

L'effet d'une réduction de la taille des classes sur la réussite scolaire en France : développements récents

Olivier Monso

Sous-direction des synthèses

Cellule des méthodes et synthèses statistiques

Direction de l'évaluation, de la prospective et de la performance

Un nombre d'élèves par classe plus faible semble permettre une meilleure réussite scolaire. Les travaux récents sur données françaises consacrés à l'impact de la taille des classes sur la réussite scolaire ont tous abouti, quoiqu'avec des nuances, à ce constat. Ces travaux ont mis en évidence un impact du nombre d'élèves par classe plus fort dans le premier degré et dans la première partie du collège, peu apparent ou inexistant au lycée général.

Une réduction du nombre d'élèves par classe bénéficie davantage, dans l'ensemble, aux élèves issus d'un environnement social ou scolaire défavorisé, et notamment à ceux scolarisés dans un établissement en éducation prioritaire. Toutefois, elle ne constitue pas forcément la réponse appropriée à tous les types de difficultés scolaires. Le lien entre taille des classes et réussite scolaire reste complexe, notamment parce que les mécanismes sous-jacents n'ont pas été clairement identifiés.

Enfin, les effets mesurés sont affectés d'une marge d'incertitude importante, ce qui pose aussi question pour mettre en œuvre une politique adaptée.

Diminuer le nombre d'élèves par classe permet-il d'améliorer les résultats scolaires des élèves ?

Cette question donne lieu à des débats récurrents entre chercheurs, institutions publiques, *think tanks*,... Deux grandes positions concurrentes peuvent être présentées de façon sommaire, en s'inspirant des travaux internationaux.

La première est sceptique sur l'efficacité d'une diminution du nombre d'élèves par classe. Ce scepticisme s'étend souvent à toutes les politiques visant à accroître les moyens (en enseignants, financiers) par élève. Cette position est fréquemment associée à Hanushek, développée dans plusieurs de ses travaux (voir par exemple [13]).

La seconde voit au contraire dans la réduction de la taille des classes un levier important pour améliorer les résultats scolaires des élèves. Cette position peut être associée à des travaux ayant mis en évidence des résultats allant dans cette direction [3, 12].

Il faut se garder de caricaturer ces positions, qui sont souvent accompagnées de nuances. Hanushek lui-même ne nie pas qu'il y ait des situations dans lesquelles une politique de réduction de taille des classes

puisse être efficace. Toutefois, dit-il, on ne sait pas vraiment sous quelles conditions cet effet existe. De façon plus générale, Hanushek met en garde contre un raisonnement, selon lui trop systématique, qui conduit à répondre aux difficultés scolaires en augmentant les moyens par élève. Un tel schéma de pensée peut empêcher de voir des solutions plus innovantes, et peut-être plus efficaces.

De même, les tenants d'une diminution de la taille des classes ne sont pas forcément partisans d'une réduction massive et pour toutes les classes. Ils mettent en évidence des types d'élèves ou d'établissements pour lesquels cette diminution serait la plus efficace. Piketty et Valdenaire [21] et Valdenaire [22], dans des travaux souvent cités sur données françaises, mettent d'ailleurs en avant que cette politique peut être appliquée à moyens constants au niveau global, en ciblant cette réduction et les moyens associés sur les élèves des zones d'éducation prioritaire.

En 2001, suite à une demande du Haut Conseil de l'évaluation de l'école, Meuret [16] a remis un rapport faisant le point sur les recherches menées sur ce sujet. Ce rapport dressait une liste de nombreux travaux français ayant abordé la question, et les mettait en regard, notamment, des

travaux américains. Les résultats issus de travaux américains étaient beaucoup plus favorables à une réduction de la taille des classes. En France, les effets d'une réduction de la taille des classes sur les performances étaient souvent très faibles, non significativement différents de zéro.

Face à ce constat, deux interprétations sont possibles. La première est que les systèmes éducatifs français et américain sont différents. Il est logique que les résultats des recherches américaines ne puissent pas s'appliquer tels quels à la France¹. La seconde interprétation s'appuie sur les différences des méthodes mises en œuvre. Les travaux américains ont, plus tôt qu'en France, utilisé des méthodes expérimentales, ou se rapprochant des conditions d'une expérience (« quasi-expérimentales »). Ils cherchaient à mieux prendre en compte une difficulté méthodologique primordiale dans cette littérature : le fait que les élèves ne sont pas répartis aléatoirement entre les classes, les élèves les plus faibles ayant plutôt tendance à être dans les classes les moins chargées.

Meuret [16] affichait sa préférence pour la seconde hypothèse : les écarts entre les conclusions des travaux français et américains qu'il cite relevaient surtout, selon lui, des différences de méthodes utilisées, plutôt que de spécificités des systèmes éducatifs. Partant de là, il concluait que l'effet d'une réduction de la taille des classes pouvait être positif et important à condition que cette réduction soit forte (en amenant les classes nettement en dessous de 20 élèves) et ciblée sur les enfants d'origine défavorisée.

Dans cet article, nous proposons un panorama des travaux sur données françaises² qui ont suivi ce

rapport. Ces travaux ont appliqué des méthodes assez similaires à celles employées dans les principales recherches internationales. Le lecteur trouvera en annexe un récapitulatif de ces travaux sous forme de tableau, avec les principaux résultats et éléments de méthode.

Mesurer les effets d'une taille des classes plus faible sur la réussite scolaire est complexe. Deux types de difficultés peuvent être identifiés. Le premier a trait à la mesure des deux objets dont il est question : qu'est-ce qu'une taille des classes ? Qu'est-ce que la réussite scolaire ? Le second est relatif à la difficulté de passer d'une corrélation entre ces deux objets, facile à observer, à une relation de causalité, beaucoup plus difficile à mettre en évidence.

TAILLE DES CLASSES, RÉUSSITE SCOLAIRE : QUE MESURE-T-ON ?

Par « taille des classes », on entend effectif d'élèves dans une classe, souvent mesuré en début d'année scolaire. La classe est ici une division administrative : par exemple la seconde C du lycée X. Cette définition n'est pas si évidente à manier, et comporte des risques de confusion. Une taille des classes ne mesure pas la même réalité dans une école, un collège ou un lycée, ou d'un établissement à un autre.

Il est important de distinguer classe et niveau. Dans le premier degré, une classe ne correspond pas forcément à un seul niveau : dans une même classe, des élèves de CP peuvent côtoyer des élèves de CE1. À la rentrée 2013, 46,5 % des classes du premier degré comportent plusieurs niveaux [1] : 57,6 % des classes pré-

élémentaires et 40,3 % des classes élémentaires. La « taille de la classe » veut-elle dire la même chose pour les élèves de ces classes que pour les autres ? Est-elle liée de façon similaire à la réussite scolaire ? Des chercheurs ont répondu à ces questions en excluant ces classes (ou les écoles comprenant ces classes) ou en intégrant explicitement cette dimension dans leurs modèles [4, 22].

Dans le second degré, la principale ambiguïté provient du fait que la taille des classes peut être assez éloignée du nombre d'élèves qui sont physiquement présents dans une salle de classe. Dire que la seconde C a une taille de 32 élèves ne signifie pas que chaque élève est, à chaque cours, entouré de 31 camarades. Tout d'abord, certains d'entre eux peuvent être absents, ou ont quitté l'établissement depuis le début de l'année (comme dans le premier degré). Ensuite, une partie des enseignements se fait en groupes : un élève peut se retrouver dans une salle avec 15 camarades pour suivre des travaux dirigés, un cours de langue... Au collège et à la rentrée 2011, 17 % des heures d'enseignement sont suivies en groupe, 47 % en lycée général ou technologique et

NOTES

1. Afsa [2], dans ce numéro, souligne la difficulté à transférer les résultats de la recherche des États-Unis à la France (en l'occurrence ayant trait à l'impact de la taille de l'établissement sur la réussite scolaire), et en appui à cet argument, présente en début d'article quelques-unes des principales différences entre les deux systèmes.

2. On inclut dans ce champ la recherche de Wößmann et West [23] sur les données des évaluations internationales TIMSS (cf. *infra*) incluant la France. Nous emploierons par commodité l'expression « travaux français ».

44 % en lycée professionnel [19].

Afin de donner une autre illustration de cette ambiguïté, nous pouvons nous appuyer sur le « nombre d'élèves par structure » (E/S) calculé par la DEPP. Il s'agit du nombre moyen d'élèves dont un enseignant a la charge pendant une heure dans une classe ou dans un groupe. Cet indicateur se rapproche donc davantage du nombre d'élèves qui sont physiquement présents, ensemble, dans une salle de classe. Au lycée général et technologique, la taille des classes (E/D) moyenne est de 28,2 élèves et la taille des structures (E/S) est de 24,1 (figure 1). On mesure ainsi l'écart entre l'indicateur « taille des classes » et le nombre d'élèves qui sont effectivement face à un enseignant. Or, dans l'hypothèse où c'est le nombre d'élèves physiquement présents dans une classe qui compterait pour la réussite scolaire, la taille des classes n'en fournirait qu'une mesure approchée (ou encore un majorant) dans le second degré. Cette difficulté se pose surtout à l'échelle du lycée : un moyen de la limiter est de ne considérer que la série générale et des matières moins susceptibles d'être enseignées en groupe [22].

Ces recherches sont confrontées à un autre problème de mesure, portant cette fois sur la variable à expliquer : qu'est-ce que la réussite scolaire ? Elles font souvent le choix

de variables mesurant l'acquisition et la maîtrise de connaissances par l'élève. Les aspects liés au développement personnel ou encore à la sociabilité de l'élève sont, à notre connaissance, plus rarement traités, rejoignant un constat déjà fait par Meuret [16]. Dans les travaux français recensés dans cet article, trois types de variables de résultats sont considérés :

- les résultats à des évaluations standardisées³ passées par les élèves en début et en fin d'année, en français et en mathématiques dans le premier degré [4, 5, 7, 22]. Une étude portant sur le second degré utilise une évaluation internationale en mathématiques et en sciences [23] ;

- le passage en classe supérieure [11] dans une étude portant sur le collège ;

- les notes aux examens, diplôme national du brevet (DNB) et baccalauréat [22].

Les études portant sur le premier degré s'appuient plus fréquemment sur des évaluations standardisées. Ce constat est en partie lié à la nature des informations disponibles dans les sources de données. En effet, les panels d'élèves utilisés dans les études portant sur le second degré (entrants en sixième 1989 et 1995), ne comprennent pas d'évaluation au cours du second degré, en dehors des évaluations nationales d'entrée en sixième.

Les évaluations standardisées sont censées fournir les mesures les plus directement liées au niveau d'acquisition de l'élève. Les notes aux examens en donnent une approche plus ou moins fine. Par exemple, Valdenaire [22] utilise à la fois les notes au contrôle continu et à l'examen final du DNB. Comme il le suggère, la seconde mesure est sans doute celle qui est la plus facilement interprétable en termes de niveau de l'élève⁴. En effet, les notes au contrôle continu d'un élève sont influencées par le niveau d'ensemble de sa classe : dans des classes de faible niveau, l'enseignant aura tendance à revoir ses exigences à la baisse, et à noter plus favorablement les élèves [17]. En fin de lycée, les notes obtenues au baccalauréat, utilisées également par Valdenaire, comprennent un aléa lié au correcteur, au sujet posé une année ou une autre, et ne sont pas forcément comparables d'une filière à une autre. Là encore, des solutions sont mises en œuvre pour tenir compte formellement de ces limites dans la modélisation, sans les effacer complètement.

Enfin, le passage en classe supérieure utilisé dans l'étude de Gary-

NOTES

3. Ces évaluations sont elles-mêmes de diverses natures. Valdenaire [22] a utilisé à la fois les évaluations spécifiques aux élèves du panel CP 1997 et les évaluations nationales à l'entrée au CE2. La DEP [7] et Bressoux et Lima [5] ont utilisé des évaluations spécifiques aux élèves de l'expérimentation « CP à effectifs réduits ». Bressoux *et al.* [4] ont utilisé une enquête ministérielle menée en 1991-1992 où les élèves passaient une évaluation spécifique en fin de CE2.

4. Nous partageons cet avis, et ce sont d'ailleurs les résultats appuyés sur l'examen final du DNB que nous présentons dans le tableau récapitulatif en annexe.

Figure 1 – Nombre d'élèves par division (« taille des classes », E/D) et nombre d'élèves par structure (E/S), rentrée 2012

	Nombre d'élèves par division	Nombre d'élèves par structure
Premier cycle	24,9	23,8
Second cycle professionnel	18,6	16,6
Second cycle général et technologique	28,2	24,1
Ensemble	24,2	22,1

Note : la ligne « Ensemble » comprend aussi les Segpa, STS, CPGE.

Champ : secteur public et privé sous contrat.

Sources : O'Prey [19].

Bobo et Mahjoub [11], de même que le redoublement, fournissent une information sur le niveau de l'élève, mais renvoient aussi à d'autres dimensions. Ce passage dépend en partie des stratégies des familles ainsi que de leur plus ou moins grande acceptation du redoublement [18]. La décision de redoublement dépend aussi du niveau de la classe et de l'appréciation de l'enseignant [6].

En abordant ces travaux, le lecteur doit ainsi garder à l'esprit que la façon de mesurer la taille des classes d'une part, les résultats scolaires d'autre part, ont une importance pour le lien qui est mesuré entre les deux. Pour un même niveau scolaire, les conclusions sur l'impact de la taille des classes sur les résultats scolaires peuvent différer suivant l'indicateur de résultats scolaires qui est regardé⁵. Ces mesures, et leurs interprétations, peuvent également contribuer à éclairer les résultats portant sur différents niveaux scolaires. Par exemple, le flou autour de la taille des classes au lycée (importance du travail en groupes) ainsi qu'une certaine imprécision liée à la mesure du niveau scolaire par les notes au baccalauréat, peuvent contribuer à la difficulté à dégager une relation significative entre taille des classes et réussite scolaire en lycée général⁶.

TAILLE DES CLASSES ET RÉUSSITE SCOLAIRE : COMMENT ISOLER UN LIEN DE CAUSALITÉ ?

Le deuxième type de problèmes auxquels ces travaux sont confrontés porte sur la façon de déterminer une relation causale allant de la taille

des classes vers la réussite scolaire. La corrélation observée, au niveau individuel, entre la taille de la classe d'un élève et ses résultats ne permet aucune déduction sur le rapport causal existant entre ces deux variables. Si le passage de la corrélation à la causalité pose un problème fréquent en évaluation, ce dernier se pose ici de façon particulièrement aigüe.

En effet, la répartition des élèves entre classes de tailles différentes n'est en général pas aléatoire. Elle est, entre autres, le produit de l'histoire scolaire et sociale de l'individu. Les élèves ayant eu les résultats les plus faibles et/ou originaires de milieux sociaux défavorisés se retrouvent plus souvent dans des petites classes, notamment pour des raisons institutionnelles. Le fait que les tailles de classes soient plus petites en éducation prioritaire y contribue [22]. Au sein d'un même établissement, le directeur d'école ou le chef d'établissement peuvent affecter les élèves les plus en difficulté dans des classes de plus petite taille. Enfin, un établissement renommé peut avoir un afflux de demandes d'élèves de milieux favorisés et y répondre par des classes de tailles plus élevées.

À ce problème de répartition non aléatoire des élèves, peut s'ajouter une répartition non aléatoire des enseignants entre classes de petite et de grande taille. Cela peut se produire, par exemple, si les enseignants novices sont plutôt orientés vers des classes de plus petite taille [4, 5]. Dans un cas comme dans l'autre, nous sommes ramenés au même problème : les écarts de résultats entre élèves des classes à petits effectifs et effectifs plus importants ne peuvent pas s'interpréter directement par un effet de la taille des classes, car les

deux ensembles d'élèves ne sont pas comparables.

Lorsqu'une telle comparaison « brute » entre élèves de classes à petits ou forts effectifs est quand même faite, elle mène fréquemment au constat suivant : les élèves ayant la meilleure réussite sont dans des classes à plus forts effectifs. Il faut y voir avant tout l'effet d'une affectation non aléatoire des élèves. Les informations disponibles sur le passé scolaire, les caractéristiques socio-démographiques des élèves et celles de l'établissement font disparaître cette corrélation, ou la rendent négative, lorsque les élèves sont rendus comparables selon ces caractéristiques, et en appliquant les méthodes économétriques les plus usuelles. Toutefois, la sélection des élèves décrite plus haut n'est alors qu'imparfaitement prise en compte, ne serait-ce que parce que les établissements ont des informations sur l'élève (comportement...) qui ne sont pas disponibles – ou le sont en partie seulement – dans les données statistiques. Or, ces facteurs inobservés, s'ils étaient connus et intégrés dans les calculs, pourraient jouer sur la corrélation ou sur l'importance de

NOTES

5. Parfois au sein d'une même étude, comme dans l'expérimentation « CP à effectifs réduits » (DEP, [7]). Les conclusions en fin de CP sur l'effet des classes de taille réduite y diffèrent selon qu'on regarde les résultats de l'évaluation standardisée (effet positif de la réduction d'effectifs) ou le taux de redoublement (pas d'effet). Les conclusions de la DEP privilégient les résultats portant sur les évaluations standardisées, de façon qui nous semble logique, au vu de ce qui précède.

6. Nous nous référons ici au résultat de Valdenaire [22] en étant bien conscients qu'il y a d'autres interprétations possibles à ce constat (*cf. infra*).

l'effet. Le signe de l'effet, ainsi que son ampleur, restent donc soumis à une forte incertitude.

DEUX TYPES DE MÉTHODES VISANT À SE RAPPROCHER LE PLUS POSSIBLE DES CONDITIONS EXPÉRIMENTALES

Afin de dépasser cette difficulté, deux grands types de méthodes ont été développés, dans des travaux de recherche menés aux États-Unis et en Israël.

Le premier type correspond à des expérimentations « contrôlées », où des classes sont concernées de façon aléatoire par une réduction du nombre d'élèves. Ceci doit permettre d'éviter que le fait de se retrouver dans une classe de taille réduite, pour un élève, soit associé à ses caractéristiques initiales. Il est souhaitable que cette affectation aléatoire concerne aussi les enseignants, afin d'éviter que les caractéristiques des enseignants diffèrent entre classes à petits et forts effectifs. Aux États-Unis, une telle expérimentation a été mise en œuvre dans le cadre du projet *Tennessee Student/Teacher Achievement Ratio* (STAR) (encadré 1).

Le second type est celui d'une expérience dite « naturelle » ou « quasi-expérience ». Il présente le point commun avec le précédent de chercher à comparer les élèves de classes de petite et de grande taille, en s'assurant que les élèves concernés sont les plus comparables possibles. Toutefois, à la différence du précédent, aucune expérimentation n'est mise en œuvre. Il faut chercher, dans les données existantes, des situations où ces comparaisons sont possibles, ce qui illustre l'expression

« expérience naturelle ». L'un des travaux les plus connus ayant utilisé une telle méthodologie sur la question de la taille des classes, appliquée aux écoles publiques d'Israël, est celui d'Angrist et Lavy [3] à partir de la « règle de Maïmonide » (encadré 2).

Soulignons d'emblée que les deux méthodes ont leurs avantages et leurs limites. Il est bien sûr intéressant qu'elles puissent être appliquées sur un même sujet, afin de confronter leurs résultats. Les travaux internationaux qui viennent d'être cités constituent un point d'entrée utile pour comprendre les orientations des travaux français ayant suivi le rapport Meuret [16], même si les méthodes mises en œuvre diffèrent parfois sensiblement.

LA MISE EN APPLICATION DE CES MÉTHODES DANS LE CAS FRANÇAIS (1) : L'EXPÉRIMENTATION « CP À EFFECTIFS RÉDUITS »

La première option méthodologique, l'expérimentation, a fait en France, à notre connaissance, l'objet d'une seule mise en application. Suivant une recommandation du rapport de Meuret [16], le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche l'a mise en œuvre [7]. 200 classes de CP ont été sélectionnées dans dix académies, accueillant principalement des enfants de familles défavorisées ou en difficulté scolaire. Il faut d'emblée souligner que la sélection des classes n'a pas été aléatoire : les classes participantes ont été proposées par les recteurs, en prenant notamment en compte les moyens disponibles localement pour permettre le dédoublement de classes. Il ne s'agit donc pas à proprement parler d'une expérimentation « contrôlée ».

Dans une moitié de ces classes, une réduction d'effectifs, conduisant à des classes comportant de 8 à 12 élèves, a été mise en place, l'autre moitié constituant les « classes témoins ». L'année suivante, les élèves de ces classes à effectifs réduits qui sont passés en CE1 ont tous été scolarisés dans des classes à effectifs habituels. Leurs résultats aux tests de langage et en mathématiques ont été comparés à ceux des autres classes en début, en cours, à la fin du CP, au début et à la fin du CE1. Un problème posé par cette expérimentation tient dans la forte réduction, au cours du temps, du nombre d'élèves et de classes suivis, en raison de la lassitude, d'absences et de mouvements sociaux. Ainsi, dans la publication de la DEP [7] portant sur les classes de CP et de CE1, les comparaisons se font entre 60 classes à effectifs réduits (sur 101 concernées au départ) et 53 classes à effectifs normaux (sur 99).

En fin de CP, les élèves ayant été dans une classe de taille réduite ont obtenu un score total (langage et mathématiques confondus) supérieur de 3,7 points sur 100 à ceux scolarisés dans des classes de CP à effectifs habituels⁷. Toutefois, cet avantage s'était considérablement érodé lorsque les mêmes épreuves ont été repassées, cette fois, en début du CE1 (1,8 point, non significativement différent de zéro). Enfin, il avait totalement disparu en fin de CE1 (figure 2).

Bressoux et Lima [5] ont procédé à un réexamen de ces données. Selon eux, la comparaison des effets moyens à laquelle procède la DEP ne

NOTE

7. Il n'est pas possible, compte tenu des données publiées, d'exprimer ces écarts en pourcentages d'écart-type des scores par élève supplémentaire.

Encadré 1 – Le projet Tennessee Student / Teacher Achievement Ratio (STAR)

Le projet STAR est une étude longitudinale menée aux États-Unis à partir de l'année scolaire 1985-1986. Il a mobilisé 11 600 élèves de 80 écoles publiques. La moitié des élèves sont rentrés au stade du *Kindergarten* (correspondant, en termes d'âge, à la grande section de maternelle), l'autre moitié entre le *1st Grade* (CP) et le *3rd Grade* (CE2), soit parce qu'ils ont rejoint une école participante à ce moment-là, soit parce qu'ils ont commencé leur scolarité en *1st Grade* (le *Kindergarten* n'étant pas obligatoire).

Au sein de chaque école participante, les enfants et les enseignants étaient répartis aléatoirement, la première année, entre trois types de classes : des classes à effectifs réduits (en principe 13-17 enfants), des classes à effectifs « normaux » (22-25 enfants) et des classes à effectifs normaux où un enseignant complémentaire aidait, à plein temps, l'enseignant déjà en place (les classes à effectifs normaux en bénéficiaient aussi, mais à temps partiel seulement). Les enfants devaient rester quatre années de suite dans le type de classe auquel ils avaient été assignés. À la fin de chaque année, les élèves passaient une batterie de tests standardisés.

De façon classique pour ce type d'expériences [10], il y a eu des écarts entre le processus tel qu'il avait été conçu et sa réalisation. Des élèves ont quitté l'école en cours d'expérience, et ces transferts ont concerné plus fréquemment les élèves qui étaient (ou auraient dû être) dans les classes à effectifs normaux. Il y a bien sûr plusieurs motifs à ces transferts, mais il est vraisemblable qu'une partie des parents, peu satisfaits que leur enfant se retrouve dans une classe à effectifs normaux, ont transféré celui-ci dans une autre école. D'autres parents ont réussi à faire en sorte que leur enfant soit transféré vers une classe à effectifs réduits. Enfin, pour les enfants qui étaient dans des classes à effectifs normaux, une nouvelle répartition aléatoire a été organisée en début de CP pour les répartir entre classes bénéficiant ou non d'un enseignant complémentaire à plein temps, en réponse aux contestations émanant des parents sur l'équité du dispositif.

Chacune de ces trois situations, comme le souligne Krueger [12], pose problème au regard du principe de l'expérience. Les deux premières le font en rompant le principe de répartition aléatoire des enfants entre les types de classe. La troisième le fait parce qu'elle introduit un nouveau paramètre qui différencie les élèves des classes à effectifs réduits des autres : ces élèves ont, par construction, des camarades de classe qui changent moins au cours du temps. Ce paramètre peut influencer sur leurs résultats.

Ces déviations exigent de tenir compte, par l'économétrie, des différences de structure qui sont apparues de façon indue entre les types de classes. Krueger, par exemple, met ainsi en évidence que les enfants dans des classes à effectifs réduits ont obtenu des résultats meilleurs que les enfants dans des classes à effectifs normaux : de 19 % à 28 % d'écart-type sur les scores en fin d'année selon le niveau considéré (voir tableau). L'élève tire surtout parti de la classe à effectifs réduits la première année où il s'y trouve, l'écart avec les élèves des autres classes augmente encore, mais plus faiblement, au cours des années suivantes. Cet effet est plus élevé pour les élèves noirs et ceux dont les familles ont des revenus modestes. Enfin, les classes avec un enseignant supplémentaire à temps plein ne présentent pas d'avantage marqué sur les autres classes à effectifs « normaux ».

Effet associé à la scolarisation dans une classe à effectifs réduits dans le dispositif STAR, langage et mathématiques

	<i>Kindergarten</i> (grande section maternelle)	<i>1st Grade</i> (CP)	<i>2nd Grade</i> (CE1)	<i>3rd Grade</i> (CE2)
En points de score (sur 100)	5,4	7,4	5,8	5,0
En % d'écart-type de la distribution des scores (élèves)	20,1	27,5	21,8	18,5

Lecture : les élèves en classe de CP à effectifs réduits dans STAR ont un score en langage et en mathématiques en fin de CP supérieur de 7,4 points à celui des élèves de classes à effectifs normaux. Cet écart représente 28 % d'écart-type de la distribution des scores des élèves en fin de CP.

Champ : élèves des écoles publiques concernées par le dispositif STAR.

Sources : Krueger [12].

peut être menée ainsi. En particulier, les conditions d'expérimentation n'ont pas permis que les élèves soient comparables du point de vue des caractéristiques de leurs enseignants. Les élèves en classe réduite avaient des enseignants à l'ancienneté plus faible dans la profession, et leur ancienneté en CP était également plus faible⁸. Cette situation pourrait s'expliquer par l'affectation non aléatoire des ensei-

gnants déjà évoquée, les enseignants plus novices ayant été plutôt affectés dans des classes à effectifs réduits. Bressoux et Lima intègrent cet effet parmi leurs variables explicatives, ainsi que le score initial des élèves et des effets aléatoires propres aux classes. Ils suggèrent finalement un effet associé à un élève en plus ou en moins dans une classe de 2,0 % ou 2,2 % d'écart-type, suivant la spécification retenue.

NOTE

8. Les auteurs de la *Note Évaluation* de la DEP [7] reconnaissent ce fait, en faisant remarquer que les enseignants de classes à effectifs réduits avaient aussi « plus fréquemment suivi une action de formation continue sur le thème de la lecture au cours des trois dernières années (25 % contre 18 %) » et avaient été « plus impliqués au cours de l'année dans des actions péri-scolaires touchant à la lecture ».

Encadré 2 – La « règle de Maïmonide » et son application à l'effet d'une réduction de la taille des classes

Au XII^e siècle, le rabbin Maïmonide proposa une règle selon laquelle la taille d'une classe ne devrait pas dépasser 40 élèves. À partir de 1969, cette règle est utilisée pour définir les seuils d'ouverture et de fermeture des classes dans les écoles publiques en Israël. Pour l'illustrer, reprenons tel quel l'exemple donné par Angrist et Lavy [3] d'une école ayant 80 élèves dans un niveau. Elle dispose alors de deux classes de 40 élèves. Si un nouvel élève arrive dans l'école, faisant passer l'effectif à 81, la règle de Maïmonide implique de remplacer les deux classes existantes par trois classes de 27 élèves chacune.

Cette situation est celle d'une expérience dite « naturelle » ou « quasi-expérience ». La règle institutionnelle produit une différence dans la taille des classes entre les établissements qui est *a priori* sans lien avec le niveau de leurs élèves. Cette variation peut alors être utilisée pour comparer les performances des classes de différentes tailles. Plus précisément, connaissant le nombre d'élèves de ce niveau dans l'école (81 dans l'exemple précédent), la règle de Maïmonide permet de déduire un nombre d'élèves par classe théorique (27). Ce nombre théorique est utilisé comme variable instrumentale pour déterminer les effets d'une taille des classes plus ou moins élevée sur les performances de ces classes. Dans une version de la méthode, seules les écoles les plus proches des seuils sont utilisées pour l'estimation (en incluant par exemple les écoles comportant de 76 à 85 élèves, mais pas celles ayant 70 ou 90 élèves), afin de renforcer encore leur comparabilité (voir tableau).

Effet d'un élève supplémentaire sur les résultats de la classe en lecture

	4th Grade (CM1)		5th Grade (CM2)	
	Échantillon complet	Échantillon restreint	Échantillon complet	Échantillon restreint
En points de score (sur 100)	0,13	0,15	0,28	0,58
En % d'écart-type de la distribution des scores (classes)	1,7	1,9	3,6	7,1

Lecture : un élève supplémentaire dans une classe de CM2 réduit le score en lecture de cette classe de 0,28 point ou 0,58 point sur 100, selon que l'estimation est menée sur l'ensemble de l'échantillon ou sur les seules écoles proches des seuils d'ouverture / fermeture de classe. Ces scores correspondent à une baisse de 3,6 % et 7,1 % en termes d'écart-types des scores moyens des classes.

Note : les résultats en pourcentages d'écart-type font référence à la distribution des scores moyens des classes. Ils ne sont donc pas directement comparables aux résultats des autres études citées, qui se réfèrent à la distribution des scores ou notes des élèves.

Champ : écoles élémentaires publiques d'Israël, année scolaire 1990-1991.

Sources : Angrist et Lavy [3].

Pour que cette méthode soit valide, le nombre théorique ainsi défini doit être un bon « prédicteur » du nombre d'élèves par classe constaté : cela implique que la règle de Maïmonide, ou toute autre règle qui joue ce rôle, soit appliquée par les responsables éducatifs de façon suffisamment stricte. Une autre condition de validité est que ce nombre théorique ne soit pas directement lié au niveau scolaire de l'élève. Son impact sur la réussite scolaire ne doit provenir que du fait qu'il induit une taille de classes plus ou moins élevée.

Sur l'année scolaire 1990-1991, et pour l'équivalent du CM2 français dans les écoles publiques d'Israël, une taille de classes réduite accroît les performances de ces classes en mathématiques et en lecture. En lecture, l'effet associé à un élève en moins dans la classe est estimé à 3,6 ou 7,1 % d'écart-type suivant la variante de la méthode qui est utilisée (la distribution de référence étant celle du score moyen des classes). En CM1, un effet est encore perceptible, mais plus faible : 1,7 ou 1,9 %. Pour ces deux classes, les effets sont plus forts dans les écoles comprenant une forte part d'élèves issus de milieux défavorisés. Pour l'équivalent du CE2 et sur l'année scolaire 1991-1992, les effets ne sont pas significatifs. Angrist et Lavy alertent toutefois sur la difficulté à comparer les tests des deux années, les conditions dans lesquelles ils se sont déroulés différant sous plusieurs aspects.

Figure 2 – Synthèse des résultats de l'expérimentation « CP à effectifs réduits »

	Score moyen sur 100			Significativité (au seuil de 5 %)
	Effectifs réduits	Effectifs habituels	Écart	
Début de CP	63,8	62,8	1,0	non
Fin de CP	78,0	74,3	3,7	oui
Début de CE1	79,1	77,3	1,8	non
Fin de CE1	84,7	84,9	- 0,2	non

Sources : MEN-DEP [7].

Enfin, Ecalle *et al.* [9] ont regardé si les effets associés à une taille de classe réduite, en termes de résultats en fin de CP, se

distinguaient suivant les caractéristiques sociodémographiques et le passé scolaire de l'élève. Leur étude s'appuie sur une partie des

épreuves de langage (reconnaissance et écriture de mots). Selon eux, l'effet positif associé à une classe à effectifs réduits ne bénéficie pas, ou guère, aux élèves des catégories les moins favorisées socialement. Pour l'illustrer, les auteurs classent les élèves en quatre groupes d'origine sociale. L'avantage des classes à effectifs réduits n'est perceptible ni pour les origines sociales les plus

favorisées, ni pour les plus défavorisées⁹. Il n'apparaît que pour des catégories « moyennes » ou « intermédiaires ». De plus, les enfants qui ne parlent pas le français à la maison ne semblent pas bénéficier de tailles de classes plus réduites. Or, ces élèves ont aussi des résultats scolaires moins bons que les autres élèves. Dans ces deux situations, la baisse des effectifs dans les classes ne semble donc pas aller dans le sens d'une réduction des inégalités scolaires.

Ces études ont été confrontées à trois limites liées à l'expérimentation. La première, déjà citée, est le caractère non aléatoire des classes sélectionnées. La deuxième, également signalée, tient à la forte baisse du nombre d'élèves et de classes suivis au cours du temps, qui conduit à faire les calculs sur un échantillon restreint et ayant peut-être des caractéristiques spécifiques. La troisième tient dans le processus très particulier de l'expérience, qui conduisait à remettre les élèves en classe à effectifs « normaux » dès l'année suivante, en CE1. On peut se demander si les élèves de classes à effectifs réduits ont pu véritablement tirer parti de cette situation, sur un temps très court, et si le retour soudain à une classe deux fois plus chargée (en moyenne) n'a pas été pénalisant. Les enseignants, de leur côté, n'ont eu qu'un an pour adapter leurs pratiques à cette forte réduction d'effectifs, avant d'arrêter là l'expérience. L'observation de leurs pratiques suggère une évolution limitée, et très diverse d'un enseignant à un autre, les enseignants ayant été laissés autonomes en la matière [7]. Un retour vers des classes à effectifs normaux différé de plusieurs années (comme dans STAR) et/ou plus progressif pourrait donner des résultats

différents. De façon plus générale, l'expérimentation menée par la DEP soulève une limite classique pour ce type de méthodes : parce qu'elle s'est faite dans un contexte particulier, elle pose la question de la généralisation des résultats.

LA MISE EN APPLICATION DE CES MÉTHODES DANS LE CAS FRANÇAIS (2) : LES MÉTHODES QUASI EXPÉRIMENTALES

La seconde option méthodologique s'appuie sur des données regroupant l'ensemble des élèves d'un niveau scolaire, représentatifs des élèves français. Dans ces données, et puisque nous ne sommes plus dans le cadre expérimental, les élèves des classes à effectifs réduits diffèrent notablement des élèves des classes plus chargées. Ces méthodes doivent donc mobiliser de façon poussée l'économétrie pour rendre les élèves de différentes tailles de classe les plus comparables possibles. Comme nous l'avons dit plus haut, on peut considérer que ce type de méthodes construit des « quasi-expériences » en s'inspirant notamment de l'approche proposée par Angrist et Lavy [3].

Les travaux de Piketty [20], Piketty et Valdenaire [21] et Valdenaire [22] sont ceux qui ont appliqué de la façon la plus fidèle la méthode d'Angrist et Lavy¹⁰. Rappelons, de façon très sommaire, qu'il s'agit de trouver un seuil, exprimé en termes de nombre d'élèves dans un niveau, tel qu'un établissement qui dépasse ce seuil se trouve dans l'obligation de créer une nouvelle classe. Un établissement qui a dépassé ce seuil bénéficie ainsi de classes moins chargées, un établissement qui est resté en deçà présente

des classes à plus forts effectifs. Comparer les élèves de ces deux types d'établissement permet d'estimer l'effet de la taille des classes sur la réussite scolaire.

Les travaux de Piketty et Valdenaire, menés à partir du panel d'entrants au CP en 1997, présentent l'intérêt de calculer selon une méthode homogène les effets de taille de classe à plusieurs niveaux d'enseignement, à partir de différentes variables de résultats scolaires : sur les résultats d'évaluations à l'entrée au CE2, les notes à l'examen du brevet des collèges, et les notes au baccalauréat. En France, pour le premier degré et le collège, il n'existe pas de seuil défini de manière légale qui fixe, partout, la taille maximale des classes. Valdenaire [22] a déterminé ces seuils de façon empirique : ce sont ceux qui permettent le mieux d'expliquer la taille effectivement observée dans les établissements. Au lycée, toutefois, un nombre

NOTES

9. Le contenu détaillé de ces catégories n'est pas donné dans l'article, mais il semble que les auteurs aient isolé, en bas de la hiérarchie sociale, un groupe très restreint : parents sans emploi, très peu qualifiés (les auteurs citant l'exemple des agents d'entretien). À titre de comparaison, les catégories « défavorisées » dans le travail de Valdenaire [22] sont beaucoup plus larges, englobant les employés, les ouvriers, et les personnes n'ayant jamais travaillé.

10. Piketty [20] a réalisé des estimations sur le premier degré, complétées pour le second degré par Piketty et Valdenaire [21]. Enfin Valdenaire, dans sa thèse [22], a quelque peu revu la méthode utilisée sur le premier degré. Ce sont ses résultats, les plus récents, que nous présentons par la suite, sachant qu'ils ne remettent pas en question, sur le fond, les résultats des travaux précédents.

d'élèves maximal a été défini en 1989¹¹ à 35 élèves, et a été utilisé pour l'étude.

Gary-Bobo et Mahjoub [11] s'inspirent également de la méthode d'Angrist et Lavy en l'adaptant quelque peu. Ils étudient l'effet d'une variation de la taille des classes sur la probabilité de passer en classe supérieure (plutôt que de redoubler ou d'être orienté vers la voie professionnelle) au collège, à partir du panel d'entrants en sixième de 1989. L'instrument d'Angrist et Lavy, ici défini avec un seuil d'ouverture de classes à 28 élèves, pose pour ces auteurs le problème de ne pas être très opérationnel pour des collèges ayant plus de 100 élèves par niveau (c'est-à-dire la majorité des collèges). Leur idée est d'utiliser également comme variable instrumentale le nombre d'élèves par niveau, tout en intégrant dans les variables de contrôle le nombre d'élèves total dans l'établissement¹². Les auteurs estiment un modèle à variables instrumentales où la variable expliquée est le fait de passer en classe supérieure en fin de sixième, cinquième, quatrième, troisième. Ils estiment également un système d'équations modélisant ensemble les tailles de classes à chaque niveau de collège (sous la forme d'équations linéaires) et les transitions correspondantes en fin d'année (sous la forme de *probits* ordonnés)¹³. Ces transitions revêtent trois situations possibles : passage en classe supérieure, redoublement, ou transition vers des classes spécialisées ou à vocation professionnelle.

Bressoux *et al.* [4] s'éloignent un peu plus de la méthode d'Angrist et Lavy, tout en restant dans le principe d'une « quasi-expérience ». Cet article porte, à l'origine, sur l'effet de la formation initiale pour les ensei-

gnants de CE2, en s'appuyant sur une enquête ministérielle menée au cours de l'année scolaire 1991-1992 sur les élèves de CE2 et leurs enseignants dans douze départements. Cette année-là, une erreur de prévision particulièrement forte sur les besoins en enseignants a conduit à rappeler des enseignants sur liste d'attente au concours, pour les envoyer enseigner sans qu'ils aient suivi les deux années réglementaires de formation. Suite à cette situation, des enseignants formés ont côtoyé des enseignants non formés, ce qui a permis de fournir une évaluation de l'impact de la formation réglementaire de deux ans.

Cette configuration a également permis aux auteurs de proposer une mesure de l'effet de la taille des classes sur les performances scolaires. Rappelons que l'une des principales difficultés pour identifier cet effet est que les élèves aux performances les plus faibles tendent à se retrouver dans les classes à petits effectifs, ne permettant pas une comparaison simple entre les classes à effectifs réduits et les autres. Toutefois, les auteurs constatent que sur l'échantillon formé par les élèves des professeurs débutants (qu'ils aient ou non suivi une formation), les élèves ayant les résultats les plus faibles en début de CE2 ne sont pas forcément dans les classes à plus faibles effectifs. Sur cet échantillon particulier, il leur apparaît donc que les élèves des classes à petits et forts effectifs peuvent être comparés, à partir d'un modèle de régression linéaire.

Enfin, Wößmann et West [23], avec une méthodologie encore différente, proposent une estimation de l'effet de la taille des classes de cinquième et quatrième, à partir d'une évaluation standardisée

qui permet aussi une mise en perspective internationale : l'évaluation *Third International Mathematics and Science Study* (TIMSS) menée en 1994-1995. Comme les travaux précédents, les auteurs souhaitent traiter le problème de la répartition non aléatoire des élèves entre classes de petite et de grande tailles. Ils décomposent ce problème en deux parties : la première correspond au fait que les élèves aux niveaux scolaires plus faibles se retrouvent en général dans les établissements où les classes sont de petite taille. La seconde est, au sein des établissements, le fait que ces élèves sont souvent affectés dans les classes à petits effectifs. Le premier type de problème est traité en introduisant des effets « établissements » dans les estimations¹⁴. Le second type de problème est traité en utilisant comme variable instrumentale (pour estimer la taille de chaque classe) la taille moyenne des classes de ce niveau dans l'établissement. En résumé, les auteurs proposent un

NOTES

11. Il est issu d'un rapport annexé à la loi d'orientation sur l'éducation du 10 juillet 1989.

12. S'ils ne mettaient pas l'effectif d'élèves dans l'établissement parmi les variables explicatives, l'emploi du nombre d'élèves par niveau comme instrument pourrait être critiqué. En effet, ce nombre d'élèves serait alors susceptible d'être corrélé à la performance des élèves.

13. Dans ce type de modèles, la variable à expliquer est qualitative et ses modalités ont la particularité de pouvoir être ordonnées. Ici, les trois modalités sont assimilées au résultat de l'élève, du meilleur au moins bon résultat : passage en classe supérieure, redoublement, ou transition vers des classes spécialisées ou à vocation professionnelle.

14. Cela est permis par le fait que TIMSS recueille des résultats sur deux classes consécutives.

effet de la taille des classes sur les scores aux évaluations en s'appuyant sur le fait que, au sein d'un même établissement, la taille moyenne des classes diffère entre la cinquième et la quatrième pour une année scolaire donnée. Cette variation-là est supposée liée à des contraintes d'ordre démographique (l'établissement recevant des cohortes d'élèves plus ou moins nombreuses) et sans lien direct avec les résultats des élèves.

Compte tenu des hypothèses à faire dans ce type de travaux, les conclusions doivent être prises avec précaution. Rappelons d'abord que ces travaux cherchent à isoler des situations où la variation de la taille des classes est présumée, et présumée seulement, intervenir sans lien avec les caractéristiques des élèves. Ensuite, leurs résultats, précisément parce qu'ils ont été obtenus à partir de ces situations spécifiques – par exemple pour la méthode d'Angrist et Lavy, celles où une variation de la demande de scolarisation dans cette école fait passer au-delà (ou en deçà) du seuil d'ouverture de classe – posent la question de leur généralisation, et donc de leur transposition à une politique globale de baisse de taille des classes. On retrouve, sous une forme différente, la limite que nous avons relevée pour l'expérimentation. Enfin, leurs résultats sont à nouveau soumis à une imprécision assez grande, d'autant que les méthodes par variables instrumentales ont tendance à augmenter l'amplitude des intervalles de confiance.

Malgré ces limites, toutefois, un grand intérêt de ces études est d'aboutir, avec des méthodes différentes, à des éléments convergents quant à l'existence d'un effet de la taille des classes sur la réussite scolaire, rejoignant

les résultats déjà obtenus aux États-Unis à partir de l'expérimentation STAR et sur données israéliennes par Angrist et Lavy [3].

UNE RÉDUCTION DE LA TAILLE DES CLASSES SERAIT PARTICULIÈREMENT FAVORABLE AUX ÉLÈVES SCOLARISÉS DANS LE PREMIER DEGRÉ ET EN DÉBUT DE COLLÈGE, MAIS N'AURAIT PAS D'EFFET AU LYCÉE GÉNÉRAL

Dans toutes ces études, une taille des classes plus élevée implique une dégradation des résultats scolaires, tout du moins pour l'année en cours et pour ce qui est du premier degré et du collège. Pour Valdenaire [22], un élève supplémentaire dans une classe de CE1 fait diminuer le score moyen aux évaluations de début de CE2 (mathématiques et français) de 2,5 à 3 % d'écart-type. Cet effet négatif dans le premier degré est conforté par Bressoux *et al.* [4]. Un élève supplémentaire dans une classe de CE2 fait à nouveau diminuer de 2,5 à 3 % d'écart-type son score aux évaluations passées en fin d'année en lecture et en mathématiques.

Au collège, des différences selon la taille de la classe sont encore apparentes au moins en début de cycle. Selon Gary-Bobo et Mahjoub [11], en classe de sixième, une hausse de la taille de la classe d'un élève diminue de 0,4 à 1,4 point (selon la méthode retenue) la probabilité de passer en classe supérieure. En cinquième, l'effet associé à un élève supplémentaire est encore négatif, de 0,3 point à 2 points sur la probabilité de passer dans la classe supérieure. Enfin, plus

aucun effet n'est apparent en classes de quatrième et de troisième.

Wößmann et West [23] estiment qu'un élève supplémentaire conduit, en France, à diminuer significativement le score en mathématiques¹⁵, en cinquième et quatrième, mais ne trouvent pas d'effet significatif sur les sciences. L'existence d'un effet significatif est ici d'autant plus remarquable que la France fait partie des quatre seuls pays, sur onze étudiés, pour lesquels un effet négatif et significatif est relevé (avec l'Islande en mathématiques, la Grèce et l'Espagne en sciences), cet effet n'étant sinon jamais significatif (à l'exception du Portugal où il est positif en mathématiques).

En fin de collège et en voie générale du lycée, les résultats disponibles ne laissent apparaître qu'un effet très modéré ou inexistant. Ainsi, Valdenaire [22] évalue à 1,3 % d'écart-type l'impact négatif d'un élève supplémentaire dans la classe sur la note moyenne à l'examen terminal du brevet des collèges. Enfin, le même exercice refait, cette fois, au lycée sur les notes au baccalauréat général, ne fait plus apparaître d'écart significatif suivant la taille des classes.

Comment interpréter le fait que l'influence de la taille des classes sur la réussite scolaire diminue au cours de la scolarité ? Cette diminution, au vu de ces travaux, interviendrait surtout vers la fin du collège et

NOTE

15. À l'échelle d'un pays donné, l'estimation de l'effet est, dans cette étude, peu précise. Elle n'est pas commentée en tant que telle par les auteurs. Ces derniers préfèrent mettre l'accent sur des groupes de pays, en mathématiques ou en sciences, pour lesquels un effet significatif peut être relevé.

en début de lycée (en tout cas pour le lycée général, les analyses ne portant pas sur le lycée technologique ou professionnel). Valdenaire [22] et Gary-Bobo et Mahjoub [11] ont cité deux pistes d'explication¹⁶. La première tient au fait que les élèves sont sélectionnés tout au long du parcours. Les élèves les plus en difficulté, qui seraient les plus susceptibles de tirer parti d'une taille de classes réduite, sortent progressivement du système scolaire ou sont orientés vers la voie professionnelle. La seconde tient à l'évolution des relations entre enseignants et élèves. En particulier, les problèmes de discipline pourraient être moins prégnants à partir de la fin du collège. Si l'on croit qu'une taille des classes réduite permet un meilleur maintien de la discipline, l'intérêt de cette réduction serait atténué pour des lycéens. Pour cette raison également, la situation du lycée professionnel mériterait sans doute des investigations, malgré les difficultés liées à la comparaison de filières différentes. Le lycée professionnel se caractérise en effet par des élèves de niveaux plus faibles et, aussi, des problèmes de discipline plus fréquents, comme en atteste la proportion plus forte d'élèves absentéistes ou encore d'incidents graves déclarés [15]¹⁷.

LA RÉDUCTION DE LA TAILLE DES CLASSES BÉNÉFICIERAIT D'AVANTAGE AUX ÉLÈVES ISSUS DE MILIEUX DÉFAVORISÉS, EN DIFFICULTÉ SCOLAIRE OU SCOLARISÉS DANS L'ÉDUCATION PRIORITAIRE

Dans les travaux qui viennent

d'être cités, de nature « quasi expérimentale », la dégradation des résultats scolaires associée à une hausse du nombre d'élèves par classe apparaît également plus marquée pour les élèves issus de milieux défavorisés et de faible niveau scolaire, et particulièrement forte pour les élèves scolarisés en éducation prioritaire. En zone d'éducation prioritaire (ZEP au moment de la réalisation des études précitées), l'impact négatif d'un élève supplémentaire en CE1 atteindrait 10 % d'écart-type des scores [22], contre 2 % d'écart-type hors ZEP. Bressoux *et al.* [4] obtiennent un effet similaire en mathématiques (8 % d'écart-type en ZEP, 3 % pour l'ensemble des élèves). En matière de notes à l'examen terminal du brevet, l'écart serait encore de 4 % d'écart-type en ZEP contre 1 % hors ZEP [22]. Enfin, le même exercice n'a pas pu être reproduit au lycée, les lycées en ZEP étant trop peu nombreux. Valdenaire [22] a toutefois mené séparément les estimations pour les enfants de catégories favorisées¹⁸ et défavorisées. Dans un cas comme dans l'autre, la taille des classes n'est pas associée à des notes au baccalauréat plus ou moins bonnes, confortant le résultat sur l'ensemble des élèves.

Le résultat portant sur les ZEP conduit Valdenaire [22] à proposer des simulations d'impact d'une réduction de la taille des classes ciblée sur les ZEP, à moyens constants au niveau global. Ces simulations incluent donc une hausse de la taille des classes dans les autres écoles et collèges, de 1,4 élève pour les écoles (CE1) et 1 élève pour les collèges (troisième), pour compenser une réduction de la taille des classes de 5 élèves en ZEP. L'inégalité initiale entre ZEP et non-ZEP serait réduite de 37 %

au primaire et de 28 % au collège, pourcentages qui constituent pour l'auteur des seuils minimum.

Pour expliquer cet avantage plus fort pour les enfants issus de milieux défavorisés ou en éducation prioritaire, on peut à nouveau envisager au moins deux pistes. Tout d'abord, ces élèves comptent parmi ceux qui manquent le plus de ressources matérielles et culturelles, et il est plausible que pour eux, toute amélioration des conditions d'accueil, et plus généralement des services offerts par l'école, revête un caractère plus crucial. Cet argument n'est pas spécifique à la question de la taille des classes et pourrait s'appliquer à d'autres situations où une politique publique s'avèrerait plus efficace pour des enfants de milieux défavorisés. Une explication plus spécifique a trait, à nouveau, au maintien de la discipline. Dans les classes qui regroupent à la fois des enfants en difficulté sociale et scolaire, et qui en plus ont de forts effectifs, il y a plus de chances de trouver des élèves qui perturbent la classe [14]. Une réduction de la taille des classes serait donc, dans

NOTES

16. Une autre explication, déjà citée, peut provenir du fait que la pertinence de la structure « classe », et donc du nombre d'élèves par classe, diminue au cours de la scolarité. Les élèves suivent en effet leurs cours de moins en moins souvent en classe entière, et de plus en plus en groupes (*cf. supra*).

17. Il s'agit de l'indicateur 17 de *L'état de l'École*.

18. Valdenaire définit les catégories favorisées ainsi : cadres, professions intermédiaires, indépendants. Les catégories défavorisées sont constituées par les employés, ouvriers, chômeurs ou inactifs n'ayant jamais travaillé.

ce contexte, plus bénéfique.

Deux nuances importantes sont à apporter à ce constat, à partir de l'expérimentation « CP à effectifs réduits ». Tout d'abord, les effets mesurés à partir de cette expérimentation peuvent sembler modérés, au regard du caractère socialement et scolairement défavorisé des écoles concernées¹⁹. Ensuite, dans les classes concernées par l'expérimentation « CP à effectifs réduits », la réduction du nombre d'élèves n'a pas profité aux élèves les plus défavorisés socialement [9]. Ces résultats vont à première vue à l'encontre de ceux obtenus à partir de méthodes quasi expérimentales, où les élèves les moins favorisés sont ceux qui bénéficient le plus d'une réduction de la taille des classes [4, 22]. En essayant, malgré les écarts de méthodes et de niveaux scolaires étudiés, de concilier ces résultats, on peut avancer l'interprétation suivante. La réduction de la taille des classes semble bénéficier aux élèves scolarisés dans des établissements et classes rencontrant le plus de difficultés. Toutefois, au sein de ces classes, ce ne seraient pas forcément les élèves les plus défavorisés ou en difficulté qui en bénéficieraient, mais plutôt des élèves aux caractéristiques « moyennes » ou aux difficultés modérées²⁰. Finalement, une lecture possible du résultat d'Écalle *et al.* [9] est qu'une réduction de la taille des classes n'est pas forcément l'outil le plus approprié pour des élèves rencontrant des difficultés sociales particulièrement prononcées. Dans un registre différent, l'absence d'effet, sur les résultats aux tests de langage en fin de CP, chez les élèves ne parlant pas français à la maison, suggérerait là aussi qu'une réduction de la taille des classes ne répond pas forcément aux besoins de tous les types d'élèves.

LES TRAVAUX FRANÇAIS RÉCENTS ONT ATTEINT UN CERTAIN CONSENSUS QUANT À L'EXISTENCE D'UN EFFET DE LA TAILLE DES CLASSES SUR LA RÉUSSITE SCOLAIRE

Depuis le rapport de Meuret [16], un consensus semble ainsi s'être dégagé, dans les travaux sur données françaises, quant au fait qu'une réduction du nombre d'élèves par classe permet une amélioration sensible des résultats scolaires, tout du moins pour l'année scolaire en cours, dans le premier degré et au collège. Cet effet apparaît, en général, plus élevé pour les élèves défavorisés ou en difficulté scolaire, et pour les élèves scolarisés en éducation prioritaire.

Il faut toutefois rappeler que le lien entre taille des classes et réussite scolaire reste complexe, notamment parce que les mécanismes sous-jacents n'ont pas été clairement identifiés. Par quels canaux passe l'effet d'une réduction de la taille des classes ? Est-ce en augmentant le temps consacré à l'élève par l'enseignant, en améliorant la communication avec les parents, ou encore parce qu'il serait plus facile d'assurer la discipline dans des classes plus petites ? De plus, pour les raisons qui ont déjà été avancées, les méthodes utilisées ne permettent pas d'anticiper avec une grande précision les effets des politiques. L'une des principales interrogations en suspens concerne le caractère cumulatif et durable de ces politiques. Une première question est de savoir si les élèves tirent ou non parti du fait de passer plusieurs années en classe à effectifs réduits. À titre d'illustration, Krueger [12] a

mené une telle analyse à partir des données de STAR. L'élève tire surtout parti de la classe à effectifs réduits la première année où il s'y trouve ; l'écart avec les élèves des autres classes augmente encore, mais plus faiblement, au cours des années suivantes. Une deuxième question porte sur le devenir scolaire des élèves après qu'ils ont quitté les classes à effectifs réduits. Le seul résultat sur données françaises est celui de la DEP [7], mais il se situait dans un cadre très particulier (les élèves en CP à effectifs réduits revenant dès l'année suivante dans une classe à effectifs habituels). Une analyse à plus long terme pourrait être engagée, pour étudier l'impact d'avoir fréquenté des classes à effectifs réduits, en cours de scolarité, sur le bilan à la fin de la scolarité, voire sur l'insertion professionnelle.

Aux États-Unis et à partir là encore des données de l'expérimentation STAR, Dynarski *et al.* [8] mettent en évidence de tels effets, significatifs et positifs. Certes, ces effets ne sont pas toujours élevés, par exemple pour la probabilité d'obtenir un diplôme du supérieur (+ 1,6 point pour les enfants ayant été en classe à effectifs réduits). Toutefois, on constate à long terme que la réduction de la taille des classes est associée également à une baisse des inégalités, les catégories les plus défavorisées du point de vue

NOTES

19. Rappelons que Bressoux et Lima [5] obtiennent un effet de l'ordre de 2 % d'écart-type en moins pour un élève en plus dans la classe.

20. Pour mémoire (*cf. supra*), la catégorie des élèves les plus « défavorisés », du point de vue du milieu social, est définie de façon différente d'une publication à l'autre. Cette catégorie est plus large chez Valdenaire [22] que chez Écalle *et al.* [9], ce qui peut aussi contribuer à la divergence des résultats.

scolaire étant celles pour lesquelles l'effet est le plus fort. De telles recherches pourraient être envisagées à partir des données françaises de panels d'élèves. Enfin, les questions du coût de la mise en œuvre, au regard

des autres politiques publiques d'éducation, demanderaient également de nouvelles investigations, ainsi que la combinaison possible avec d'autres politiques. Selon Ecalle *et al.* [9], une politique de réduction de la taille

des classes ne sera par exemple pleinement efficace que si les enseignants bénéficient d'une formation qui leur permet d'adapter au mieux leurs pratiques pédagogiques à ce nouveau contexte. ■

Annexe – Récapitulatif des études récentes sur données françaises

Résultats présentés sous la forme « un élève supplémentaire dans la classe = ... » sauf précision contraire. Les écarts-types correspondent à la distribution des notes ou scores des élèves

	Primaire			Collège				Lycée		
	CP	CE1	CE2	6 ^e	5 ^e	4 ^e	3 ^e	1 ^{re}	Terminale	
Ensemble des élèves		2,5 à 3 % d'écart-type de baisse des résultats aux tests à l'entrée du CE2 (français et mathématiques)	2,5 à 3 % d'écart-type de baisse des résultats aux tests à la fin du CE2 (lecture et mathématiques)	0,4 à 1,4 point de baisse dans la probabilité de passer en 5 ^e	0,3 à 2 points de baisse dans la probabilité de passer en 4 ^e générale	Pas d'effet significatif sur la probabilité de passer en 3 ^e générale	Pas d'effet significatif sur la probabilité de passer en 2 ^{es} générale ou technologique. 1,3 % d'écart-type en moins sur la note moyenne à l'examen terminal du brevet	Pas d'effet significatif sur la note au baccalauréat général à l'épreuve anticipée de français	Pas d'effet significatif sur la note moyenne aux épreuves de terminale du baccalauréat général	
Élèves les plus faibles-défavorisés	Classes de CP d'écoles à faibles niveaux scolaires : un effet significatif en fin de CP (langage et mathématiques) mais pas d'écart significatif en fin de CE1	Classes de CP d'écoles à faibles niveaux scolaires : pas d'effet significatif pour les élèves ayant les origines sociales les plus défavorisées (langage)	En éducation prioritaire (ZEP), 10 % d'écart-type de baisse des résultats aux tests à l'entrée du CE2	En éducation prioritaire (ZEP), 5 % (lecture) et 8 % (mathématiques) d'écart-type de baisse des résultats aux tests à la fin du CE2				En éducation prioritaire (ZEP), 4 % de moins sur la note moyenne à l'examen terminal du brevet		Pas d'effet significatif pour les enfants d'origine sociale défavorisée
Source	[7]	[9]	[22]	[11]	[11] sur le passage en classe supérieure, [23] sur les scores en mathématiques et en sciences		[11] sur le passage en classe supérieure, [22] sur les notes au brevet	[22]		
Type de travaux	Expérimentation Expérience « naturelle »									
Données utilisées	Expérimentation CP à effectifs réduits; dix académies, 2002-2004	Panel d'élèves recrutés au CP en 1997, enquêtes 19, 1998-1999 et 1999-2000, MEN-DEPP	Enquête ministérielle de 1991-1992 sur les élèves de CE2 et leurs enseignants (12 départements)	Panel d'élèves recrutés en 6 ^e en 1989, MEN-DEPP [11]. Enquête <i>Third International Mathematics and Science Study</i> (TIMSS) 1994-1995, <i>International Association for the Evaluation of Educational Achievement</i> (IEA) [23]. Base Diplôme national du brevet 2004, base Scolarité 2003-2004, fichiers Thèmes 2003-2004, MEN-DEPP [22]	Panel d'élèves recrutés en 6 ^e en 1995, bases Scolarité et fichiers Thèmes 2001 à 2003, MEN-DEPP					

BIBLIOGRAPHIE

- [1] **Abdouni S.** (2013), Les élèves du premier degré à la rentrée 2013 : un effort pour la scolarisation à deux ans dans l'éducation prioritaire, *Note d'Information*, 13.33, MEN-DEPP.
- [2] **Afsa C.** (2014), Une question de taille, *Éducation & formations*, ce numéro, MENESR-DEPP.
- [3] **Angrist J. D., Lavy V.** (1999), Using Maimonides' Rule to Estimate the Effect of Class-Size on Scholastic Achievement, *The Quarterly Journal of Economics*, 114 (2):533-575.
- [4] **Bressoux P., Kramarz F., Prost C.** (2009), Teachers' Training, Class-Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes, *The Economic Journal*, 119:540-561.
- [5] **Bressoux P., Lima L.** (2011), La place de l'évaluation dans les politiques éducatives : le cas de la taille des classes à l'école primaire en France, *Raisons éducatives*, 15:99-123.
- [6] **Cosnefroy O., Rocher T.** (2005), Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats, *Les Dossiers - Enseignement scolaire*, n° 166, 118 p., MEN-DEP.
- [7] **Direction de l'évaluation et de la prospective** (2005), L'expérimentation d'une réduction des effectifs en cours préparatoires, *Note Évaluation*, 05.03, MEN-DEP.
- [8] **Dynarski S., Hyman J. M., Schanzenbach D. W.** (2011), *Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion*, NBER Working Paper, 17533.
- [9] **Ecalte J., Magnan A., Gibert F.** (2006), Class-size effects on literacy skills and literacy interest in first grade: A large-scale investigation, *Journal of School Psychology*, 44:191-209.
- [10] **Fougère D.** (2012), Les méthodes d'expérimentation en question, *Éducation & formations*, n° 81, p. 41-47, MEN-DEPP.
- [11] **Gary-Bobo R., Mahjoub M. B.** (2013), Estimation of Class-Size Effects, Using "Maimonides' Rule" and Other Instruments: the Case of French Junior High Schools, *Annales d'Économie et de Statistique*, 111-112:193-225.
- [12] **Krueger A. B.** (1999), Experimental Estimates of Education Production Functions, *The Quarterly Journal of Economics*, 114 (2):497-532.
- [13] **Hanushek E. A.** (2003), The Failure of Input-Based Schooling Policies, *The Economic Journal*, 113:F64-F98.
- [14] **Lazear E. P.** (2001), Educational Production, *The Quarterly Journal of Economics*, 116 (3):777-803.
- [15] **L'état de l'École** (2013), 78 p., MEN-DEPP.
- [16] **Meuret D.** (2001), *Les recherches sur la réduction de la taille des classes*, Rapport pour le Haut Conseil de l'évaluation de l'école, 43 p.
- [17] **Murat F.** (1998), Les différentes façons d'évaluer le niveau des élèves en fin de collège, *Éducation & formations*, n° 53, p. 35-49, MEN-DEP.
- [18] **Murat F.** (2009), Le retard scolaire en fonction du milieu parental : l'influence des compétences des parents, *Économie et Statistique*, 424-425:103-124.
- [19] **O'Prey S.** (2012), La taille des structures pédagogiques dans les établissements du second degré, *Note d'Information*, 12.27, MEN-DEPP.
- [20] **Piketty T.** (2004), *L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises. Une estimation à partir du panel primaire 1997*, Document de travail Paris-Jourdan, 71 p.
- [21] **Piketty T., Valdenaire M.** (2006), L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. Estimations à partir du panel primaire 1997 et du panel secondaire 1995, *Les Dossiers - Enseignement scolaire*, n° 173, 153 p., MEN-DEP.
- [22] **Valdenaire M.** (2011), *Essais en économie de l'éducation*, Thèse de doctorat, École des hautes études en sciences sociales, 337 p.
- [23] **Wößmann L., West M.** (2006), Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS, *European Economic Review*, 50 (3):695-736.