice d'entropie ne respecte pas parfaitement le principe d'invariance à la composition, en pratique, selon REARDON et FIREBAUGH [2002], il ne serait pas trop sensible à des variations de composition. Il sera toutefois nécessaire de vérifier que les variations constatées entre académies, ou dans le temps, ne sont pas liées en partie à des différences de composition. Une première approche serait de se limiter à des comparaisons entre des populations dont la composition est « proche ». De manière plus systématique, il est important de vérifier que les messages obtenus sont robustes au choix de l'un ou l'autre des indicateurs. On trouvera en **annexe 2** (p. 46) les résultats obtenus par département et académie en utilisant l'indice d'entropie, l'indice de dissimilarité « généralisé » et l'indice d'information mutuelle⁵. Nous avons aussi ajouté à la comparaison l'indice d'Atkinson, dérivé des travaux de ce dernier dans le domaine des inégalités de revenus, qui dans sa version la plus simple⁶, est défini par :

$$A = 1 - \sum_{k=1}^K \left(t_a^k t_b^k t_c^k t_d^k \right)^{1/4} \quad (18)$$

où t_g^k correspond à la proportion des élèves du groupe g qui sont scolarisés dans le collège k . Cet indicateur a notamment comme intérêt de respecter l'invariance à la composition telle que nous l'avons définie plus haut.

Il s'avère sur nos données que la plupart des indices sont très corrélés entre eux (entre 0,97 et 0,99), mais, et c'est normal, des différences apparaissent également sur certains cas. Cela permet de souligner, s'il était nécessaire, la relativité de tout classement. Partant de ce constat, nous retenons donc pour l'analyse qui suit l'indice d'entropie normalisé, dont les propriétés nous semblent les plus adéquates, à la fois, au vu de la variable utilisée (catégorie sociale à quatre modalités) et de la possibilité de pouvoir décomposer cet indice selon plusieurs dimensions (selon le secteur public et privé, comme dans l'article compagnon de ce numéro, ou selon différents regroupements de catégories sociales).

5. L'indice d'exposition normalisé n'a pas été retenu, car il est surtout pertinent pour les variables binaires.

6. C'est-à-dire symétrique : on peut également le définir en utilisant une version non symétrique, en pondérant différemment les groupes.

UNE APPLICATION DE L'INDICE D'ENTROPIE À L'ÉTUDE DE LA SÉGRÉGATION DANS LES COLLÈGES FRANÇAIS

Comparer la ségrégation entre territoires : des disparités entre collèges plus fortes dans les départements urbains

Afin d'illustrer l'apport d'un indice de ségrégation, et notamment de l'indice d'entropie, aux constats et analyses sur la ségrégation dans le système éducatif, nous proposons une application aux écarts de composition sociale entre les collèges français, à partir des données des bases Scolarité ↘ **Encadré 1**. Le milieu social est défini à partir de la catégorie sociale du « parent de référence » de l'élève (le père si sa profession est renseignée, la mère sinon) et en utilisant les quatre regroupements usuels des catégories sociales de la DEPP (très favorisées, favorisées, moyennes, défavorisées⁷). Mesurer la ségrégation entre collèges, sur un territoire et en termes de milieu social des élèves, a pour premier atout celui d'apporter un élément important de description de ce territoire et ainsi d'aller au-delà du profil « moyen » constitué par les parts moyennes des différentes catégories sociales. Les indices de ségrégation, rappelons-le, permettent d'établir dans quelle mesure les collèges se distinguent les uns des autres en termes de composition sociale. Pour des départements et des académies présentant un profil moyen similaire, les disparités entre collèges peuvent différer fortement.

C'est ce que nous illustrons ci-dessous à partir de l'indice d'entropie, calculé en utilisant les quatre catégories sociales présentées en première partie de cet article. Les départements des Vosges (académie de Nancy-Metz) et de la Somme (académie d'Amiens) ont une tonalité sociale « moyenne » similaire : les parts des différentes catégories sociales parmi les collégiens y sont proches, la part des collégiens défavorisés avoisinant 50 %⁸ ↘ **Tableau 1**. Fort logiquement, l'entropie, mesurant la diversité des catégories sociales présentes sur l'ensemble du territoire, est proche⁹. C'est lorsqu'on considère les disparités entre les collèges (indice d'entropie) qu'une différence notable apparaît : les collèges de la Somme sont beaucoup plus segmentés que ceux des Vosges, l'indice d'entropie y est plus de deux fois supérieur¹⁰.

De façon similaire, à l'échelle académique, les académies de Grenoble et de Lyon ont un profil social « moyen » similaire et plutôt favorisé. Ces deux académies présentent par ailleurs des

7. Pour une présentation et une discussion du contenu de ces catégories, voir ROCHER [2016]. Dans notre étude, la catégorie « origine sociale non renseignée » est par ailleurs agrégée avec l'« origine défavorisée ». En effet, les résultats scolaires des enfants dont l'origine sociale est non renseignée sont en général proches des élèves de catégorie « défavorisée ». Leur assimilation aux élèves d'origine défavorisée est donc raisonnable, sauf dans le cas de collèges à forts taux de non-réponse sur l'origine sociale. C'est pourquoi les collèges dont plus de 25 % des élèves n'ont pas de catégorie sociale renseignée ont été exclus du champ.

8. Notons que la proximité entre ces deux départements peut être attestée à partir des valeurs moyennes prises par d'autres variables caractérisant l'environnement socio-économique des enfants, et publiées dans *Géographie de l'École* [MENESR-DEPP, 2014] : part des enfants de 0 à 17 ans en situation de pauvreté, dont les parents sont au chômage.

9. L'inverse n'est pas vrai : deux territoires peuvent avoir une entropie égale, mais des structures sociales différentes. C'est par exemple le cas d'un territoire A où la répartition entre les quatre groupes parmi les collégiens (en partant des « très favorisés » vers les « défavorisés ») serait de 70, 10, 10, 10 %, et un territoire B où elle serait de 10, 10, 10, 70 %. Le calcul de l'indice d'entropie ne tient pas compte de la hiérarchie des groupes : cela peut constituer une limite (partagée par la plupart des indices de ségrégation, qui sont souvent symétriques), mais peut être aussi un choix.

10. Rappelons que la valeur de l'indice n'a pas de sens en elle-même. Si comme la plupart des indices de ségrégation, l'indice d'entropie varie de 0 à 1, il prend rarement des valeurs très élevées. Cela peut s'expliquer par sa forme fonctionnelle, qui le rend très sensible à des valeurs extrêmes – de très petites proportions de telles ou telles catégories – qui se retrouvent rarement dans les collèges français. À titre d'illustration, les valeurs départementales de l'indice vont de 0,017 à 0,147.

Encadré 1

LES BASES SCOLARITÉ

Depuis 2004, la base Scolarité fournit des informations détaillées sur l'ensemble des élèves du second degré, relevant du ministère de l'Éducation nationale, pour les établissements publics et privés, en France métropolitaine et dans les départements d'outre-mer. Les données concernant les établissements publics et privés sous contrat sont issues des systèmes de gestion (Sicle) compilés en bases statistiques (Bases élèves académiques). Pour l'enseignement privé hors contrat, une enquête spécifique est menée (Scolege), mais les données sur la PCS étant moins fiables, il a été exclu du champ de cette

étude. Les données sont remplies par les services administratifs des établissements et contrôlées par les services en académies. Les PCS des parents sont donc connues par des codes dans une nomenclature en 32 postes, renseignés à partir des libellés en clair donnés par les parents. Les autres informations disponibles sont notamment, concernant l'élève, le sexe, la date de naissance, le statut de boursier et concernant sa scolarité, l'établissement fréquenté, la formation suivie, la classe fréquentée, les options suivies et ces mêmes informations pour l'année précédente. Au niveau national, ce fichier est anonyme.

Tableau 1 Exemples de départements ou d'académies au profil social « moyen » similaire mais présentant une ségrégation plus ou moins forte de leurs collèges, rentrée 2015

		Part de collégiens de catégorie... [en %]					Entropie h(P)	Indice d'entropie H (ensemble des collèges)
		Nombre de collèges	Très favorisée	Favorisée	Moyenne	Défavorisée		
Départements	Vosges	48	14,8	13,1	24,6	47,6	1,247	0,038
	Somme	67	15,6	11,0	24,7	48,7	1,228	0,087
Académies	Grenoble	323	24,7	14,5	27,1	33,6	1,346	0,054
	Lyon	302	27,0	13,6	23,8	35,6	1,334	0,092
France hors Mayotte		6 890	22,7	12,5	26,9	37,8	1,318	0,097

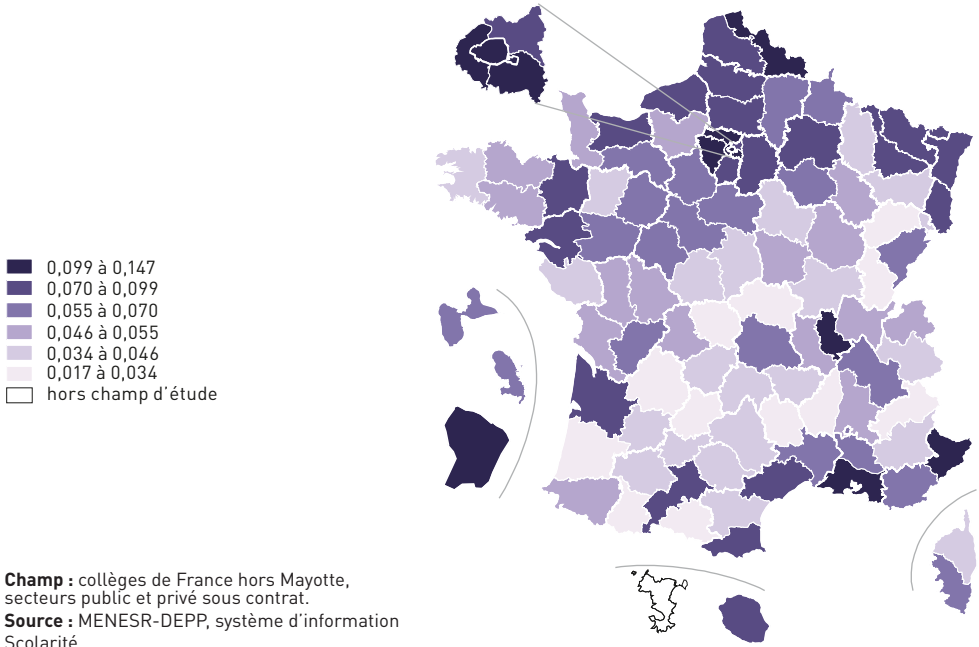
Champ : collèges de France hors Mayotte, secteurs public et privé sous contrat.

Source : MENESR-DEPP, système d'information Scolarité.

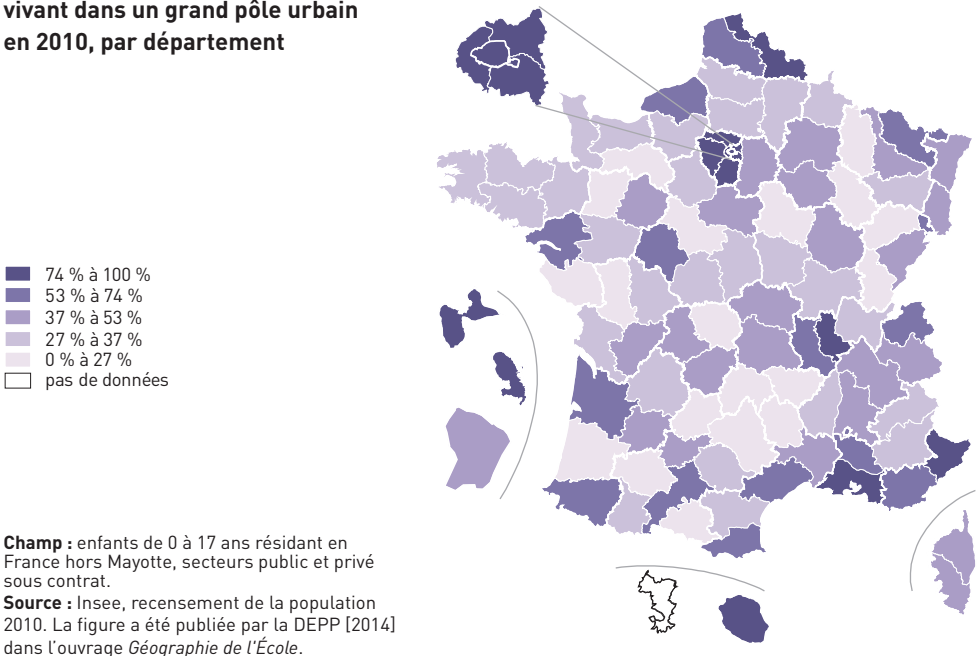
disparités internes assez fortes de contexte socio-économique [MENESR-DEPP, 2014] mais, pour ce qui concerne la ségrégation entre les collèges, c'est dans l'académie de Lyon qu'elle est, de loin, la plus forte.

Dans cet exemple, l'indice d'entropie permet d'enrichir la comparaison portant sur deux territoires dont la composition sociale moyenne est similaire. Il doit aussi, sous certaines précautions déjà évoquées, permettre une telle comparaison sur des territoires dont la composition sociale moyenne diffère : c'est précisément l'apport d'un indice de ségrégation, au regard d'un « simple » indicateur de dispersion, comme la variance ou l'écart-type. Le calcul de l'indice par département suggère un lien fort entre ségrégation sociale et degré d'urbanisation. Les départements où la ségrégation entre collèges est la plus forte sont aussi, en général,

📄 **Figure 4** Indice d'entropie à la rentrée 2015



📄 **Figure 5** Part d'enfants de 0 à 17 ans vivant dans un grand pôle urbain en 2010, par département



les plus urbains ↘ **Figure 4 et Figure 5**. En métropole, les valeurs les plus élevées sont atteintes à Paris, dans les Hauts-de-Seine, les Yvelines, le Val-de-Marne, le Val-d'Oise, le Rhône, les Bouches-du-Rhône, le Nord et les Alpes-Maritimes. De l'autre côté, c'est dans les Hautes-Alpes, la Lozère et la Creuse que les collèges se ressemblent le plus. Le coefficient de corrélation entre l'indice d'entropie et la part d'enfants de 0 à 17 ans vivant dans un grand pôle urbain est de 0,78. Au sein d'une académie, c'est souvent dans le département le plus urbain que la ségrégation est la plus forte : par exemple dans la Marne (académie de Reims), en Gironde (académie de Bordeaux), ou encore dans l'Hérault (académie de Montpellier).

Ce constat rejoint celui déjà fait dans les travaux sur les typologies de collèges publics, constatant des disparités très tranchées entre collèges dans les grandes agglomérations [THOMAS, 2005]. Il a aussi été illustré dans une étude récente menée à partir de l'indice d'exposition normalisé [LY et RIEGERT, 2015]. Deux types d'interprétation, au moins, peuvent être avancés. La première interprétation tient à une ségrégation résidentielle plus forte dans les grandes villes. L'économie géographique a proposé un cadre théorique, dont nous allons donner un aperçu, permettant d'expliquer pourquoi les regroupements de population suivant le statut socio-économique sont plus marqués dans les grandes villes qu'ailleurs. La pression foncière y engendre de fortes variations des prix du logement entre le centre-ville et la périphérie. Les ménages se répartissent entre les différents quartiers et communes d'une façon qui est fortement associée à leur statut socio-économique, car ce dernier détermine leur capacité à payer leurs dépenses de logement. CHARLOT, HILAL et SCHMITT [2009] présentent de façon plus détaillée ce cadre théorique et ses implications. Ils mettent aussi en évidence que, conformément à ce modèle, la ségrégation est d'autant plus forte que les aires urbaines sont de taille importante¹¹. Ces explications n'excluent nullement d'autres facteurs, historiques et culturels, ayant engendré, dans un quartier donné, des regroupements de populations aux caractéristiques similaires. Cette ségrégation résidentielle prend aussi des formes différentes d'une aire urbaine à une autre, comme l'a montré DASRÉ [2012]. Quelles que soient ses modalités, elle agit comme une contrainte forte sur le recrutement des collèges, par l'intermédiaire de leurs secteurs de recrutement. La ségrégation entre collèges en termes de milieu social traduit en bonne partie la ségrégation résidentielle.

La seconde interprétation tient à des stratégies de choix de collège par les parents, qui accentuent les effets de la ségrégation résidentielle. L'existence de ces stratégies d'évitement a été mise en évidence par les travaux des sociologues, géographes et économistes [POUPEAU et FRANÇOIS, 2008 ; VAN ZANTEN, 2009 ; DAVEZIES et GARROUSTE, 2014]. Ces travaux ont aussi montré que les pratiques d'évitement sont d'abord le fait des catégories sociales les plus aisées. Enfin, elles sont sans doute exacerbées en milieu urbain, où l'offre scolaire est plus dense. Les possibilités d'« éviter » le collège de secteur, par une dérogation pour accéder à un autre collège public ou par une inscription dans un collège privé, y sont plus étendues. Il est donc probable que les stratégies d'évitement accentuent encore le lien entre ségrégation des collèges et degré d'urbanisation.

Les liens entre ségrégation résidentielle et ségrégation des collèges restent toutefois complexes. Nous venons de proposer une vision « séquentielle », où la seconde reproduit et renforce la première. Un mécanisme inverse est sans doute aussi à l'œuvre, dans la mesure où

11. En particulier, « avec une plus forte densité, la pression foncière devient telle qu'elle pousse les familles à revenu moyen, cherchant à devenir propriétaires, à s'éloigner du centre, contribuant ainsi à un tri spatial des différentes catégories sociales » [CHARLOT, HILAL, SCHMITT, 2009, p. 33].

les choix de résidence sont influencés par la réputation des collèges avoisinants, et donc leur composition sociale. FACK et GRENET [2009] ont ainsi mis en évidence la sensibilité des prix des logements parisiens aux résultats obtenus par les collèges au diplôme national du brevet. Notre objet n'est pas de départager et quantifier ces différents canaux, mais plutôt d'en retenir le constat vers lequel tous semblent converger : la ségrégation sociale entre collèges est plus forte en milieu urbain.

Décomposer l'évolution de la ségrégation suivant le secteur : une ségrégation stable depuis 2003 mais une baisse dans les collèges publics

La ségrégation entre collèges a-t-elle augmenté au cours des années récentes ? À notre connaissance, peu de travaux permettent de se faire une idée sur le sujet. TRANCART [2012] propose une telle analyse pour la période 1993-2008, sur les collèges publics de France métropolitaine. Pour ce qui est des écarts de composition sociale, elle met en avant une hausse de la ségrégation au cours du temps. Toutefois, elle illustre aussi la façon dont ce constat est sensible aux catégories définies pour l'analyse : en considérant la ségrégation entre les élèves d'origine sociale défavorisée et les autres, les indices utilisés (incluant l'indice de dissimilarité et le coefficient de Gini) sont stables au cours du temps. C'est en isolant les « très défavorisés » que la ségrégation est croissante, c'est-à-dire en excluant les élèves dont le responsable est ouvrier, pour ne garder que ceux dont le responsable est chômeur ou inactif n'ayant jamais travaillé.

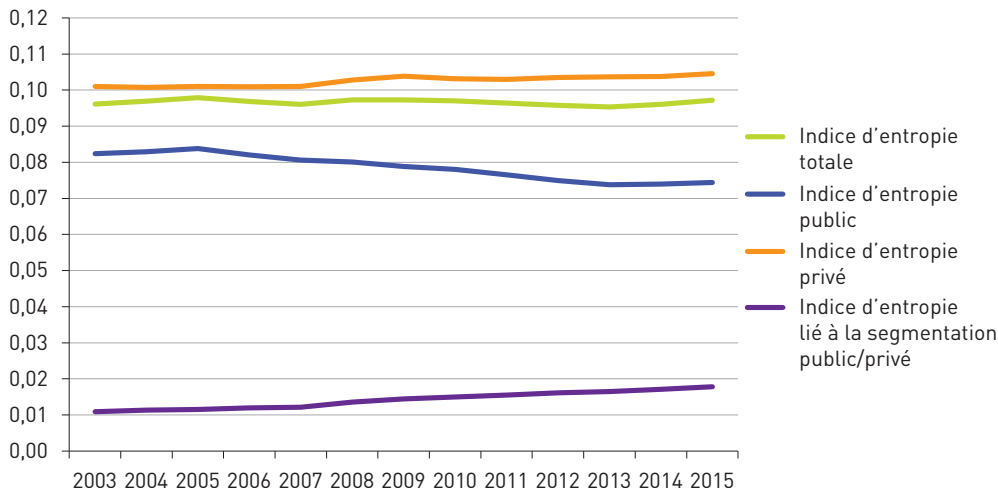
Sur une période plus récente (2004-2011), et prenant en compte les collèges privés, THAUREL-RICHARD et MURAT [2013] constatent une très légère baisse des disparités entre collèges, en se référant à l'évolution de l'écart-type de la part d'élèves d'origine défavorisée dans les collèges¹². La baisse est toutefois plus prononcée au sein de chaque secteur, public et privé. Ils interprètent cette divergence par le fait que, dans le même temps, les écarts de composition sociale entre les deux secteurs s'accroissent : la composition du privé est de plus en plus favorisée, en particulier la part de collégiens d'origine défavorisée a diminué rapidement entre 2004 et 2011 dans le secteur privé, alors qu'elle est restée stable dans le public. Toutefois, il ne s'agit pas à proprement parler d'une analyse en termes de « ségrégation » : en effet, l'indicateur de disparité retenu (l'écart-type) est par construction sensible à la composition sociale des collèges, et à la façon dont celle-ci évolue au cours du temps.

Par rapport aux travaux cités, notre contribution est de proposer une analyse pour la période récente (2003-2015), sur l'ensemble du territoire national et intégrant les collèges privés. L'apport d'un indice de ségrégation est d'abord de permettre une analyse de l'évolution des disparités entre collèges tenant compte de l'évolution de la composition sociale d'ensemble et dans chaque secteur. En utilisant l'indice d'entropie dans sa version multigroupe, nous souhaitons aussi apporter un diagnostic portant sur l'ensemble de la distribution des quatre catégories sociales, sans en choisir une en particulier. Enfin, les propriétés de décomposition de cet indicateur constituent, à nouveau, un atout pour ce type de démarche au regard, notamment, de l'indice de dissimilarité.

Entre 2003 et 2015, la ségrégation des collèges, ainsi mesurée par l'indice d'entropie, apparaît très stable ↘ **Figure 6**. L'indice d'entropie pour les collèges publics a diminué, celui ayant trait

12. Dans cette étude, une approche multigroupe est aussi proposée en s'appuyant sur l'écart-type de l'indice de synergie socio-scolaire moyen par établissement [voir *supra* et LE DONNÉ et ROCHER, 2010].

↳ **Figure 6 Évolution de l'indice d'entropie parmi l'ensemble des collèges, au sein des collèges publics et privés, et entre collèges publics et privés, entre 2003 et 2015**



Champ : collèges de France hors Mayotte, secteurs public et privé sous contrat.

Source : MENESR-DEPP, système d'information Scolarité.

aux collèges privés a peu varié. D'un autre côté, l'indice d'entropie qui mesure les différences de composition sociale entre collèges publics et privés a augmenté. Pour aboutir à une décomposition de l'indice d'entropie global sur ces trois composantes, il faut également tenir compte des poids à leur attribuer¹³ (voir *supra* l'équation (11) p. 31). La variation de + 0,001 de l'indice d'entropie entre 2003 et 2015 (de 0,096 à 0,097) se décompose en :

- 1) - 0,008 correspondant à la baisse de l'indice d'entropie parmi les collèges publics ;
- 2) + 0,007 correspondant à la hausse des écarts de composition sociale entre secteurs public et privé ;
- 3) + 0,002 dû à la très légère hausse de l'indice d'entropie parmi les collèges privés. Ainsi, la stabilité de l'indice d'entropie pour l'ensemble des collèges masque deux mouvements principaux qui se sont compensés : tout d'abord, parmi les seuls collèges publics, l'indice d'entropie a diminué. D'un autre côté, l'écart de composition sociale entre secteurs public et privé s'est accentué : plus concrètement, cet écart croissant s'interprète par le fait que les collèges privés accueillent des élèves de plus en plus favorisés au cours du temps, dans l'absolu et relativement à la moyenne des collèges publics¹⁴.

Une façon alternative de présenter ces résultats est de dire que, si la ségrégation sociale est restée stable, sa structure a sensiblement évolué : si l'essentiel de la ségrégation est dû

¹³. Ainsi, l'évolution de l'entropie parmi les collèges privés est non seulement de moindre ampleur que les autres, mais elle a aussi un poids plus faible dans l'indice d'entropie global (ce poids dépend notamment de la part du secteur privé parmi les élèves). C'est ce qui permet de comprendre pourquoi son rôle dans l'évolution de l'indice d'entropie global est ici très faible.

¹⁴. Cette information-là n'est pas contenue dans l'indice d'entropie « lié à la segmentation public/privé », qui en toute rigueur ne mesure que des écarts de « diversité ». C'est l'examen des parts des différentes origines sociales dans le public et dans le privé qui permet d'aboutir à cette interprétation.

à celle observée dans les collèges publics (59 % de la ségrégation¹⁵), cette part a fortement diminué, en lien avec la baisse de la ségrégation parmi les collèges publics (elle était de 68 % en 2003). Inversement, la part de la ségrégation attribuable aux écarts de diversité entre collèges publics et privés a augmenté, passant de 11 % à 18 %. Il s'agit d'un résultat à l'échelle nationale : l'article compagnon de ce numéro montre que cette structure de la ségrégation et l'évolution de cette dernière prennent des formes différentes d'une académie à l'autre.

Décomposer la ségrégation suivant les catégories sociales pour repérer des évolutions différenciées : l'exemple de Paris

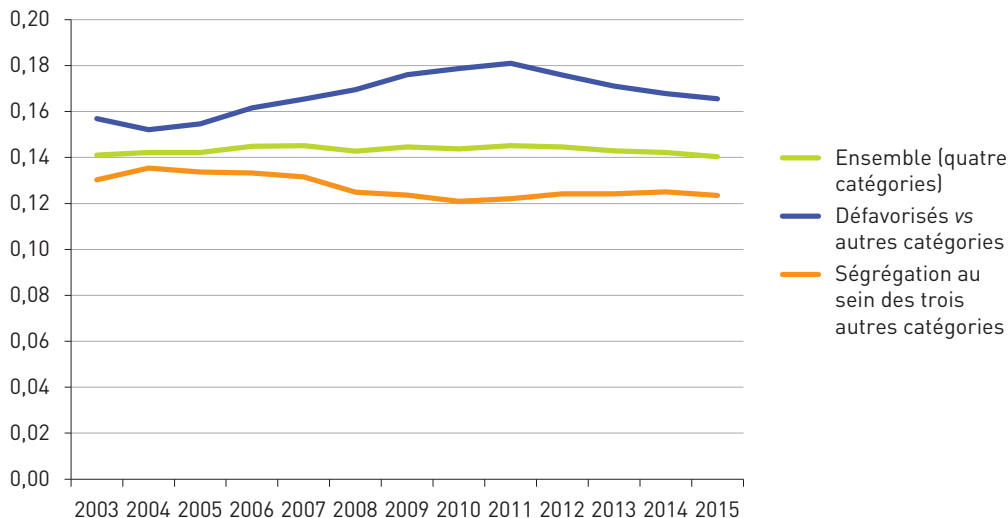
Nous avons présenté l'indice multigroupe comme ayant l'atout de ne privilégier aucune catégorie sociale en particulier dans l'étude de la ségrégation ↘ **Figure 6**. Comme cela a déjà été indiqué, cela ne signifie nullement que les approches « binaires », considérant la ségrégation envers une catégorie particulière soient d'un moindre intérêt. La plupart des travaux français s'appuient sur de telles approches, souvent en considérant la ségrégation envers les élèves d'origine défavorisée, comme TRANCART [2012], ou parfois en isolant la catégorie sociale « opposée », c'est-à-dire les élèves très favorisés [LY et RIEGERT, 2015]. L'apport d'un indice multigroupe comme l'indice d'entropie est d'abord de pouvoir prendre en compte simultanément les différents groupes sociaux dans l'analyse. Il est aussi de pouvoir calculer la ségrégation s'exerçant envers chacune des catégories sociales, prise isolément, et notamment les deux catégories « extrêmes » qui viennent d'être mentionnées. Cette étape-là est nécessaire afin de vérifier que le constat fait globalement sur les quatre catégories sociales ne « masque » pas un constat remarquable sur l'une des catégories. Enfin, le caractère décomposable de l'indice d'entropie permet de faire le lien entre ces deux approches : en effet, il permet non seulement de calculer la ségrégation s'exerçant envers une catégorie particulière, mais aussi d'explicitier sa contribution à la ségrégation d'ensemble, calculée sur les quatre catégories.

Nous en donnons une illustration à partir de la ségrégation entre collèges étudiée sur la ville de Paris ↘ **Figure 7**. MERLE [2010 ; 2012] a mis en avant une hausse de la ségrégation sociale entre les collèges de Paris entre 2005 et 2010, à partir d'un indice de dissimilarité (binaire) calculé en opposant les élèves d'origine « défavorisée » à ceux des autres catégories. Ayant mené le calcul avec l'indice d'entropie multigroupe, nous ne trouvons guère d'évolution dans la ségrégation entre collèges sur cette période, et de façon plus large sur la période 2003-2015. Est-il possible, par-delà les différences entre les indices utilisés, de réconcilier ces résultats ? Nous utilisons pour cela les propriétés de décomposition de l'indice d'entropie multigroupe suivant la ségrégation s'exerçant envers les différents groupes concernés (voir l'équation (9) p. 30). L'indice d'entropie portant sur les quatre catégories sociales regroupées peut s'exprimer en distinguant l'indice binaire calculé en considérant le groupe des élèves de milieu défavorisé *versus* les autres élèves, d'une part, et l'indice multigroupe calculé sur les trois autres catégories (en excluant donc les enfants de milieu défavorisé du calcul), d'autre part.

Pour les seuls enfants d'origine défavorisée, nous mettons en évidence une légère hausse de la ségrégation depuis 2003 dans les collèges parisiens, plus marquée sur la période de

15. Résultat obtenu là encore à partir de l'indice d'entropie dans le public, pondéré comme dans l'équation (11) (p. 31), et divisé par la ségrégation totale. L'interprétation est facilitée par le fait que les poids relatifs aux différentes composantes ont peu varié au cours du temps (la part du privé et la répartition des catégories sociales sont restées assez stables). Autrement dit, il revient ici au même de commenter l'évolution de l'indice d'entropie parmi les collèges publics et celle de la contribution correspondante dans la décomposition de l'indice d'entropie.

📉 **Figure 7 Évolution de la ségrégation dans les collèges parisiens et de ses composantes, entre 2003 et 2015**



Champ : collèges de Paris, secteurs public et privé sous contrat.

Source : MENESR-DEPP, système d'information Scolarité.

cing ans (2005-2010) étudiée par MERLE. D'un autre côté, la ségrégation au sein des trois autres catégories sociales est plutôt orientée à la baisse, tout du moins sur la période 2005-2010. Ces deux évolutions se sont compensées¹⁶ et permettent ainsi d'interpréter la stabilité de l'indice d'entropie portant sur les quatre groupes sociaux. Sur la période 2005-2010, la quasi-stabilité de l'indice d'entropie (+ 0,002) se décompose ainsi : + 0,007 en lien avec la ségrégation des élèves d'origine défavorisée (*versus* les autres catégories) et - 0,006 en lien avec la ségrégation des trois autres catégories (les élèves d'origine défavorisée étant exclus du calcul).

Enfin, en 2015, la ségrégation des élèves d'origine défavorisée contribue à 47 % de l'indice d'entropie à Paris. Le fait que la ségrégation de ces élèves ne rende compte « que » de la moitié de la ségrégation entre catégories sociales à Paris illustre un autre intérêt de l'utilisation d'un indice multigroupe. En poursuivant la décomposition, la ségrégation des élèves « très favorisés » vis-à-vis des catégories « favorisées » et « moyennes » prises ensemble correspond à 49 % de la ségrégation totale¹⁷. À nouveau, il n'est pas question, à travers cet exemple, d'apporter un jugement quant à la « bonne » mesure de la ségrégation sociale. Notre apport est plutôt d'illustrer comment, à partir de l'indice d'entropie, nous pouvons mener à la fois une analyse multigroupe et une analyse binaire, et surtout, réconcilier les deux approches, en apportant une interprétation aux différences de résultats auxquelles elles peuvent aboutir.

¹⁶. Ces deux composantes sont pondérées par des poids dépendant des parts des différentes catégories sociales parmi les parents d'élèves. On peut toutefois, comme pour le précédent exercice de décomposition, vérifier que ces poids sont très peu variables sur la période considérée et n'interviennent pas dans l'interprétation du résultat.

¹⁷. Il reste donc 4 % de la ségrégation imputable à la seule ségrégation entre élèves « favorisés » et « moyens », ce qui présente une contribution et un intérêt limités.

L'apport de l'indice d'entropie aux analyses sur la ségrégation entre collègues

En conclusion, si l'analyse des disparités de composition sociale des collègues est un sujet récurrent dans les travaux de la DEPP, une telle analyse, lorsqu'elle est menée en termes de ségrégation sociale, exige l'appropriation et l'utilisation d'outils spécifiques, sous la forme d'indices de ségrégation. Ces indices permettent de comparer l'hétérogénéité des collègues, en termes de composition sociale, entre différents territoires ou, pour un même territoire, au cours du temps, en tenant compte du fait que les compositions sociales des territoires concernés sont différentes.

Plusieurs indices sont à même de remplir cet objectif et il est bien difficile, étant donné la littérature existante, d'en définir un qui serait « meilleur » que les autres. Notre contribution propose quelques éléments et références au lecteur qui veut se faire sa propre opinion. En même temps, notre démarche se veut pragmatique : nous recommandons de calculer plusieurs indices pour assurer la robustesse des résultats. Constatant ici que les indices sont très corrélés les uns aux autres, nous avons cherché celui qui nous semblait apporter la plus grande richesse d'interprétation, au vu de la problématique posée, ayant trait à la ségrégation sociale entre les collègues.

Une première exigence, au vu de la variable utilisée (la catégorie sociale en quatre catégories) était que l'indice puisse être appliqué de façon assez « naturelle » à la présence de plusieurs groupes sociaux. Il est vrai que d'autres indices de ségrégation ont une version « multi-groupe », mais pour plusieurs d'entre eux, nous avons vu que ce passage au « multigroupe » est complexe, soit parce qu'il implique une combinaison d'indices sur des variables binaires (ségrégation envers les enfants d'origine défavorisée, très favorisée, etc.) et qu'il soulève donc le problème du choix de pondérations à leur accorder ; soit parce que ce passage s'accompagne d'une perte de propriétés ou de lisibilité. L'intérêt de l'indice d'entropie apparaît donc surtout dans l'étude de la ségrégation entre plusieurs catégories sociales.

Une deuxième exigence était que l'indice permette des décompositions utiles à l'analyse. Dans le cas de la ségrégation entre collègues, une telle analyse doit notamment tenir compte de l'existence du secteur privé sous contrat et de l'intérêt à le distinguer dans l'analyse. L'indice choisi devait donc permettre de mesurer séparément la ségrégation dans le public et dans le privé, d'isoler leur contribution respective à la ségrégation d'ensemble, ainsi que d'évaluer la ségrégation liée aux différences de composition sociale entre collègues publics et privés. L'indice d'entropie permet de faire une telle décomposition, et d'apporter notamment ainsi un éclairage sur l'évolution de la ségrégation sociale entre collègues au cours des dix dernières années. Il permet également d'isoler, dans la ségrégation totale, la contribution relative à la ségrégation s'exerçant envers telle ou telle catégorie sociale.

La principale limite de l'indice d'entropie est un certain manque de lisibilité, même si nous avons cherché à expliciter au mieux ses fondements ↘ **Annexe 1** p. 44. Rappelons que le choix de l'indice de ségrégation ne peut être univoque, il dépend des données utilisées (variables binaires ou à plus de deux modalités), de la problématique, du contexte de l'analyse. D'autres travaux ont fait le choix d'une approche de la ségrégation ciblant une catégorie sociale particulière, et à partir d'indices présentant dans ce cas une interprétation plus directe, comme l'indice de dissimilarité ou l'indice d'exposition normalisé. Par ailleurs, dans l'outil « carte scolaire » que la DEPP a mis à disposition des acteurs locaux, c'est l'indice d'exposition normalisé qui a été retenu. En l'occurrence c'est la volonté de donner aux acteurs un indice le plus lisible possible qui a prévalu.

Une autre limite, qui est partagée par tous les indices présentés ici, est qu'ils sont « aspatiaux », au sens où ils ne prennent pas en compte la répartition spatiale des élèves des différents groupes sociaux. Si cette dimension spatiale n'a pas de sens au niveau des collèges (on peut considérer qu'au sein d'un établissement scolaire les élèves sont tous « regroupés »), la ségrégation entre collèges est aussi, du fait de la sectorisation, le résultat d'une ségrégation urbaine. Celle-ci peut être étudiée par le biais d'indices de concentration intégrant plus directement ces dimensions à partir des lieux de résidence des élèves [voir par exemple RATHELOT et SILLARD, 2010 ; MARCON et PUECH, 2014]. Par ailleurs, la concurrence locale entre établissements, et particulièrement entre établissements privés et publics, peut créer des phénomènes de polarisation locale. Le diagnostic doit aussi descendre à un niveau plus local encore, et permettre de situer un collège dans son environnement. C'est en effet l'information dont l'expert et le décideur public ont besoin pour mieux appréhender les logiques de la ségrégation et, le cas échéant, agir sur la carte scolaire. Ces points sont investigués dans un article « compagnon » dans ce même numéro (p. 53).

Les auteurs remercient Cédric Afsa (DEPP) pour sa contribution aux échanges et ses conseils, ainsi que les participants des Journées de Méthodologie Statistique, de l'atelier de la DEPP, du Séminaire de l'Observation Urbaine, du séminaire Politiques Éducatives du LIEPP et particulièrement les discutants, Danièle Trancart et Jean-Christophe Vergnaud, enfin deux relecteurs anonymes de la revue. Une première version de ce travail [GIVORD, GUILLERM *et alii*, 2015] est parue dans les Actes des XII^e Journées de Méthodologie Statistique (Paris, 31 mars-2 avril 2015).

Annexe 1

POURQUOI L'ENTROPIE EST-ELLE UNE MESURE DE LA DIVERSITÉ ?

L'un des principes des indices de ségrégation est de caractériser la mixité sociale dans un établissement comme la présence d'une grande diversité de milieux sociaux. Le concept d'entropie, correspondant au degré de désorganisation d'un système, peut être d'un tel usage. Il est relié à la théorie de l'information de SHANNON [1948]. Ce dernier a élaboré une théorie mathématique de l'information et montré que l'entropie telle que nous l'avons définie dans le texte était une fonction qui avait les bonnes propriétés pour mesurer la quantité d'information d'une source (plus précisément, l'information minimale que l'on doit conserver afin de représenter ces données sans perte d'information)¹⁸. On peut essayer de transposer ce concept à la mesure de la diversité sociale des collèges. La théorie de l'information peut se résumer dans le fait qu'un système est d'autant plus complexe à décrire qu'il est grand et qu'il contient des éléments différents. Pour illustrer cela en l'adaptant au cadre qui est le nôtre, supposons que l'objectif soit de trouver un élève particulier au sein d'un collège et qu'on définisse la complexité comme le nombre minimal de questions qu'il faudra poser pour trouver cet élève particulier parmi la liste de l'ensemble des élèves. On va alors s'intéresser à l'information apportée par la connaissance *a priori* de sa catégorie sociale (de combien cette information permet de réduire le nombre de questions à poser pour trouver cet élève).

Intuitivement, le nombre de questions à poser pour trouver un élève particulier est d'autant plus élevé que la taille de l'établissement est grande. Cette relation n'est cependant pas linéaire : à une recherche séquentielle (consistant à parcourir la liste des noms, avec des questions : « *Le premier élève est-il l'élève recherché ?* », si ce n'est pas le cas, poursuivre par : « *Est-ce le deuxième ?* », etc.), on peut adopter une stratégie itérative plus efficace, appelée recherche dichotomique. Supposons, pour simplifier l'exposé, que l'établissement comporte t élèves avec $t = 2^n$. Pour trouver un élève, on partage la liste d'élèves par moitié et on demande dans quelle moitié se trouve l'élève recherché. On va ensuite réitérer ce processus avec la moitié où il se trouve et ainsi de suite, jusqu'à tomber sur lui. Cette procédure permet d'arriver à trouver l'élève en n questions. Or $n = \log(t) / \log(2)$. On peut généraliser ce résultat à un nombre t quelconque : $\log(t) / \log(2)$ est une bonne estimation du nombre moyen de questions à poser pour trouver un élève. Ce nombre est bien sûr croissant avec la taille de la population, mais pas de façon proportionnelle, car des stratégies, comme la recherche dichotomique, permettent de gagner en efficacité par rapport à une recherche séquentielle, pour se repérer dans les gros établissements.

Supposons maintenant que la population de l'établissement puisse être divisée en C catégories (par exemple liées aux catégories sociales), avec un effectif de n_i pour la catégorie i . En quoi connaître la catégorie de l'élève recherché va-t-il réduire la complexité de la démarche – dit autrement, quelle information nous apporte le fait

18. Pour une présentation plus proche de l'original, voir l'article Wikipedia à l'adresse https://fr.wikipedia.org/wiki/Entropie_de_Shannon.

de savoir la catégorie à laquelle l'élève recherché appartient ? Avoir cette information va nous permettre de réduire nos recherches parmi les élèves de la catégorie i : le nombre de questions à poser en utilisant une recherche dichotomique décrite plus haut sera donc d'ordre $\log(n_i)$.

Le gain apporté par la connaissance de la catégorie d'un élève quelconque sur le nombre de questions à poser pour le trouver parmi l'ensemble des élèves de l'établissement, par rapport à la situation où on ne connaît pas cette catégorie, sera donc d'ordre :

$$\begin{aligned} \log(t) - \sum_{i=1}^C \frac{n_i}{t} \log(n_i) &= \sum_{i=1}^C \frac{n_i}{t} (\log(t) - \log(n_i)) \\ &= \sum_{i=1}^C \frac{n_i}{t} \log\left(\frac{t}{n_i}\right) \\ &= \sum_{i=1}^C q_i \log\left(\frac{1}{q_i}\right) \end{aligned} \quad [19]$$

où $q_i = \frac{n_i}{t}$

En effet un élève quelconque appartient à une catégorie i dans n_i cas sur t , et en moyenne le nombre de questions à poser dans un cas où on a une information initiale sur la catégorie de l'élève est d'ordre :

$$\sum_{i=1}^C \frac{n_i}{t} \log(n_i)$$

Cette formule correspond à celle de l'entropie indiquée dans le texte pour un établissement. On notera que la réduction du nombre de questions (l'information apportée par la connaissance de la catégorie sociale) et donc l'entropie sont indépendants de la taille de l'établissement, puisqu'elle ne dépend que des proportions des différentes catégories, ce qui est une propriété souhaitable.

Les indices d'entropie utilisés correspondent alors à la comparaison entre la diversité sur l'ensemble de la population (mesurée par l'entropie globale) et la moyenne de ces mesures au niveau des établissements.

Annexe 2

LA SÉGRÉGATION DANS LES COLLÈGES
SELON L'ACADÉMIE ET LE DÉPARTEMENT, RENTRÉE 2015

Départements, académies	Valeur des indices				Indice d'entropie selon le secteur			Contributions à l'indice d'entropie selon le secteur		
	Indice d'entropie	Indice d'information mutuelle	Indice de dissimilarité	Indice d'Atkinson	Collèges publics	Collèges privés	Segm. public/ privé	Part de l'indice coll. publics	Part de l'indice coll. privés	Part de l'indice segm. public/ privé
Alpes-de- Haute- Provence	0,043	0,057	0,172	0,054	0,037	0,018	0,007	79 %	3 %	17 %
Hautes-Alpes	0,017	0,023	0,114	0,026	0,013	0,000	0,006	68 %	0 %	32 %
Bouches-du- Rhône	0,123	0,162	0,296	0,153	0,096	0,118	0,026	60 %	19 %	21 %
Vaucluse	0,067	0,086	0,224	0,085	0,051	0,039	0,020	58 %	12 %	30 %
Aix-Marseille	0,105	0,138	0,271	0,131	0,082	0,100	0,022	61 %	18 %	21 %
Aisne	0,057	0,068	0,209	0,077	0,036	0,032	0,023	52 %	8 %	40 %
Oise	0,078	0,102	0,229	0,102	0,048	0,078	0,027	51 %	14 %	35 %
Somme	0,087	0,106	0,262	0,104	0,046	0,058	0,039	38 %	16 %	45 %
Amiens	0,079	0,100	0,239	0,102	0,049	0,069	0,028	50 %	15 %	35 %
Doubs	0,056	0,073	0,208	0,069	0,049	0,077	0,003	71 %	24 %	5 %
Jura	0,026	0,033	0,138	0,034	0,023	0,027	0,002	71 %	22 %	7 %
Haute-Saône	0,024	0,030	0,127	0,036	0,019	0,031	0,003	71 %	18 %	11 %
Territoire de Belfort	0,043	0,056	0,182	0,049	0,028	0,013	0,018	52 %	5 %	42 %
Besançon	0,045	0,058	0,180	0,057	0,037	0,064	0,003	68 %	25 %	7 %
Dordogne	0,031	0,040	0,152	0,045	0,024	0,044	0,005	65 %	19 %	17 %
Gironde	0,081	0,108	0,239	0,101	0,057	0,086	0,021	58 %	16 %	26 %
Landes	0,029	0,037	0,146	0,043	0,024	0,037	0,004	75 %	12 %	13 %
Lot-et- Garonne	0,034	0,043	0,153	0,048	0,018	0,041	0,013	44 %	19 %	37 %
Pyrénées- Atlantiques	0,050	0,067	0,194	0,068	0,033	0,061	0,008	44 %	39 %	17 %
Bordeaux	0,066	0,087	0,215	0,086	0,046	0,087	0,013	56 %	24 %	20 %
Calvados	0,076	0,099	0,237	0,097	0,054	0,078	0,018	55 %	22 %	23 %
Manche	0,046	0,060	0,190	0,064	0,036	0,040	0,010	56 %	22 %	22 %
Orne	0,057	0,070	0,210	0,074	0,034	0,051	0,020	44 %	22 %	34 %
Caen	0,067	0,086	0,224	0,088	0,049	0,070	0,014	54 %	24 %	21 %
Allier	0,033	0,041	0,150	0,047	0,025	0,024	0,008	68 %	7 %	25 %
Cantal	0,044	0,056	0,179	0,074	0,042	0,034	0,003	81 %	11 %	7 %
Haute-Loire	0,040	0,051	0,178	0,055	0,030	0,045	0,003	44 %	48 %	9 %
Puy-de-Dôme	0,066	0,089	0,211	0,083	0,051	0,077	0,011	61 %	22 %	17 %
Clermont- Ferrand	0,057	0,075	0,200	0,076	0,044	0,076	0,007	61 %	28 %	12 %
Corse-du-Sud	0,058	0,074	0,213	0,084	0,052	0,000	0,009	85 %	0 %	15 %
Haute-Corse	0,038	0,047	0,162	0,055	0,027	0,000	0,013	65 %	0 %	35 %
Corse	0,048	0,061	0,187	0,070	0,040	0,013	0,010	78 %	2 %	20 %

COMMENT MESURER LA SÉGRÉGATION DANS LE SYSTÈME ÉDUCATIF ?

Départements, académies	Valeur des indices				Indice d'entropie selon le secteur			Contributions à l'indice d'entropie selon le secteur		
	Indice d'entropie	Indice d'information mutuelle	Indice de dissimilarité	Indice d'Atkinson	Collèges publics	Collèges privés	Segm. public/privé	Part de l'indice coll. publics	Part de l'indice coll. privés	Part de l'indice segm. public/privé
Seine-et-Marne	0,070	0,094	0,218	0,091	0,053	0,022	0,022	65 %	3 %	31 %
Seine-Saint-Denis	0,094	0,112	0,258	0,131	0,064	0,063	0,032	56 %	10 %	34 %
Val-de-Marne	0,116	0,154	0,297	0,147	0,089	0,065	0,033	64 %	7 %	29 %
Créteil	0,105	0,137	0,272	0,137	0,081	0,068	0,028	65 %	8 %	26 %
Côte-d'Or	0,050	0,068	0,188	0,066	0,038	0,022	0,015	63 %	7 %	31 %
Nièvre	0,047	0,060	0,185	0,069	0,036	0,038	0,011	68 %	8 %	24 %
Saône-et-Loire	0,040	0,050	0,170	0,053	0,028	0,034	0,012	60 %	10 %	29 %
Yonne	0,038	0,047	0,161	0,051	0,023	0,007	0,017	53 %	2 %	45 %
Dijon	0,049	0,064	0,189	0,066	0,035	0,034	0,014	62 %	9 %	29 %
Ardèche	0,030	0,040	0,162	0,039	0,022	0,038	0,003	50 %	39 %	11 %
Drôme	0,048	0,065	0,191	0,062	0,036	0,031	0,014	55 %	16 %	29 %
Isère	0,064	0,086	0,214	0,078	0,055	0,054	0,010	70 %	15 %	16 %
Savoie	0,038	0,051	0,162	0,056	0,030	0,047	0,006	67 %	17 %	16 %
Haute-Savoie	0,048	0,064	0,178	0,062	0,029	0,063	0,011	45 %	33 %	22 %
Grenoble	0,054	0,073	0,195	0,070	0,043	0,060	0,008	62 %	23 %	15 %
Guadeloupe	0,062	0,074	0,188	0,092	0,030	0,091	0,025	42 %	18 %	40 %
Guyane	0,119	0,115	0,306	0,145	0,078	0,043	0,048	56 %	4 %	40 %
La Réunion	0,082	0,090	0,239	0,115	0,046	0,045	0,039	48 %	5 %	47 %
Nord	0,119	0,151	0,300	0,144	0,070	0,111	0,038	37 %	31 %	32 %
Pas-de-Calais	0,078	0,093	0,242	0,098	0,039	0,049	0,039	37 %	13 %	49 %
Lille	0,108	0,134	0,286	0,133	0,060	0,098	0,040	37 %	26 %	37 %
Corrèze	0,040	0,051	0,160	0,050	0,033	0,033	0,007	71 %	11 %	18 %
Creuse	0,021	0,025	0,115	0,041	0,021	0,000	0,000	100 %	0 %	0 %
Haute-Vienne	0,054	0,070	0,191	0,071	0,037	0,021	0,019	60 %	4 %	36 %
Limoges	0,048	0,061	0,179	0,064	0,036	0,039	0,012	67 %	8 %	25 %
Ain	0,054	0,072	0,199	0,063	0,043	0,041	0,012	64 %	14 %	22 %
Loire	0,054	0,069	0,201	0,066	0,041	0,057	0,008	53 %	31 %	16 %
Rhône	0,112	0,150	0,285	0,135	0,086	0,080	0,030	51 %	22 %	27 %
Lyon	0,092	0,123	0,255	0,110	0,068	0,084	0,021	51 %	26 %	22 %
Martinique	0,067	0,085	0,206	0,095	0,037	0,052	0,029	48 %	8 %	44 %
Aude	0,037	0,047	0,150	0,054	0,016	0,046	0,018	37 %	14 %	49 %
Gard	0,068	0,089	0,227	0,090	0,062	0,032	0,013	71 %	10 %	19 %
Hérault	0,083	0,109	0,243	0,102	0,068	0,059	0,017	68 %	11 %	20 %
Lozère	0,024	0,031	0,137	0,035	0,023	0,015	0,003	67 %	19 %	14 %
Pyrénées-Orientales	0,071	0,090	0,207	0,091	0,043	0,044	0,029	48 %	11 %	41 %
Montpellier	0,073	0,095	0,224	0,094	0,059	0,056	0,016	65 %	13 %	22 %

Départements, académies	Valeur des indices				Indice d'entropie selon le secteur			Contributions à l'indice d'entropie selon le secteur		
	Indice d'entropie	Indice d'information mutuelle	Indice de dissimilarité	Indice d'Atkinson	Collèges publics	Collèges privés	Segm. public/privé	Part de l'indice coll. publics	Part de l'indice coll. privés	Part de l'indice segm. public/privé
Meurthe-et-Moselle	0,071	0,093	0,221	0,090	0,046	0,100	0,017	53%	23%	24%
Meuse	0,042	0,052	0,180	0,066	0,027	0,039	0,014	53%	14%	33%
Moselle	0,076	0,097	0,239	0,091	0,054	0,059	0,022	59%	12%	29%
Vosges	0,038	0,047	0,160	0,053	0,025	0,063	0,007	57%	23%	20%
Nancy-Metz	0,068	0,088	0,221	0,087	0,047	0,079	0,017	57%	17%	25%
Loire-Atlantique	0,073	0,099	0,222	0,090	0,059	0,077	0,008	47%	43%	10%
Maine-et-Loire	0,061	0,080	0,210	0,077	0,046	0,059	0,009	39%	46%	15%
Mayenne	0,045	0,057	0,180	0,062	0,025	0,055	0,008	31%	52%	17%
Sarthe	0,060	0,078	0,209	0,078	0,038	0,067	0,016	47%	26%	27%
Vendée	0,037	0,048	0,170	0,051	0,022	0,037	0,007	26%	56%	18%
Nantes	0,069	0,091	0,222	0,087	0,053	0,071	0,008	44%	44%	12%
Alpes-Maritimes	0,104	0,138	0,267	0,134	0,087	0,060	0,024	69%	8%	23%
Var	0,058	0,076	0,193	0,081	0,040	0,032	0,021	58%	7%	35%
Nice	0,084	0,110	0,233	0,111	0,065	0,049	0,022	65%	8%	27%
Cher	0,042	0,053	0,174	0,055	0,029	0,035	0,012	62%	9%	29%
Eure-et-Loir	0,062	0,082	0,214	0,080	0,045	0,066	0,014	59%	18%	23%
Indre	0,038	0,046	0,163	0,054	0,027	0,030	0,011	63%	9%	29%
Indre-et-Loire	0,061	0,082	0,206	0,080	0,043	0,059	0,016	56%	17%	27%
Loir-et-Cher	0,057	0,073	0,199	0,076	0,033	0,063	0,019	48%	19%	33%
Loiret	0,065	0,085	0,212	0,084	0,045	0,062	0,018	59%	13%	28%
Orléans-Tours	0,062	0,081	0,211	0,082	0,043	0,064	0,016	58%	15%	26%
Paris	0,140	0,170	0,337	0,163	0,086	0,130	0,047	43%	24%	33%
Charente	0,056	0,071	0,193	0,074	0,036	0,046	0,019	53%	13%	34%
Charente-Maritime	0,047	0,061	0,189	0,064	0,037	0,047	0,009	69%	12%	20%
Deux-Sèvres	0,052	0,068	0,192	0,069	0,041	0,077	0,003	60%	34%	6%
Vienne	0,052	0,069	0,190	0,071	0,039	0,067	0,009	59%	24%	17%
Poitiers	0,054	0,070	0,197	0,071	0,041	0,070	0,008	63%	22%	15%
Ardennes	0,063	0,074	0,218	0,086	0,040	0,023	0,027	51%	6%	43%
Aube	0,059	0,073	0,204	0,074	0,032	0,072	0,021	45%	20%	35%
Marne	0,083	0,107	0,242	0,106	0,056	0,069	0,026	50%	18%	31%
Haute-Marne	0,048	0,057	0,188	0,071	0,027	0,012	0,023	48%	3%	49%
Reims	0,074	0,092	0,229	0,096	0,047	0,061	0,026	50%	15%	35%

COMMENT MESURER LA SÉGRÉGATION DANS LE SYSTÈME ÉDUCATIF ?

Départements, académies	Valeur des indices				Indice d'entropie selon le secteur			Contributions à l'indice d'entropie selon le secteur		
	Indice d'entropie	Indice d'information mutuelle	Indice de dissimilarité	Indice d'Atkinson	Collèges publics	Collèges privés	Segm. public/privé	Part de l'indice coll. publics	Part de l'indice coll. privés	Part de l'indice segm. public/privé
Côtes-d'Armor	0,050	0,065	0,195	0,067	0,034	0,060	0,007	46 %	40 %	14 %
Finistère	0,044	0,059	0,179	0,059	0,032	0,043	0,008	41 %	41 %	18 %
Ille-et-Vilaine	0,074	0,100	0,234	0,092	0,061	0,077	0,007	48 %	43 %	9 %
Morbihan	0,051	0,068	0,193	0,070	0,040	0,053	0,005	37 %	54 %	9 %
Rennes	0,060	0,080	0,209	0,078	0,047	0,063	0,006	45 %	45 %	10 %
Eure	0,046	0,059	0,174	0,061	0,027	0,045	0,018	50 %	12 %	39 %
Seine-Maritime	0,086	0,111	0,253	0,110	0,063	0,074	0,023	58 %	15 %	27 %
Rouen	0,073	0,093	0,226	0,093	0,050	0,066	0,021	56 %	15 %	29 %
Bas-Rhin	0,077	0,101	0,234	0,092	0,059	0,054	0,020	66 %	9 %	26 %
Haut-Rhin	0,072	0,092	0,227	0,089	0,052	0,024	0,027	56 %	7 %	38 %
Strasbourg	0,077	0,100	0,233	0,093	0,059	0,044	0,022	63 %	9 %	28 %
Ariège	0,018	0,023	0,115	0,024	0,014	0,017	0,003	68 %	12 %	19 %
Aveyron	0,042	0,054	0,182	0,059	0,029	0,057	0,003	44 %	47 %	8 %
Haute-Garonne	0,070	0,094	0,224	0,085	0,057	0,077	0,012	69 %	14 %	17 %
Gers	0,043	0,057	0,179	0,057	0,028	0,045	0,013	52 %	19 %	29 %
Lot	0,028	0,036	0,134	0,044	0,019	0,021	0,009	60 %	9 %	31 %
Hautes-Pyrénées	0,031	0,041	0,158	0,043	0,023	0,028	0,008	59 %	17 %	24 %
Tarn	0,039	0,051	0,167	0,054	0,030	0,034	0,009	61 %	17 %	22 %
Tarn-et-Garonne	0,035	0,044	0,161	0,045	0,020	0,043	0,010	45 %	25 %	29 %
Toulouse	0,068	0,091	0,221	0,086	0,055	0,092	0,007	67 %	24 %	10 %
Yvelines	0,133	0,169	0,321	0,156	0,108	0,077	0,032	69 %	7 %	24 %
Essonne	0,090	0,121	0,248	0,112	0,073	0,042	0,022	70 %	6 %	25 %
Hauts-de-Seine	0,147	0,182	0,342	0,169	0,097	0,115	0,051	52 %	13 %	35 %
Val-d'Oise	0,099	0,132	0,264	0,126	0,073	0,082	0,027	61 %	12 %	27 %
Versailles	0,133	0,175	0,316	0,158	0,101	0,109	0,035	62 %	12 %	26 %
Ensemble France, hors Mayotte	0,097	0,128	0,264	0,121	0,074	0,105	0,018	59 %	23 %	18 %

Champ : collèges de France hors Mayotte, secteurs public et privé sous contrat.

Source : MENESR-DEPP, système d'information Scolarité.

▣ BIBLIOGRAPHIE

- CARRINGTON W. J., TROSKE K. R., 1997, "On measuring segregation in samples with small units", *Journal of Business and Economic Statistics*, vol. 15, n° 4, p. 402-409.
- CHARLOT S., HILAL M., SCHMITT B., 2009, « La périurbanisation renforce-t-elle la ségrégation résidentielle urbaine en France ? », *Espace, Populations, Sociétés*, 2009-1, p. 29-44.
- DASRÉ A., 2012, *Les mesures du regroupement spatial des populations : aspects méthodologiques et applications aux grandes aires urbaines françaises*, Thèse de doctorat en démographie de l'Université Bordeaux 4.
- DAVEZIES L., GARROUSTE M., 2014, *More harm than good? Sorting effects in a compensatory education program*, Document de travail du CREST, n° 2014-42.
- DUNCAN O. D., DUNCAN B., 1955, "A methodological analysis of segregation indexes", *American Sociological Review*, vol. 20, n° 2, p. 210-217.
- FACK G., GRENET J., 2009, « Sectorisation des collèges et prix des logements à Paris », *Actes de la recherche en sciences sociales*, 2009-5, n° 108, p. 44-62.
- FACK G., GRENET J., BENHENDA A., 2014, *L'impact des procédures de sectorisation et d'affectation sur la mixité sociale et scolaire dans les lycées d'Île-de-France*, rapport IPP, n° 3.
- FRANKEL D. M., VOLIJ O., 2011, "Measuring school segregation", *Journal of Economic Theory*, vol. 146, n° 1, p. 1-38.
- GIRARD P., GILOTTE O., 2005, « La sectorisation, l'affectation et l'évitement scolaire dans les classes de sixième à Paris en 2003 », *Éducation & formations*, n° 71, MEN-DEP, p. 137-149.
- GIVORD P., GUILLERM M., MONSO O., MURAT F., AFSA C., 2015, « Comment mesurer la ségrégation dans le système éducatif ? Une application à la composition sociale des collèges français », *Actes des XII^e Journées de Méthodologie Statistique*, Paris, 31 mars-2 avril 2015.
- JASPAR M.-L., 2007, « Le recueil des professions et catégories sociales des parents des élèves dans le système d'information du second degré », *Éducation & formations*, n° 74, MENESR-DEPP, p. 19-20.
- LE DONNÉ N., ROCHER T., 2010, « Une meilleure mesure du contexte socio-éducatif des élèves et des écoles », *Éducation & formations*, n° 79, MENJVA-DEPP, p. 103-115.
- LY S. T., MAURIN E., RIEGERT A., 2014, *La mixité sociale et scolaire en Île-de-France : le rôle des établissements*, rapport IPP, n° 4.
- LY S. T., RIEGERT A., 2015, *Mixité sociale et scolaire, ségrégation inter et intra-établissement dans les collèges et lycées français*, rapport pour le conseil national d'évaluation du système scolaire.
- MARCON E., PUECH F., 2014, « Mesures de la concentration spatiale en espace continu : théorie et applications », *Économie et Statistique*, n° 474, Insee, p. 105-132.
- MENESR-DEPP, 2014, *Géographie de l'école*, n° 11, Paris, MENESR-DEPP.
- MERLE P., 2012, *La ségrégation scolaire*, Paris, La Découverte, coll. « Repères ».
- MERLE P., 2010, « Structure et dynamique de la ségrégation sociale dans les collèges parisiens », *Revue Française de Pédagogie*, n° 170, p. 73-85.
- MORGAN B. S., 1975, "The segregation of socioeconomic groups in urban areas: a comparative analysis", *Urban Studies*, vol. 12, n° 1, p. 47-60.
- POUPEAU F., FRANÇOIS J.-C., 2008, *Le sens du placement. Ségrégation résidentielle et ségrégation scolaire*, Paris, Raisons d'Agir, coll. « Cours et travaux ».
- RATHELOT R., 2012, "Measuring Segregation When Units are Small: A Parametric Approach", *Journal of Business and Economic Statistics*, 30(4), p. 546-553.
- RATHELOT R., SILLARD P., 2010, *L'apport des méthodes à noyaux pour mesurer la concentration géographique. Application à la concentration des immigrés en France de 1968 à 1999*, Document de travail de la direction des études et synthèses économiques, G 2010 / 11, Insee.
- REARDON S. F., FIREBAUGH G., 2002, "Measures of multigroup segregation", *Sociological methodology*, vol. 32, n° 1, p. 33-67.
- ROCHER T., 2016, « Construction d'un indice de position sociale des élèves », *Éducation & formations*, n° 90, MENESR-DEPP, p. 5-27.
- SHANNON C. E., 1948, « A mathematical theory of communication », *Bell System Technical Journal*, vol. 27, n° 3, p. 379-423 et 623-656.

THAUREL-RICHARD M., 2005, « Présentation du dossier "Typologie des collèges publics" », *Éducation & formations*, n° 71, MEN-DEP, p. 95-103.

THAUREL-RICHARD M., MURAT F., 2013, « Évolutions des caractéristiques des collèges durant la mise en œuvre de l'assouplissement de la carte scolaire de 2007 », *Éducation & formations*, n° 83, MEN-DEPP, p. 11-23.

THOMAS F. 2005, « Typologie des collèges publics. Disparité des collèges publics en 2003-2004 », *Éducation & formations*, n°71, MEN-DEP, p. 117-135.

TRANCART D., 2012, « Quel impact des ségrégations socio-spatiales sur la réussite scolaire au collège ? », *Formation Emploi*, n° 120, Céreq, p. 35-55.

VAN ZANTEN A., 2009, *Choisir son école. Stratégies familiales et médiations locales*, Paris, PUF, coll. « Le Lien Social ».

