

# CONFÉRENCE DE CONSENSUS

## LUTTER CONTRE LES DIFFICULTÉS SCOLAIRES : LE REDOUBLEMENT ET SES ALTERNATIVES ?

Partie 2. Le redoublement en France  
et dans le monde : de l'étude de ses  
impacts à la croyance en son utilité

Décembre 2014



# **Lutter contre les difficultés scolaires : le redoublement et ses alternatives ?**

## **Partie 2 :**

**Le redoublement en France et dans le monde :  
de l'étude de ses impacts à la croyance en son utilité**



---

# Table des matières

<b>Avant propos</b>	<b>i</b>
<b>Résumé</b>	<b>iii</b>
<b>1 Les effets du redoublement : une revue de littérature</b>	<b>1</b>
I    Un consensus scientifique jusqu'au milieu des années 2000 . . . . .	2
1    Les synthèses de littérature jusqu'aux années 2000 . . . . .	2
2    Les limites méthodologiques de ces études . . . . .	8
II   Les effets du redoublement revisités . . . . .	14
1    Les effets du redoublement en maternelle . . . . .	14
2    Les effets du redoublement au primaire . . . . .	17
3    Les effets du redoublement dans le secondaire . . . . .	26
III  Performances des systèmes scolaires :	
Le redoublement à l'épreuve des comparaisons internationales . . . . .	30
1    Les performances des élèves . . . . .	30
2    Le taux de retard . . . . .	32
3    Origine sociale . . . . .	38
IV  Les effets du redoublement sur les parcours scolaires et l'insertion dans la vie active . . . . .	42
1    L'effet du redoublement sur le décrochage et l'éducation supérieure . . . . .	42
2    Les effets du redoublement sur l'insertion dans la vie active . . . . .	45
<b>2 Les rôles du redoublement</b>	<b>47</b>
I    Les croyances et attitudes des enseignants et des familles vis-à-vis du redoublement . . . . .	48
II   Les fonctions latentes du redoublement . . . . .	52
III  La connaissance de la recherche sur le redoublement par les enseignants . . . . .	55
<b>Conclusion</b>	<b>57</b>
<b>Annexes</b>	<b>59</b>
I    La correction du biais de sélection dans l'estimation des effets du redoublement . . . . .	60

II	Commentaire technique sur l'ajout de <i>mauvaises</i> variables de contrôle . . . . .	66
	<b>Bibliographie</b>	<b>68</b>

Ce document s'inscrit dans une série de rapports publiés par le Conseil national d'évaluation du système scolaire (Cnesco) autour de la question :

### **Lutter contre les difficultés scolaires : le redoublement et ses alternatives ?**

Ils ont été portés par Arthur Heim et Claire Steinmetz, chargés de missions au Cnesco, sous la direction de Nathalie Mons, présidente du Cnesco et professeure à l'université de Cergy Pontoise. Une assistance analytique et éditoriale a été apportée par Pascal Bressoux, professeur à l'université de Grenoble et membre du Cnesco, Marcel Crahay, professeur à l'université de Genève, Dominique Goux, chercheuse au CREST et membre du Cnesco, Marc Gurgand, professeur à l'école d'économie de Paris et membre du Cnesco, Pierre Vrignaud, professeur émérite à l'université Paris Ouest et président du cycle de conférences de consensus du Cnesco et de l'IFÉ (ENS de Lyon), et Anne West, professeure à London School of Economics et membre du Cnesco, et avec la participation de Alain Lopes et Philippe Romain, chargés de mission au Cnesco.

Les données exploitées dans le présent rapport proviennent de la Base Centrale de Scolarité (MENESR-DEPP), du Panel d'élèves entrés en 6<sup>ème</sup> en 1995 (MENESR-DEPP [producteur], Centre Maurice Halbwachs [diffuseur]) et des enquêtes PISA (OCDE).

Pour toute question relative au contenu de ce rapport, contacter [arthur.heim@education.gouv.fr](mailto:arthur.heim@education.gouv.fr)

Les opinions et arguments exprimés n'engagent que le Cnesco.

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

Publié en décembre 2014

Conseil National d'évaluation du système scolaire

Carré Suffren - 31-35, rue de la Fédération

75 015 Paris



---

# Avant propos

Le présent rapport s'inscrit dans le processus de réflexion de la conférence de consensus organisée par le Conseil national d'évaluation du système scolaire (Cnesco) et l'Institut Français d'Éducation (IFÉ/ENS de Lyon) dont le titre est :

## **Lutter contre les difficultés scolaires : le redoublement et ses alternatives ?**

Très débattu dans les salles de classe et l'opinion publique, le redoublement a également généré une littérature scientifique abondante dont de récents développements ont fait évoluer les conclusions. En France, sa pratique est en pleine mutation. Parfois contesté par les parents - notamment au primaire - ou exigé dans une visée stratégique au collège et au lycée, le redoublement a nettement reculé ces dix dernières années, faisant évoluer les profils des élèves qui y sont sujets. Pour autant, la France demeure dans le groupe de tête des pays qui font le plus redoubler leurs élèves.

La phase exploratoire de la conférence de consensus - dont l'objectif principal était d'identifier, d'analyser et de synthétiser la littérature scientifique internationale étudiant le redoublement et ses alternatives - aboutit à une série de trois rapports :

### **Le redoublement en France et dans le monde :**

*une comparaison statistique et réglementaire ;*

### **Le redoublement en France et dans le monde :**

*de l'étude de ses impacts à la croyance en son utilité ;*

### **Le redoublement en France et dans le monde :**

*quelles alternatives au redoublement ?*

Cette série permet en premier lieu aux participants à la conférence de consensus de s'approprier la thématique du redoublement et de ses alternatives en disposant d'une revue exhaustive de la littérature scientifique à ce sujet. Elle offre également à toute personne, utilisateur ou acteur de l'éducation, professeur ou étudiant, les éléments statistiques et réglementaires les plus récents, une analyse scientifique de l'impact et des perceptions du redoublement et enfin un ensemble de dispositifs et d'organisations qui complètent ou se substituent au redoublement et permettent de lutter efficacement contre les difficultés scolaires.

Ce rapport est le deuxième de la série. Nous résumons ici ses principaux apports.



---

# Résumé

## La recherche sur les effets du redoublement : des résultats récents affinés

Veine de recherche très féconde, les études scientifiques sur les effets du redoublement se sont encore enrichies de nouveaux résultats depuis les années 2000.

### Un consensus de longue date autour de l'inefficacité du redoublement

Jusqu'au début des années 2000, des chercheurs ont mis en évidence de façon récurrente des effets négatifs du redoublement à court et à long terme, notamment sur les résultats scolaires. Cependant, ces études souffraient de plusieurs limites méthodologiques. La diversité des pratiques du redoublement dans le temps et dans le monde a depuis longtemps généré une demande forte pour une évaluation scientifique des effets du redoublement. De nombreux auteurs ont ainsi étudié ses effets sur les résultats des élèves ou sur d'autres indicateurs de développement, de motivation, d'aspiration des élèves,... Les résultats des recherches conduites entre 1911 et 1989 ne plaident pas en faveur du redoublement comme moyen de remédier aux difficultés ou à l'échec scolaire. La première revue de littérature menée par Jackson (1975) recensait 30 études, aux méthodologies très variées, cherchant à comparer les performances des redoublants et non-redoublants et concluait qu'"il n'y a pas un ensemble de preuves fiables pour indiquer que le redoublement est plus favorable que la promotion automatique pour les élèves ayant des difficultés académiques sérieuses ou des difficultés d'adaptation".

Holmes et Matthews (1984) ont fait une méta-analyse<sup>1</sup> de la littérature sur les effets du redoublement aux États-Unis entre 1929 et 1981 et obtiennent à chaque fois des différences statistiquement notables entre élèves promus et redoublants dans tous les champs de comparaison. Ils montrent alors qu'à QI donné, les redoublants ont de moins bons résultats en mathématiques et en lecture, ont une plus faible estime d'eux-même, sont plus fréquemment absents et ont plus souvent une attitude négative envers l'école. Ces résultats ont été confirmés par Holmes (1989) qui a conduit une seconde méta-analyse

---

1. Une méta-analyse est une démarche statistique combinant les résultats d'un ensemble d'études indépendantes sur un sujet de recherche. Elle permet une analyse plus précise des données par l'augmentation du nombre de cas étudiés et de tirer une conclusion globale.

incluant 19 études supplémentaires. Enfin, Jimerson (2001) a synthétisé la littérature produite entre 1990 et 1999 dans une méta-analyse structurée de façon à tenir compte de la rigueur méthodologique. Ses résultats montrent que la comparaison des redoublants avec des non-redoublants aux caractéristiques observables similaires produit dans la majorité des cas des différences négligeables (dans 48 % des mesures de performance scolaire et dans 86 % des analyses de mesures non-cognitives) ou en faveur des élèves promus (respectivement 47 % et 9 %).

En somme, la littérature jusqu'au début des années 2000 aboutit à un consensus fort mettant en évidence des effets négatifs du redoublement. Au mieux, ce dernier n'a pas d'effet mais il s'avère dans bien des cas nocif pour la réussite scolaire des élèves, pour le développement de leur estime de soi et pour diverses capacités non cognitives.

Néanmoins, cette littérature souffre de plusieurs faiblesses méthodologiques qu'ont révélées plusieurs articles récents.

Tout d'abord, Allen et al. (2009) font une méta-analyse des études publiées entre 1990 et 2007 en Amérique du Nord sur l'effet du redoublement sur les performances scolaires. L'apport principal de ce travail est de s'intéresser à la rigueur de la méthodologie employée dans les articles recensés pour déterminer l'impact du redoublement. Ils s'intéressent également au moment où l'analyse est menée (i.e. combien de temps après le redoublement) et à la façon dont les élèves redoublants et non redoublants sont comparés. Ils montrent alors que les études qui utilisent une méthodologie plus rigoureuse obtiennent des effets du redoublement à chaque fois négligeables. Ainsi, lorsque les chercheurs tentent de réduire les différences préexistantes entre le groupe de redoublants et le groupe de non redoublants avec lesquels ils le comparent, l'effet du redoublement à court terme est proche de zéro. Par contre, les auteurs montrent que plus on mesure l'effet longtemps après le redoublement, plus il est négatif. Ce résultat est davantage marqué pour les études qui comparent les résultats à niveau égal<sup>2</sup> qu'à âge égal<sup>3</sup>. Au final, Allen et al. (2009) montrent que lorsque les chercheurs contrôlent davantage les différences préexistantes entre redoublants et non-redoublants, l'impact n'est pas aussi négatif qu'il n'y paraît, mais il semble plus nocif à long terme.

Ces résultats montrent que la comparaison entre les redoublants et non-redoublants ne mesure pas uniquement l'impact du redoublement. En effet, il existe des caractéristiques (observables ou non) qui influencent à la fois les chances de redoubler, le niveau scolaire, la motivation de l'élève, ... Si les chercheurs n'en tiennent pas compte, ils obtiennent des résultats qui confondent tous ces paramètres. C'est ce qu'on appelle le biais de sélection. La littérature récente que nous synthétisons ci-après emploie des méthodes quasi-expérimentales qui exploitent l'existence naturelle de groupes cibles affectés par le redoublement et comparent leurs résultats à ceux de l'ensemble de la population à laquelle ils appartiennent, en tenant sous contrôle les différences par des techniques statistiques.

---

2. Qui comparent les redoublants avec les élèves qui sont de le même niveau scolaire qu'eux, et donc plus jeunes.

3. Qui comparent les redoublants avec les élèves de la même génération qui sont donc dans un niveau scolaire plus élevé.

## Un renouvellement de la recherche sur le redoublement depuis 2000

Depuis 2000, de nouvelles méthodes statistiques et économétriques permises par les progrès technologiques (en puissance de calculs notamment) aboutissent à des résultats plus affinés sur les effets du redoublement. On constate désormais une distinction importante entre l'impact à court terme et à plus long terme. Ces recherches mettent en évidence des effets qui peuvent être positifs en termes de résultats scolaires à court terme - notamment l'année du redoublement - mais toujours négatifs à long terme. Ces nouvelles recherches explorent, de plus, des nouveaux niveaux scolaires, comme le collège. Le lycée reste cependant un champ totalement vierge.

Les années 2000 ont vu se développer de nouvelles méthodes d'estimations de l'impact causal de politiques publiques qui ont été appliquées au cas du redoublement. Ainsi, les articles de [Hong et Raudenbush \(2006\)](#) et de [Dong \(2010\)](#) analysent l'effet du maintien en dernière année de maternelle aux États-Unis avec des méthodes très différentes. De leur côté, [Hong et Raudenbush \(2006\)](#) concluent que la rétention en maternelle ne produit pas d'effets bénéfiques aux élèves même à très court terme. Au contraire, [Dong \(2010\)](#) emploie une méthode différente - visant notamment à éliminer l'effet de caractéristiques inobservables pour le chercheur<sup>4</sup>- et montre que les élèves ayant été maintenus en maternelle ont des résultats environ 14 % plus élevés au CP en mathématiques et en lecture que s'ils avaient été promus directement. L'effet sur les performances en mathématiques diminue fortement mais reste non-négligeable en CE<sub>2</sub> ; celui sur la lecture disparaît complètement en CE<sub>2</sub>. Ainsi, il semble que si l'effet du redoublement en maternelle peut être bénéfique à court terme aux élèves, son effet semble s'atténuer très rapidement.

Les articles étudiant les effets du redoublement au primaire sont très nombreux, mais sont également très spécifiques aux systèmes scolaires qu'ils étudient. Ainsi, une partie importante de la littérature étudie les effets à court et à moyen termes du redoublement lorsque celui-ci est décidé sur la base de notes à un test standardisé ([Jacob et Lefgren, 2004, 2007](#); [Roderick et Nagaoka, 2005](#); [Schwerd et West, 2012](#)). Ce type de règle fournit un contexte idéal pour une évaluation rigoureuse de l'effet du redoublement. Les résultats de ces études vont généralement dans le même sens : le redoublement engendre des gains positifs ou nuls à court terme sur les performances en mathématiques et en lecture pour les élèves à la limite des exigences de performances attendues. Cependant, le redoublement n'est généralement pas la seule remédiation dont ils bénéficient. Dans certains cas, ils sont placés dans des classes moins nombreuses, avec des enseignants expérimentés, suivent des cours d'été, etc. Ces éléments indiquent donc que le redoublement au primaire peut avoir des effets positifs à court terme s'il est accompagné d'autres remédiations. Cependant, les effets s'estompent là aussi très rapidement (trois ans après).

Plus proche de nous, [Goos \(2013\)](#) a récemment consacré une thèse de doctorat à l'étude des effets du redoublement au CP sur les apprentissages, les aspects psychosociaux (motivation, absentéisme,

---

4. Le praticien qui décide de faire redoubler dispose d'informations personnelles sur l'élève (son assiduité, sa participation, ses efforts,...) qui ne sont pas disponibles pour les chercheurs mais affectent la probabilité de redoubler. Il faut donc que les méthodes d'évaluation en tiennent compte.

ambition, etc.) et la carrière scolaire ultérieure des élèves en Belgique flamande. Elle montre alors que les élèves redoublants réussissent mieux en mathématiques et en lecture l'année de leur redoublement que leurs pairs à risque qui ont été promus. Cependant, cet effet disparaît lors de la seconde année, c'est-à-dire lorsque les redoublants doivent à nouveau apprendre au même rythme que les autres. Par ailleurs, l'auteur apporte des éléments solides qui indiquent que les redoublants auraient progressé plus rapidement en mathématiques et en lecture s'ils n'avaient pas été retenus. Les indicateurs de développements psychosociaux révèlent des profils moins avancés pour les redoublants (ce qui ne surprendra pas le praticien) mais leurs estimations montrent qu'ils se seraient développés plus rapidement s'ils n'avaient pas redoublé leur CP. [Goos \(2013\)](#) met en relation ces résultats avec les études américaines précédemment évoquées et suppose que dans ses données, "*beaucoup de redoublants recyclent simplement la matière, au lieu de recevoir une instruction à leur rythme, dans leur groupe de niveau*".

En France, [Cosnefroy et Rocher \(2005\)](#) ont réalisé plusieurs travaux sur les effets du redoublement en employant différentes méthodes. Leurs conclusions pointent du doigt les effets délétères du redoublement en matière de performance scolaire et sa nocivité en termes de motivation et de sentiment d'efficacité. Par ailleurs, les auteurs comparent les notes données par les enseignants aux élèves et leurs résultats à des tests standardisés et constatent que les notes données par les enseignants au sein d'une même classe varient moins que les performances absolues des élèves. En revanche, l'écart réel entre les classes est gommé par la notation des enseignants. Dès lors, un élève peut être noté très bas dans une classe donnée alors qu'en réalité, son niveau absolu<sup>5</sup> est plutôt bon, et réciproquement. En l'absence de repère extérieur à la classe tel que des tests communs, la décision du redoublement se fait presque à l'aveugle et confère un caractère arbitraire à la décision.

Alors que la littérature sur les effets du redoublement au primaire abonde, peu d'études rigoureuses portent sur le second degré ([Crahay, 2004](#)). En France, [Gary-Bobo et al. \(2013\)](#) ont étudié le redoublement au collège. Ils mettent en évidence des effets en général négatifs sauf pour les élèves les plus faibles qui, d'après leurs calculs, bénéficient légèrement du redoublement. Cependant, la probabilité d'accéder à la 3<sup>ème</sup> est toujours négativement affectée par le redoublement (sauf pour les élèves les plus compétents pour qui il n'y a pas d'effet). Les auteurs concluent donc que si le redoublement peut avoir un impact positif sur les résultats des élèves les plus faibles, cet effet n'est pas très important et ne compense pas l'augmentation du risque de décrochage avant la fin de la 3<sup>ème</sup>.

À la marge de ces résultats, [Hill \(2014\)](#) pose la question des nuisances que peuvent générer un nombre de redoublants trop élevé dans une classe. À partir de données américaines, il montre alors qu'une part trop élevée de redoublants a tendance à faire redoubler les élèves qui suivent ces cours pour la première fois. La littérature sur les effets de voisinage est très active. Cependant, nous n'avons pas connaissance d'autres articles traitant ce sujet. Nous prenons donc ces résultats avec prudence.

Au total, les premières recherches sur les effets du redoublement trouvaient des effets systématiquement négatifs à court et à long termes du redoublement mais souffraient de plusieurs problèmes

---

5. Mesuré par les tests standardisés.

méthodologiques et de données parfois lacunaires ou difficilement comparables. Les évolutions récentes ont permis de corriger ces problèmes et d'affiner sensiblement les conclusions. Dans la majorité des études recensées, le redoublement semble n'avoir aucun effet sur les variables étudiées à moyen terme. Quelques études obtiennent des effets positifs à court terme dans des contextes très particuliers, sans pour autant conclure que le redoublement est une remédiation efficace contre les difficultés d'apprentissage. On remarquera que si le redoublement a donné lieu à de nombreuses recherches au primaire, le secondaire reste aujourd'hui très peu investigué.

### Performances des systèmes scolaires : le redoublement à l'épreuve des comparaisons internationales

Si au niveau des élèves, le redoublement ne permet pas, au moins dans la durée, d'améliorer les performances des élèves qui y sont sujets, on peut s'interroger sur les effets de cette pratique sur les performances des systèmes scolaires. Les enquêtes internationales telles que PISA (OCDE) sont particulièrement adaptées pour répondre à cette question.

#### Redoublement et performances : une analyse macro

Les vagues successives d'enquêtes PISA - auxquelles la France a participé - permettent de mettre en valeur quelques résultats importants quant au lien entre redoublement et performances globales du système. [Crahay \(2004\)](#) exploite ainsi la première vague de l'enquête PISA (2000) et montre qu'il y a d'avantage d'hétérogénéité entre les élèves dans les pays pratiquant plus le redoublement, même si certains font exception.

Par ailleurs, l'étude des compétences des élèves met aussi en évidence une disparité des performances entre ceux à l'heure et ceux ayant déjà redoublé. En effet, [Cosnefroy et Rocher \(2005\)](#) montrent que *"quel que soit le moment du cursus scolaire, les élèves en retard ont, en moyenne, des résultats nettement inférieurs aux élèves à l'heure"*. En prenant le cas de la France (enquête PISA 2000), les élèves à l'heure à 15 ans (2<sup>nd</sup>e générale ou technologique : 48,2 %) ont les meilleurs résultats en compréhension de l'écrit (devant la Finlande) alors que ceux en retard d'un ou deux ans (3<sup>ème</sup> : 36,5 % ou 4<sup>ème</sup> : 7,11 %) ont de très mauvais résultats, bien inférieurs à la moyenne de l'OCDE. Des études menées en France, en Espagne et en Belgique (Communauté Française) montrent à chaque fois les mêmes résultats : ces pays qui pratiquent le redoublement révèlent que les élèves en retard ont de moins bons résultats que les élèves à l'heure. Nos propres travaux sur les données PISA 2012 confirment ce diagnostic.

#### Taux de retard

La dernière enquête PISA révèle l'importante hétérogénéité du taux d'élèves ayant déjà redoublé suivant les pays. Parmi les 65 nations participant à l'enquête, 55 % ont un taux inférieur à 10 %. On trouve notamment dans cette liste les pays nordiques (Norvège, Suède, Finlande, Danemark), les pays

baltes et slaves, le Japon et les dragons asiatiques (Taïwan, Corée du Nord, Malaisie, Thaïlande, Vietnam, Shanghai) et les pays du Common Wealth Britanique (Angleterre, Irlande, Australie, Nouvelle-Zélande, Canada). La plupart de ces pays pratiquent la promotion automatique mais autorisent le redoublement dans certains cas exceptionnels (maladie, fort absentéisme,...).

Les pays d'Europe occidentale font partie des pays ayant le plus fort taux de redoublement avec l'Amérique Latine. Ainsi, les taux de redoublement en France et aux Pays-Bas sont de 28 %, le même niveau que le Pérou. Le Luxembourg et la Belgique ont le même taux de redoublement que le Brésil et l'Argentine, soit 35 % environ (PISA, 2012).

Si on rapporte ces taux aux performances moyennes des élèves à 15 ans dans ces pays, on observe une relation statistique faible mais notable qui indique que les pays à fort taux de redoublement ont tendance à être moins performants que les pays ayant un taux de redoublement plus faible. Ce résultat a déjà été obtenu sur des données PISA (2000) par Nathalie Mons ([Duru-Bellat et al., 2004](#)), ce qui montre la stabilité de ce résultat dans le temps, malgré l'évolution des pratiques du redoublement. Les auteurs concluaient alors que *"la pratique du redoublement, non seulement n'améliore pas les résultats des élèves, mais tend à les amoindrir, quand on les mesure chez les élèves d'un âge donné"* ([Duru-Bellat et al. \(2004\)](#) p. 65 cité dans [Paul et Troncin \(2004\)](#) p. 28). Cette relation n'est toutefois plus aussi marquée (mais toujours négative) lorsqu'on restreint l'échantillon aux seuls pays membres de l'OCDE.

### **Origine sociale**

Les enquêtes révèlent aussi un lien entre performances des élèves et origine sociale. En France, l'indice de statut économique, social et culturel (SESC) de l'OCDE montre que la performance des élèves est davantage liée aux origines socio-économiques que dans les autres pays ([Fumel et al., 2010](#)). Les élèves provenant de milieux défavorisés ont à la fois un taux de redoublement plus élevé et un niveau de performance plus faible que les élèves d'origines favorisées. De même, les élèves ayant des parents d'origines étrangères ont plus de chance d'être en retard comparés aux autres enfants ([Désir, 2014](#)).

D'un point de vue plus statistique, l'analyse de l'[OCDE \(2013\)](#) montre que 26 % de l'effet du statut socio-économique de l'élève sur ses performances en mathématiques peuvent être expliqués par des différences dans les taux de redoublement entre les pays.

Ainsi, il existe une relation entre hétérogénéité des performances et inégalités sociales : plus les pays ont des écarts de performances importants, plus les inégalités sociales y sont fortes. À l'inverse, Nathalie Mons (2004) montre que les pays avec des élèves aux performances homogènes présentent moins d'inégalités sociales ([Duru-Bellat et al., 2004](#)).

## Les effets du redoublement sur les parcours scolaires et l'insertion dans la vie active

De nombreuses recherches montrent, dans une variété de pays pratiquant le redoublement, qu'avoir été retenu une année est associé à un risque de décrochage plus élevé. Cette littérature est recensée dans Jimerson (1999) qui discute également les limites méthodologiques de ces études. Il montre par ailleurs que les élèves ayant redoublé sont significativement plus nombreux (en pourcentage) à abandonner l'école avant 19 ans et moins nombreux à obtenir un diplôme au lycée ainsi qu'à être inscrits à 20 ans dans l'enseignement supérieur.

Fine et Davis (2003) étudient les trajectoires scolaires des élèves ayant redoublé mais ayant cependant obtenu un diplôme du secondaire. Ils montrent que ces élèves, ayant pourtant déjà "battu les statistiques", ont une probabilité nettement plus faible d'être inscrits dans le supérieur. Ces résultats indiquent que même les élèves les plus persistants dans le supérieur parmi les redoublants ont des trajectoires d'éducation différentes des élèves ayant toujours été promus, toutes choses égales par ailleurs<sup>6</sup>.

En France, Afsa (2011) fait plusieurs estimations de l'effet d'avoir redoublé la 6<sup>ème</sup> sur la probabilité d'obtenir un diplôme du secondaire à partir des données du Panel 1995 de la DEPP (MENESR). Il montre notamment que le redoublement a globalement un effet négatif marqué sur la probabilité d'obtenir un diplôme du secondaire.

Enfin, le travail de Brodaty et al. (2008) est, à notre connaissance, le seul à estimer l'effet du redoublement sur l'insertion dans la vie active. Les auteurs trouvent qu'une année de retard entraîne une diminution du revenu moyen pendant les 5 premières années dans la vie active d'environ 9 %. En comparant ces estimations avec ce que rapporte une année d'étude supplémentaire calculé dans d'autres travaux, ils concluent qu'un étudiant qui redouble une année pour obtenir un diplôme l'année suivante fait en réalité une opération blanche en termes de perspectives de gains salariaux futurs. Les auteurs interprètent ces résultats comme une preuve empirique que le redoublement (ou plutôt le retard) est utilisé par les employeurs comme un signal de productivité des candidats durant les toutes premières années d'entrée dans la vie active.

Au final, les effets du redoublement sur les performances scolaires et les indicateurs psychosociaux étudiés peuvent dans certains contextes scolaires apparaître positifs à court terme mais sont toujours négatifs sur le long terme. Le redoublement a, par contre, toujours un effet négatif sur les trajectoires scolaires et demeure le meilleur déterminant du décrochage. Il semble également impacter négativement le revenu futur du jeune adulte en agissant comme un signal de faible performance du salarié pour les entreprises.

---

6. C'est à dire, à caractéristiques observables données.

Cependant, cette pratique pédagogique très critiquée par la recherche persiste dans de nombreux pays, notamment en France, ce qui révèle un certain attachement des enseignants et des parents.

## Les usages et les représentations sociales du redoublement chez les enseignants et les parents

Des études récentes ont cherché à comprendre l'attachement des enseignants et des parents au redoublement. Ceux-ci considèrent encore le redoublement comme un moyen de remédiation efficace. Ils lui attribuent par ailleurs un rôle instrumental. Cependant, les représentations sont en train de changer dans ce domaine.

Le récent sondage de [OpinionWay \(2012\)](#) révèle que 70 % des parents et 64 % des enseignants interrogés sont d'accord avec la phrase "*Le redoublement permet réellement à l'élève de rattraper son retard et d'être mieux préparé pour les classes supérieures*". Ils sont toutefois une large majorité à penser que le redoublement pourrait "*être remplacé par d'autres mesures plus efficaces et mieux adaptées, telles que l'accompagnement personnalisé pour les enfants en difficulté*" (77 %). Cet élément révèle une forme de paradoxe de l'opinion public : le redoublement est une bonne chose, mais il est remplaçable par quelque chose de plus efficace. Ce type de réponse indique que l'adhésion des parents au redoublement se fait faute de mieux. Le point de vue des parents est peu étudié dans la littérature scientifique contrairement à celui des enseignants. [Paul et Troncin \(2004\)](#) citent cependant plusieurs études anciennes qui révèlent une adhésion massive au redoublement. Les recherches de [Troncin \(2005\)](#) montrent que très peu de parents sont prêts à s'opposer au redoublement de leur enfant en CP, ce que l'auteur interprète comme un signe de confiance des parents envers les enseignants à ce niveau scolaire. Cependant, les auteurs notent que les familles dites "favorisées" sont nettement moins disposées à accepter d'emblée la décision de redoublement, ce qui peut indiquer un plus grand scepticisme à l'égard du redoublement. En particulier, de nombreux redoublements sont aujourd'hui contestés par les parents au primaire. Ces observations ont été confortées par les premières observations des panels exploratoires organisés par le Cnesco.

Les résultats du sondage d'[OpinionWay \(2012\)](#) en matière de croyance des enseignants sont en adéquation avec la littérature scientifique à ce sujet ([Draelants, 2006](#); [Crahay, 2007](#); [Boraita et Marcoux, 2013](#); [Range et al., 2011](#); [Terry, 2011](#); [Wynn, 2010](#)). La plupart de ces études sont teintées d'une forme de surprise, voire d'exaspération des chercheurs ([Draelants, 2008](#)) qui se manifeste très bien dans le titre de l'article de [Marcoux et Crahay \(2008\)](#) : "*Mais pourquoi continuent-ils à faire redoubler ?*"

Pourtant, les enseignants ont de bonnes raisons de croire aux vertus du redoublement ([Xia et Glennie, 2005](#)), notamment parce que "*dans la plupart des cas, l'élève redoublant sera un peu meilleur durant son année de redoublement*" ([Draelants, 2006](#)). Le jugement de l'efficacité pédagogique apparaît alors comme un mélange de bon sens et d'observation qui est très différent de l'approche des chercheurs sur la question.

Marcoux et Crahay (2008) expliquent aussi l'adhésion des enseignants au redoublement par une vision séquentielle des apprentissages qui implique qu'un élève ne maîtrisant pas une des étapes de cette séquence ne peut pas progresser davantage. Dès lors, le redoublement apparaît comme une solution adaptée pour solidifier les bases. Il est de plus perçu comme un moyen permettant aux élèves de gagner en maturité afin d'être plus aptes à comprendre et acquérir les compétences, puisque leurs capacités cognitives seront plus développées que l'année précédente. Aussi, le redoublement permettrait d'éviter qu'ils ne soient complètement perdus l'année suivante en ayant accumulé des lacunes.

Pour Dubet (2002), les enseignants sont attachés au redoublement car il est conforme à leur vision morale de l'école républicaine. Le sociologue écrit : "*l'école démocratique repose sur deux piliers [...] : un principe d'égalité, tous les élèves sont fondamentalement égaux et peuvent prétendre aux mêmes choses ; un principe de mérite, fondant des inégalités justes*". Or ces deux piliers sont contradictoires car ils impliquent de classer les élèves et d'affirmer leur égalité, ce qui oblige à expliquer leurs inégalités de performances comme les conséquences de leur liberté (Draelants, 2008). Au final, cette conception morale de l'enseignement crée un attachement au redoublement qui s'intègre alors comme une institution conforme aux valeurs de l'École démocratique dans lesquelles "*tous les élèves sont égaux et les meilleurs sont ceux qui travaillent plus*" (Dubet, 2002, p.18).

Par ailleurs, Draelants (2006, 2008) a mis en exergue d'autres rôles associés au redoublement qui expliquent l'attachement des professeurs et des parents à cette pratique. Il les désigne comme des "*fonctions latentes*", c'est-à-dire des fonctions cachées du redoublement dont l'usage ne vise pas seulement l'objectif explicite de remédiation scolaire.

Ainsi, le redoublement joue un rôle incitatif sur le travail des élèves et certains voient dans le redoublement une forme de menace, de punition incitant les élèves à augmenter leurs efforts afin de ne pas redoubler. Draelants (2008) constate ainsi en Belgique que "*sans redoublement les notes perdent beaucoup de leur pouvoir régulateur*".

Le redoublement peut aussi apparaître comme un moyen de gérer l'hétérogénéité au sein des classes et des établissements. Cet argument repose sur l'idée que l'élève redoublant est moins en difficulté parmi les élèves plus jeunes qui découvrent le programme. Pour Draelants (2008), l'interdiction du redoublement en 2001 en Belgique a compliqué le travail des enseignants pour qui la gestion de l'hétérogénéité est la principale difficulté. Ainsi, l'attachement au redoublement peut aussi refléter une certaine crainte de dégradation des conditions de travail.

En outre, le redoublement joue un rôle de signal envoyé par l'enseignant à ses pairs et par l'établissement aux autres établissements, aux enseignants et aux parents. En effet, Gary-Bobo et Robin (2014); Crahay (2007) expliquent que les enseignants peuvent être favorables au redoublement par peur d'être jugés par leurs pairs et d'être considérés comme laxistes. En laissant passer un élève trop faible dans le niveau supérieur dont un collègue est en charge, l'enseignant s'expose aux critiques, place l'autre enseignant dans une situation délicate, et apparaît complaisant avec les élèves en difficulté.

Le redoublement peut également être utilisé par le chef d'établissement pour signaler aux parents

et aux élèves le niveau d'exigence scolaire qu'il attend de son établissement. Dans un contexte de concurrence entre établissements et dans un environnement institutionnel de quasi-marché, un fort taux de redoublement permet d'exclure, sans l'afficher, des élèves d'un niveau insuffisant, car ces derniers pourront s'inscrire ailleurs afin d'éviter de redoubler. Les parents privilégiant une stratégie élitiste seront sensibles à cette hiérarchie d'excellence. S'ils jugent un établissement non sur sa capacité à faire réussir le plus grand nombre mais sur sa sélectivité, le redoublement leur apparaît comme un signal crédible de performance de l'établissement.

Les croyances des enseignants évoluent dans le temps, comme le montre la revue de littérature et les analyses de [Crahay et al. \(2010\)](#). D'après leurs études, ces croyances se fondent principalement sur l'expérience passée des enseignants et non sur les résultats de la recherche. Malgré les preuves de l'inefficacité du redoublement, la plupart des enseignants reste persuadée que faire redoubler un élève en difficulté est la solution la plus adaptée et la plus efficace [Terry \(2011\)](#); [Marcoux et Crahay \(2008\)](#).

Les connaissances des enseignants sur la recherche sur le redoublement ne permettent pas de les faire changer d'avis quant à l'inefficacité du redoublement ([Crahay et Ory, 2006](#); [Crahay et al., 2010](#)).

En somme, les enseignants, les parents et les chefs d'établissement semblent attachés au redoublement car ils croient en son efficacité pédagogique. L'évolution des performances de l'élève redoublant l'année de son redoublement, est en général positive, ce qui les conforte dans leurs jugements. Le redoublement s'inscrit, par ailleurs, dans un système de valeurs méritocratiques dans lequel il joue un rôle structurant. Enfin, il a plusieurs fonctions latentes qui expliquent également l'attachement des parties prenantes à cette institution scolaire.

**Le redoublement en France et dans le  
Monde : de l'étude de ses impacts à la  
croyance en son utilité**



---

## Chapitre 1

# Les effets du redoublement : une revue de littérature

La question de l'efficacité du redoublement est peut être le point le plus débattu et le plus controversé de cette thématique. Alors que les opposants considèrent non seulement que répéter une année d'étude n'aide pas un élève à améliorer ses performances, mais qu'en plus cette pratique affecte négativement l'estime qu'il a pour lui-même et augmente ses chances d'abandon, les promoteurs y voient un mécanisme d'incitation puissant et un outil de gestion de classe vertueux. Dans ces conditions, la demande pour une évaluation scientifique des effets du redoublement est extrêmement forte, et ce depuis de nombreuses années.

Jusqu'au début des années 2000, les chercheurs ont ainsi mis en évidence, de façon récurrente, des effets négatifs du redoublement à court et à long terme, notamment sur les résultats scolaires. Cependant, ces études souffraient de plusieurs limites méthodologiques. Depuis, de nouvelles méthodes statistiques et économétriques ont été développées. Elles aboutissent à des résultats plus affinés sur les effets du redoublement qui apparaissent désormais distincts selon qu'on les considère à court et à long terme.

---

Dans la section qui suit, nous synthétisons la littérature que l'on qualifie *d'ancienne* et présentons les principales limites méthodologiques dont les méthodes de cette époque souffraient. La section suivante décrit brièvement l'approche conceptuelle sur laquelle s'appuient les nouvelles méthodes d'évaluation employées et présente les résultats des articles les mettant en œuvre.

## I Un consensus scientifique jusqu'au milieu des années 2000

Jusqu'aux années 2000, les recherches sur les effets du redoublement sont principalement concentrées aux États-Unis. En effet, une vague importante de littérature émerge au milieu des années 1980 lorsque dans plusieurs États, le redoublement est réintroduit et se substitue aux politiques de promotion sociale (ou promotion automatique). L'augmentation du taux de redoublement a souvent été attribuée à la mise en place quasi systématique de tests standardisés (Byrnes et Yamamoto, 1986; United States Department of Education, 1999). En effet, en 1983, la commission nationale sur l'excellence en éducation publie un rapport (National Commission on Excellence in Education, 1983) en faveur d'une réforme du système éducatif et plaide pour la mise en place de tests standardisés afin de mieux piloter le système et les établissements. Ce mouvement insiste sur la définition de standards de compétences à acquérir à chaque niveau. Cette réforme a naturellement conduit à l'abandon dans de nombreux États de la promotion automatique - pratique qui consiste à faire passer un élève même s'il n'a pas acquis les compétences attendues - au profit du redoublement et/ou d'autres formes de remédiation.

Dans la section qui suit, nous présentons les résultats des principales études estimant les effets du redoublements publiées avant les années 2000. Puis, nous discutons des défauts que des travaux plus récents ont mis en exergue et des moyens actuels de les résoudre.

### 1 Les synthèses de littérature jusqu'aux années 2000

Jusqu'au début des années 2000, les chercheurs ont mis en évidence de façon récurrente des effets négatifs du redoublement à court et à long terme, notamment sur les performances des élèves à des tests standardisés. Ces études révèlent que le redoublement n'est pas un moyen de remédiation efficace pour lutter contre les difficultés scolaires. Cependant, ces dernières souffraient de plusieurs limites méthodologiques, ce qui ne permet pas d'apporter de preuves crédibles pour affirmer que le redoublement est plus favorable que la promotion automatique pour permettre aux élèves en difficulté de réussir. En plus des effets sur les résultats scolaires, la recherche s'est intéressée aux effets du redoublement sur d'autres variables telles que la motivation de l'élève, l'estime de soi, l'absentéisme,... De même que pour les performances scolaires, les résultats des études montrent un effet négatif du redoublement sur ces différentes variables.

Les recherches conduites entre 1911 et 1989 et toutes les preuves qu'elles ont accumulées ne plaident pas en faveur du redoublement comme moyen de remédier aux difficultés ou à l'échec scolaire. La première revue systématique de la littérature a été menée par Jackson (1975) sur 30 études publiées entre 1911 et 1973. Dans cette analyse, Jackson classe les études en 3 catégories selon leurs approches *expérimentale*, *comparative pré-post* et *naturelle* (pour une revue plus en détail des résultats de Jackson, voir Jimerson (2001)). Les études *naturelles* comparent les élèves qui redoublent dans des conditions

scolaires normales avec ceux qui sont promus. Les études de *pré-post* tests comparent les performances des redoublants avant et après avoir été promus<sup>1</sup>. enfin, les études expérimentales comparent les élèves en difficulté dont une partie est aléatoirement choisie pour redoubler.

Dans le groupe d'études naturelles, Jackson (1975) reporte 51 analyses (parmi 204) présentant des résultats meilleurs pour les élèves promus (24 pour des mesures de performances scolaires et 27 pour des mesures socio-émotionnelles) et 5 analyses avec des résultats favorables pour les redoublants (2 en performances et 3 mesures socio-émotionnelles). Les 204 études restantes ne trouvent pas de résultats significativement différents entre les deux groupes. Les études comparant les progrès des redoublants (*pré-post*) rapportent des gains pour les élèves redoublants dans 98 analyses parmi les 114 retenues (69 en performances, 29 en gain socio-émotionnels), les autres études sont globalement non significatives. Enfin, les études expérimentales ne s'intéressent qu'aux performances scolaires. elles sont rares (car présentent des problèmes de mise en place et d'éthique notamment) et peu concluantes puisque parmi les 4 études retenues, seule 1 analyse reporte des résultats significatifs en faveur des élèves promus. les 3 autres ne trouvent pas de différences significatives.

Les résultats recensés par Jackson sont donc assez contrastés suivant la méthode employée. Les études expérimentales et naturelles obtiennent en majorité des résultats nuls ou tout du moins non significatifs<sup>2</sup> ou négatifs pour les redoublants. Les travaux employant une méthodologie *pré/post* tests obtiennent majoritairement des progrès de la part des redoublants.

Cependant, ces travaux sont très loin de répondre aux exigences méthodologiques actuelles et ne comparent pas les mêmes choses. Les études dites naturelles comparent *naïvement* les redoublants et les non-redoublants sans tenir compte des différences entre ces groupes alors que les études *pré-post* se focalisent sur les progrès parmi la population des redoublants. Les points de références ne sont donc pas les mêmes. Ainsi, les études *pré-post* tests sont plus à même de trouver un impact positif du redoublement puisque les élèves vont certainement progresser, même si ces gains risquent d'être très inférieurs à ceux des élèves promus avec qui on ne les compare pas. *A contrario*, les études naturelles risquent de trouver davantage d'effets négatifs que le véritable impact du redoublement puisque ces études comparent des élèves *intrinsèquement* différents, notamment en terme de développement cognitif (que des résultats initiaux ne mesurent pas forcément), de motivation, d'ambition, de capacité de concentration, etc. L'écart observé est en partie du au redoublement mais aussi aux caractéristiques de chaque élève associées à la fois à leur probabilité de redoubler et à leur performance scolaire (motivations, intérêt pour l'école,...). Nous reviendrons en détail sur ce problème que l'on qualifie de **biais de sélection**. Enfin, les études expérimentales sont celles qui s'approchent le plus des critères de validité méthodologique

---

1. Par exemple, un élève redoublant le CP passe un test au début de son second CP et plus tard dans son cursus (en CE<sub>1</sub> ou plus tard). La différence de réussite est alors utilisée pour mesurer les progrès.

2. La significativité d'un résultat est la conclusion d'un test rejetant ou non la nullité du paramètre. Il dépend principalement de la taille de l'effet et de la précision de l'estimation. Une différence non significative peut-être due à un effet *réellement* nul précisément estimé ou à un effet dont les données ne permettent pas d'estimer suffisamment finement le paramètre. Dans ce cas, on ne peut ni conclure sur l'absence d'effet, ni sur sa présence.

appliqués aujourd'hui. Leur faible nombre ne permet pas d'être aussi affirmatif dans leurs conclusions que les autres catégories. Néanmoins, leur plus grande rigueur théorique conduit à valoriser davantage leurs résultats. Ainsi, la conclusion de Jackson (1975) est "qu'il n'y a pas un ensemble de preuve fiable pour indiquer que le redoublement est plus favorable que la promotion automatique pour les élèves ayant des difficultés académiques sérieuses ou des difficultés d'adaptation"<sup>3</sup>.

Holmes et Matthews (1984) ont fait une méta-analyse<sup>4</sup> de la littérature sur les effets du redoublement aux États-Unis entre 1929 et 1981, soit 44 études dont 18 utilisant un groupe de contrôle apparié<sup>5</sup> sur des caractéristiques initiales. Cette méthode vise à réduire l'effet des différences initiales entre redoublants et non-redoublants en raisonnant à *caractéristiques données*. Leur analyse révèle des différences significatives en faveur des élèves promus dans tous les champs de comparaison (résultats scolaires en mathématiques, lecture, estime de soi, attitude envers l'école, absentéisme,...) y compris dans les études plus rigoureuses appariant les redoublants et non-redoublants sur des caractéristiques observables avant le redoublement. Ces résultats ont été confirmés par Holmes (1989) qui a conduit une seconde méta-analyse incluant 19 études supplémentaires (soit 63 études). Holmes a rapporté 54 études présentant des résultats négatifs du redoublement à la fois sur des variables mesurant des difficultés d'ajustement socio-émotionnels et des performances scolaires. L'auteur trouve également que les résultats sont plus négatifs lorsque les élèves redoublants et non redoublants sont appariés sur leurs QI ou leurs résultats passés.

Enfin, Jimerson (2001) a synthétisé la littérature scientifique produite entre 1990 et 1999 dans une nouvelle méta-analyse. Parmi les critères d'inclusion, l'auteur s'assure que les résultats s'appuient sur la comparaison d'un groupe de redoublants avec un groupe de contrôle identifiable. Il a ainsi recensé 20 études répondant aux critères d'inclusion qu'il a classées en deux groupes : mesurant l'effet du redoublement sur les performances scolaires ou socio-émotionnelles.

Les 20 études présentent 175 résultats concernant les résultats académiques dont 91 présentent des différences significatives entre les élèves. Parmi elles, 9 sont en faveur du groupe de redoublants, les 82 autres sont en faveur des élèves promus. Les 84 analyses restantes ont des résultats non significatifs. Ainsi, 47 % des études indiquent que le redoublement a un effet négatif et 48 % ne trouvent pas d'effets. Seules 5 % des études obtiennent un effet positif dont en réalité seulement 3 études des résultats après l'année redoublée<sup>6</sup>. L'effet moyen obtenu est de -.39 indiquant que les groupes redoublants ont un score moyen inférieur de 39 % de l'écart-type par rapport aux groupes promus (voir tableau 1.1).

---

3. Jackson (1975) p. 672, traduction de l'auteur.

4. Une méta-analyse est une démarche statistique combinant les résultats d'un ensemble d'études indépendantes sur un sujet de recherche. Elle permet une analyse plus précise des données par l'augmentation du nombre de cas étudiés et de tirer une conclusion globale.

5. L'appariement consiste à identifier des élèves promus ayant certaines caractéristiques importantes identiques (ou presque) à celles d'élèves redoublants, et de comparer leurs trajectoires et évolutions.

6. Les 6 autres études analysent les scores des élèves l'année redoublée.

16 des 20 études présentent 148 analyses de facteurs socio-émotionnels. Il s'agit dans ce cas de mesurer l'effet du redoublement sur l'image de soi des élèves, leur motivation, leur attitude envers l'école,... Il s'agit par exemple de réponses à des assertions du type<sup>7</sup> :

- *Je me sens largement capable d'apprendre ce que l'enseignant nous enseigne ;*
- *Je sens que je peux réussir tout le travail, même si c'est difficile ;*
- *En classe, je travaille, parce que je n'ai pas le choix ;*
- *Je me demande bien ce que je fais à l'école. Si je pouvais je n'irais pas.*

Les résultats dans ce domaine sont plus mitigés puisque 5 % sont significativement en faveur du groupe redoublant (8 analyses), 9 % sont significativement en faveur du groupe de contrôle (13 études) et 86 % ne trouvent pas de différences significatives.

**En somme, cette étude de Jimerson (2001) montre que la comparaison des redoublants avec des non-redoublants aux caractéristiques observables similaires produit dans la majorité des cas des différences non significatives (dans 86 % des analyses de mesures socio-émotionnelles et dans 48 % des mesures de performance scolaire) ou en faveur des élèves promus (dans 47 % des analyses de réussite scolaire et 9 % des mesures socio-émotionnelles).** L'auteur ne conclut toutefois pas (et à raison) à un impact causal et direct du redoublement. Ces résultats reflètent la complexité des interactions entre individualités, milieux et expériences scolaires. Ainsi, *"simplement faire redoubler un élève a peu de chance de résoudre les multiples facteurs qui induisent ses faibles performances ou les orientations qui l'ont conduit à redoubler. [...] Ni la promotion sociale ni le redoublement ne résoudront [les] maux de l'Éducation nationale ni ne faciliteront le succès académique des enfants. Au contraire, on doit attirer l'attention vers des stratégies de remédiation alternatives"* (Jimerson, 2001, p.435).

Les résultats de ces trois méta-analyses ont été synthétisés dans le tableau 1.1. De façon générale, les tailles des effets recensés sont assez comparables, notamment celles de Jimerson (2001) et Holmes et Matthews (1984). Elles sont presque toutes négatives et de magnitude faible à moyenne si on se réfère à la classification de Cohen (1988).

En France, il n'existe à notre connaissance que deux études sur la période considérée : Seibel et Levasseur (1983) pour le redoublement en début de primaire et Grisy (1993) pour le redoublement au début du collège. Dans la première, des élèves de CP passent un test de français et de mathématiques avant la fin de l'année. Parmi ces élèves, les auteurs identifient des élèves qualifiés de "faibles" aux vues de ces résultats, dont une partie est amenée à redoubler, l'autre à passer dans la classe supérieure. Ces mêmes élèves sont ensuite re-testés environ 6 mois plus tard et leurs taux de réussite aux tests sont comparés parmi les élèves faibles ayant redoublé et ceux n'ayant pas redoublé. Dans cette étude, si on considère que les caractéristiques des deux groupes sont comparables et stables dans le temps,

---

7. Ces questions ne sont pas issues des études recensé par Jimerson (2001). Ce sont des exemples tirés des travaux de Cosnefroy et Rocher (2005).

### Taille de l'effet moyen obtenu dans 3 méta-analyses analysant l'effet du redoublement sur les différentes variables d'intérêt

Variabiles mesurées	Jimerson (2001)	Holmes (1989)	Holmes et Matthews (1984)
<b>Taille globale de l'effet</b>	<b>-.31 [246]</b>	<b>-.15 [861]</b>	<b>-.37 [575]</b>
<b>Performances scolaires</b>	<b>-.39 [169]</b>	<b>-.19 [536]</b>	<b>-.44 [367]</b>
Maitrise de la langue	-.36 [11]	.16 [106]	-.40 [85]
Lecture	-.54 [52]	-.08 [144]	-.48 [75]
Mathématiques	-.49 [48]	-.11 [137]	-.33 [77]
Total/score composite	-.20 [13]	NA	NA
Notes	-.18 [45]	-.58 [4]	-.58 [4]
<b>Adaptation socio-émotionnelle</b>	<b>-.22 [77]</b>	<b>-.09 [234]</b>	<b>-.27 [142]</b>
Social	-.08 [12]	-.09 [101]	-.27 [60]
Émotionnelle	-.28 [13]	-.03 [33]	-.37 [9]
Comportementale	-.11 [30]	-.13 [24]	-.31 [13]
Image de soi	-.04 [16]	-.13 [45]	-.19 [34]
Adaptation (composite)	-.15 [4]	NA	NA
Attitude envers l'école	NA	-.05 [39]	-.16 [26]
Absentéisme	-.65 [2]	-.18 [7]	-.12 [6]

Sources : (Jimerson et al., 2006, p.88), traduction de l'auteur

Lecture : Un nombre négatif indique que l'effet mesuré est favorable au groupe de contrôle, c'est-à-dire, au groupe d'élèves promus. Les nombres entre crochets indiquent combien de taille d'effets ont été utilisés pour calculer l'effet moyen reporté.

**Tableau 1.1 – Synthèse des résultats des méta-analyses**

et que hormis le redoublement, il n'y a pas d'autres changements affectant les deux groupes, on peut estimer une valeur *contrefactuelle*, c'est-à-dire, ce qui se serait passé pour les redoublants s'ils avaient été promus. En comparant les scores avec ces valeurs théoriques estimées, on peut ainsi mesurer l'effet moyen du redoublement sur la population des redoublants. En faisant cela, on suppose que, si les élèves qui ont redoublé avaient été promus, ils auraient eu la même progression que ceux, faibles, qui sont réellement entrés dans la classe supérieure. Les résultats sont présentés dans le tableau 1.2.

Ces résultats corroborent bien ceux obtenus dans les articles précédemment cités. Les élèves qui redoublent progressent entre juin et décembre en mathématiques et en français, mais nettement moins vite que les élèves faibles promus dont la progression est ici assez impressionnante (+17.3 % en français et +11.7 % en mathématiques). Si on suppose que les élèves qui ont redoublé ont eu la même progression, alors le redoublement réduit leurs scores de 13.1 % en français et 8.6 % en mathématiques<sup>8</sup>. Ces résultats ne corrigent pas pour autant tous les biais. Il est d'ailleurs probable que la comparaison entre les deux

8. Ces résultats ne sont pas présentés par l'auteur de l'étude, ils relèvent de la littérature de l'inférence causale initiée par l'article de Rubin (1974) qui s'est beaucoup développée depuis. Il s'agit d'un estimateur en *double différences*. Il repose sur l'hypothèse que les différences entre deux groupes (traité et contrôle) ne varient pas dans le temps et que la façon dont le traitement (ici le redoublement) est administré n'a pas d'influence sur les résultats potentiels. Ainsi, si la seule variation entre deux groupes à deux dates différentes est le traitement alors on peut supposer que le groupe traité aurait progressé de la même façon que le groupe de contrôle en l'absence de traitement. En mesurant la différence de la variation des scores (avant-après) entre les groupes, on obtient l'indicateur de double différences (ou différence en différence).

### Synthèses des taux de réussite des élèves aux tests standardisés dans l'étude de Seibel et Levasseur (1983)

groupe d'élèves	résultats en Juin	résultats en décembre	gains
Français			
"forts"	86.5 %	84.0 %	-2.5 %
"moyens"	69.5 %	75.5 %	5.6 %
faibles promus	36.3 %	53.6 %	17.3 %
faibles redoublants	33.4 %	37.5 %	4.1 %
	Effet du redoublement		-13.1 %
Maths			
"forts"	98.2 %	99.1 %	0.9 %
"moyens"	77.9 %	85.9 %	8 %
faibles promus	51.1 %	62.8 %	11.7 %
faibles redoublants	47.8 %	51.1 %	3.1 %
	Effet du redoublement		-8.6 %

Sources : *Seibel et Levasseur (1983) construit à partir du graphique V page 16*

**Tableau 1.2 – Résultats principaux de l'étude de Seibel et Levasseur (1983)**

groupes de redoublants et non redoublants ne soit pas valide car les non-redoublants possèdent des caractéristiques inobservables pour le statisticien qui affectent à la fois leur probabilité de redoubler et leurs progression. En présence de tels caractéristiques, l'estimation précédente est biaisée car l'hypothèse de même progression en l'absence de redoublement n'est pas plausible.

Grisay (1993) aboutit aux mêmes résultats pour le collège : un élève qui redouble progresse bel et bien durant sa seconde année, cependant, il progresse nettement moins vite qu'un élève qui a les mêmes caractéristiques *observables* mais qui passe dans la classe supérieure.

En somme, la littérature scientifique jusqu'aux années 2000 aboutit à un consensus fort concernant les effets du redoublement. Au mieux, ce dernier n'a pas d'effet mais il s'avère dans bien des cas nocif pour la réussite scolaire des élèves et pour le développement de leur estime de soi, et bien d'autres mesures non cognitives. Toutefois, cette littérature souffre de nombreuses faiblesses méthodologiques que nous allons à présent décrire.

## 2 Les limites méthodologiques de ces études

Des articles plus récents ont mis en exergue des faiblesses méthodologiques dans les articles plus anciens. Les résultats obtenus sont en effet affectés par le moment auquel le chercheur estime l'effet du redoublement (immédiatement après, plusieurs années après,...), s'il compare les élèves à âge égal ou à niveau égal, si les mesures utilisées sont comparables et si les groupes sont suffisamment proches pour être comparés. Les études plus rigoureuses méthodologiquement obtiennent des effets du redoublement systématiquement non significatifs comparés aux études précédentes. Plus l'année où sont mesurés les résultats est éloignée du redoublement, plus ses effets sont importants et négatifs. Les estimations de l'effet du redoublement sont différentes lorsque les redoublants et non-redoublants sont comparés à âge égal ou à niveau égal. Dans le premier cas, les résultats à court terme sont plutôt négatifs alors que dans le second cas les effets à court terme sont plutôt positifs ou nuls.

Les études que nous venons de présenter sont souvent citées dans la littérature comme ayant des preuves concluantes sur les effets du redoublement. Cependant, comme nous l'avons évoqué dans la section précédente, peu d'études recensées par [Holmes et Matthews \(1984\)](#), [Holmes \(1989\)](#) et [Jimerson \(2001\)](#) remplissent les critères actuels de rigueur méthodologique dans l'étude d'impact. De plus, les études plus récentes (nous le verrons plus loin) fournissent souvent des résultats plus contrastés.

Pour commencer, [Allen et al. \(2009\)](#) font une méta-analyse des études publiées entre 1990 et 2007 en Amérique du Nord sur l'effet du redoublement sur les performances scolaires. L'apport principal de ce papier - au delà du fait qu'il présente des résultats plus récents - est de s'intéresser à la rigueur de la méthodologie employée dans les articles recensés pour déterminer l'impact du redoublement. Ils s'intéressent également au moment où l'analyse est menée (i.e. combien de temps après le redoublement) et à la sélection du groupe de contrôle. Ils montrent alors que les études qui utilisent une méthodologie plus rigoureuse obtiennent des effets du redoublement systématiquement non significatifs et en moyenne .34 écart type plus élevés (moins négatif) que les études de moins bonne qualité. Si on reprend les effets moyens recensés dans les autres méta-analyses synthétisés dans le tableau 1.1, l'effet moyen passe donc de -.31 à .03 lorsque la méthodologie est plus rigoureuse. Ainsi, lorsque les chercheurs tentent de réduire les différences préexistantes entre le groupe de redoublants et le groupe de non redoublants avec lequel est faite la comparaison, l'effet du redoublement est proche de 0. Par ailleurs, les auteurs montrent que plus l'analyse est menée longtemps après l'année redoublée, plus l'effet du redoublement est important et délétère. L'impact négatif du nombre d'année après le redoublement est plus fort pour les études qui comparent les résultats à niveau égal<sup>9</sup> qu'à âge égal<sup>10</sup> (voir la section suivante).

Au final, l'étude montre que lorsque les chercheurs contrôlent davantage pour les différences pré-

---

9. Qui comparent les redoublants avec les élèves qui sont de le même niveau scolaire qu'eux, et donc plus jeunes.

10. Qui comparent les redoublants avec les élèves de la même génération qui sont donc dans un niveau scolaire plus élevé.

existantes entre redoublants et non-redoublants, l'impact n'est pas aussi négatif. Il y a trois facteurs principaux expliquant ce résultat (Roderick et Nagaoka, 2005) :

- a) le moment auquel le chercheur estime l'effet sur la performance ;
- b) la comparabilité des résultats des tests ;
- c) la capacité des chercheurs à construire un groupe de comparaison valide.

L'analyse de Allen et al. (2009) nous a déjà fourni quelques résultats pour le premier et le dernier point. Nous allons les décrire plus en détail dans les sous-sections qui suivent.

### **A Les comparaisons à âge égal ou à niveau égal**

Les comparaisons à âge égal étudient l'effet du redoublement sur les performances des élèves (ou tout autre paramètre) à un âge donné (par exemple 15 ans). Dans cette configuration, les élèves qui ont redoublé seront dans une classe inférieure par rapport à leurs pairs qui ont été promus. Au contraire, les comparaisons à niveau égal mesurent l'effet du redoublement en calculant l'écart de performance entre les redoublants et les élèves, plus jeunes, qui se trouvent dans ce niveau.

Il y a plusieurs raisons pour lesquels les estimations de l'effet du redoublement peuvent différer entre ces deux méthodologies. La plus évidente est qu'un élève qui a redoublé connaît déjà le programme et a donc un avantage sur ses pairs, plus jeunes, l'année de son redoublement<sup>11</sup>. La seconde est qu'un élève plus âgé sera plus mature, ce qui peut lui permettre de mieux appréhender certains concepts (l'abstraction,...). Ainsi, les comparaisons à niveau égal peuvent donner un avantage au redoublant l'année répétée. Au contraire, les comparaisons à âge égal mesurent les différences de performances entre deux élèves dans des niveaux différents. Or les enfants promu dans le niveau supérieur apprennent de nouveaux concepts et renforcent ceux déjà acquis, alors que les redoublants répètent, plus ou moins à l'identique, le programme de l'année passée. On comprend alors que les élèves promus, même faibles, ont plus de chances de réussir à un test commun que les élèves qui ont redoublé.

Les estimations de l'impact du redoublement sont effectivement différentes lorsque les redoublants et non-redoublants sont comparés à *âge égal* ou à *niveau égal*. Dans le premier cas, les résultats à court terme sont plutôt négatifs alors que dans le second cas - c'est-à-dire lorsqu'on compare les résultats d'élèves ayant effectués deux fois une classe avec ceux ne l'ayant fait qu'une fois - les effets à court terme sont plutôt positifs ou nuls (Holmes, 1989; Roderick et Nagaoka, 2005).

Par ailleurs, la question de recherche derrière ces deux comparaisons n'est pas exactement la même. Les comparaisons à âge égal testent si deux années passées dans le même niveau induisent une progression plus forte que deux ans passés dans des niveaux différents (la seconde année étant dans une classe supérieure). Lorsqu'on compare des élèves à niveau égal, on s'intéresse plutôt au fait de savoir si les élèves en difficulté apprennent davantage en deux ans qu'en un an et s'ils parviennent à rattraper leur retard à l'issue de cette deuxième année.

---

11. Cet avantage est naturellement l'un des buts du redoublement.

Ces deux questions sont intéressantes et pertinentes ; néanmoins le choix de la méthode a une influence forte sur l'extrapolation possible des résultats. Pour Roderick et Nagaoka (2005), la question de l'impact causal du redoublement ne peut se mesurer qu'avec une comparaison à âge égal. L'effet du redoublement se mesure en comparant les résultats potentiels pour un même individu lorsque celui-ci redouble ou lorsqu'il passe dans la classe supérieure. On ne peut estimer la valeur contre-factuelle<sup>12</sup> qu'en comparant avec un groupe d'élèves *statistiquement identiques* qui ont été promus. La comparaison à niveau égal interroge le bénéfice d'une année d'étude supplémentaire, mais le redoublement n'est qu'un moyen d'ajouter une année d'étude parmi d'autres. On pense en particulier aux parcours de transition qui répartissent les études sur une année supplémentaire pour permettre aux élèves d'apprendre plus lentement comme c'est le cas au lycée en Finlande (Jost, 2007). La question la plus pertinente dans ce cas est plus de se demander si le redoublement est la meilleure utilisation d'une année d'étude supplémentaire. À notre connaissance, il n'y a pas de travaux comparant deux formes de remédiation de la sorte.

En outre, si les comparaisons à niveau égal produisent des résultats plutôt positifs à court terme, l'analyse de Allen et al. (2009) montre que plus la mesure est faite longtemps après le redoublement, plus l'effet est négatif, ce qui est encore plus marqué pour les comparaisons à niveau égal. Ce résultat indique que les élèves redoublant bénéficient d'un léger avantage l'année de leur redoublement lorsqu'on les compare aux élèves de même niveau (donc plus jeune). Cependant, cet avantage disparaît l'année suivante. Il s'ensuit alors une trajectoire descendante pouvant entraîner un désengagement de l'élève, voire un décrochage. Au contraire, les comparaisons à âge égal montrent un effet plutôt négatif initialement (car l'élève redoublant est en moyenne plus faible), mais qui ne se détériore pas aussi vite.

Si la comparaison à âge égal permet de mieux répondre à la question de l'effet causal du redoublement, sa mise en place nécessite de dépasser un problème d'organisation et de validité interne crucial : la comparabilité des résultats.

## **B La comparabilité des résultats des tests entre les niveaux**

Lorsqu'on souhaite comparer des élèves ayant redoublé avec des élèves ayant été promus, il est crucial que les deux groupes, pourtant dans des niveaux scolaires différents, se voient administrer le **même** questionnaire à la même date (ou à peu près). La raison principale étant que les programmes scolaires sont différents selon les niveaux et que les tests passés ne sanctionnent pas les mêmes acquis de compétences.

Ce type d'organisation est malheureusement très peu courant puisqu'il est très compliqué d'organiser de tels évaluations. L'enquête PISA de l'OCDE fait exception en évaluant les compétences des élèves à 15 ans, peu importe la classe dans laquelle ils se trouvent.

Si les tests ne sont pas identiques (même s'ils sont standardisés), les estimations de l'effet du

---

12. Ce qui se serait passé en l'absence de redoublement.

redoublement risquent d'être très influencées par les différences entre les items dans la comparaison des scores et des gains en termes de progression entre les différents niveaux. Par exemple, [Reynold \(1992\)](#) estime l'effet du redoublement en comparant les résultats de redoublants en classe équivalente au CE<sub>2</sub><sup>13</sup> avec ceux d'élèves de niveau équivalent au CM<sub>1</sub><sup>14</sup>, à partir de tests standardisés à Chicago<sup>15</sup>. Or, ces tests ne peuvent pas servir à mesurer les progrès d'un élève ou comparer les performances entre les niveaux car les scores ne sont pas comparables directement ([Roderick et Nagaoka, 2005](#)). Pourtant, [Reynold \(1992\)](#) obtient des résultats significativement négatifs pour les redoublants et conclut que le redoublement a un effet délétère sur les performances scolaires, sans discuter la comparabilité de ces résultats. De façon similaire, [Alexander et al. \(1994\)](#) estiment l'effet du redoublement sur le test standardisé californien<sup>16</sup> et obtiennent un impact positif du redoublement. Cependant, ces tests souffrent du même problème que les tests à Chicago : ils ne sont pas conçus pour comparer les niveaux. Conscients de ce problème, [Shepard et al. \(1996\)](#) ont reproduit les travaux de [Alexander et al. \(1994\)](#) en convertissant les résultats de façon à rendre les tests comparables. Ils trouvent également un rebond après une année redoublé, mais les résultats obtenus par [Alexander et al. \(1994\)](#) étaient réellement surestimés. Au final, le problème de comparabilité des tests influence bel et bien les estimations, mais en plus, le sens du biais de comparabilité dépend du test utilisé pour l'analyse.

Ainsi, une partie de la littérature fournit des résultats différents car elle compare à niveau égal ou à âge égal, et utilise des tests qui ne sont pas toujours comparables. Il se peut enfin que même en utilisant des tests comparables, les groupes de comparaisons ne soient pas suffisamment proches.

### C La construction des groupes de comparaison

Les études diffèrent de façon significatives dans leur capacité à construire un groupe de comparaison adéquat et à résoudre le problème de sélection qui affecte les estimations de l'effet du redoublement. Cette différence est certainement la source la plus importante de biais affectant les différents résultats.

En effet, la comparaison des performances des élèves redoublants et non redoublants mesure d'une part, l'effet du redoublement (ce que le chercheur veut identifier) et d'autre part, l'impact des différences entre les élèves. En effet, ces deux groupes possèdent des caractéristiques (observables ou non) qui influencent à la fois leurs résultats scolaires et la probabilité qu'ils redoublent. Ainsi, les élèves en difficulté ont en générale des résultats plus faibles et ont plus de chances de redoubler. Ces différences créent un **biais de sélection** qui, en l'absence de méthode de correction, fait apparaître les redoublants plus faibles que les promus.

Pour corriger ce biais de sélection, les chercheurs utilisent en général trois sortes de stratégie ([Grisay, 2001](#)) :

---

13. 3<sup>rd</sup> graders, pour 3<sup>ème</sup> année d'école élémentaire.

14. 4<sup>th</sup> graders, pour 4<sup>ème</sup> année d'école élémentaire.

15. tests ITBS exprimé en *équivalent de niveau*, c'est-à-dire par rapport à la norme nationale attendue pour un niveau donné.

16. *California Achievement test*, dont le test de CM<sub>1</sub> est réputé plus difficile que celui de CE<sub>2</sub>

1. Une stratégie expérimentale pure ;
2. Une stratégie expérimentale en ligne de base ;
3. Une stratégie quasi-expérimentale.

La stratégie expérimentale pure est souvent considérée comme le "*gold standard*" de l'évaluation. Elle consiste à allouer de façon aléatoire un traitement dans un échantillon représentatif. Le tirage aléatoire résout le problème du biais de sélection puisque le traitement est indépendant des caractéristiques des individus. Si le tirage est satisfaisant, les deux groupes (traités et non-traités) sont statistiquement identiques et la seule différence entre ses participants est le statut (traité/non-traité). Il est cependant très compliqué de mettre en place une expérimentation de la sorte, pour des raisons organisationnelles mais aussi très souvent éthique (il n'est moralement pas admissible de faire redoubler au hasard des élèves).

Dans l'encadré ci-dessous, nous définissons la notion d'effet du redoublement et de biais de sélection.

#### Une définition simple de la causalité inspirée de Rubin (1974)

Supposons qu'un chercheur s'intéresse à l'effet du redoublement sur la moyenne générale au brevet. Il dispose des scores de chaque élève  $y_i$  et d'une variable  $R_i$  qui vaut 1 s'il a redoublé, et 0 sinon. En théorie, il existe pour chaque personne une valeur réelle, correspondant à sa situation vis-à-vis du redoublement ; et une valeur *contrefactuelle*, qui indique ce qui se serait passé dans la situation opposée. Ainsi, chaque élève a virtuellement deux résultats possibles au brevet :

$y(1)$  S'il a redoublé

$y(0)$  S'il n'a jamais redoublé.

L'effet du redoublement est la différence entre ces deux valeurs. Cependant, on ne peut observer qu'une valeur pour chaque individu : celle qui lui est réellement arrivée. Concrètement, un même élève ne peut pas à la fois redoubler et être promu. L'effet du redoublement est donc inobservable et nécessite, pour l'estimer, d'avoir recours à une comparaison des moyennes entre des élèves ayant redoublé et d'autres ayant toujours été promu.

En notant, de façon très simple  $Y|_R$  la moyenne au brevet des élèves ayant redoublé et  $Y|_P$  la moyenne des élèves ayant toujours été promu, on peut calculer la différence  $D$  :

$$D = Y|_R - Y|_P$$

Chaque moyenne estime correctement la moyenne théorique correspondante et donc :

$$D = Y(1)|_R - Y(0)|_P$$

Cependant, cette différence ne mesure pas l'effet causal du redoublement sur la moyenne des notes au brevet car les valeurs virtuelles de chaque groupe ne sont pas les mêmes. En effet, les redoublants et non-redoublants possèdent des caractéristiques inobservables (capacités cognitives et caractéristiques conatives, intérêt pour l'école, motivation,...) différentes qui modifient la distribution des notes potentielles théoriques définies précédemment.

Pour s'en convaincre, on ajoute et soustrait (pour respecter l'égalité de l'équation) la moyenne théorique du groupe des redoublants **qu'ils auraient eu s'ils n'avaient pas redoublé** :  $+Y(0)|_R - Y(0)|_R$ . On les ajoute à l'équation précédente :

$$D = Y(1)|_R - Y(0)|_P + Y(0)|_R - Y(0)|_P$$

En re-arrangeant les termes dans l'équation, on obtient :

$$D = \underbrace{Y(1)|_R - Y(0)|_R}_{\text{Effet du redoublement sur les redoublants}} + \underbrace{Y(0)|_R - Y(0)|_P}_{\text{Biais de sélection}}$$

Ainsi, en comparant la moyenne des résultats des redoublants et non-redoublants, on mesure bien l'effet du redoublement, mais cette valeur est *polluée* par le biais de sélection. Celui-ci correspond au fait que le redoublement n'est pas indépendant des résultats potentiels au brevet.

Comme les élèves en difficulté ont plus de chance de redoubler, les redoublants ont en moyenne des résultats potentielles  $Y(0)|_R$  plus faible que les non-redoublants ( $Y(0)|_P$ ).

Ainsi,  $Y(0)|_R - Y(0)|_P$  est négatif, ce qui tire la comparaison entre les redoublants et non-redoublants vers des valeurs négatives, faisant croire à un effet délétère du redoublement sur les notes au brevet. Si on ne corrige pas le biais de sélection, on peut être amené à conclure, à tort, que le redoublement a des effets négatifs alors qu'en réalité, le biais de sélection est en (grande) partie responsable de ce résultat.

Les stratégies expérimentales pures font redoubler au hasard des élèves, peu importe leurs niveaux ou caractéristiques. De ce fait, il n'y a aucune raison pour que les résultats potentiels des deux groupes soient différents. En somme,  $Y(0)|_R = Y(0)|_P$ , la comparaison des moyennes des deux groupes tirés au hasard mesure directement l'effet du redoublement.

La stratégie expérimentale en ligne de base ne comporte pas de groupe de contrôle, mais une variation délibérée dans l'intensité de l'intervention. L'objectif est alors de vérifier si les effets attendus coïncident dans le temps avec la mise en œuvre de l'innovation (s'ils s'accroissent lorsqu'on l'intensifie, cessent ou s'atténuent quand on la suspend, etc.). Cette méthode n'est pas applicable à la question du redoublement directement mais pourrait être appliquée dans la comparaison de programmes de remédiation alternatifs (la durée de cours d'été par exemple).

Enfin, les études quasi-expérimentales exploitent l'existence *naturelle* de groupes cibles affectés par un traitement ou une politique et comparent leurs résultats à ceux de l'ensemble de la population à laquelle ils appartiennent, en tenant sous contrôle les différences entre les groupes par des techniques statistiques.

Dans le cas du redoublement, la très grande majorité des études s'inscrit dans la catégorie des études quasi-expérimentales. Cependant, les méthodes statistiques employées pour corriger le biais de sélection ne se valent pas toutes et dépendent du contexte et des données disponibles.

Il existe plusieurs techniques qui permettent, sous des hypothèses raisonnables, de corriger au moins en partie le biais de sélection. Cependant, toutes ne sont pas utilisables en toutes circonstances et toutes ne permettent pas d'inférer les résultats obtenus sur l'ensemble de la population. Nous présentons ces méthodes dans l'annexe I.

## II Les effets du redoublement revisités

Nous distinguons les articles qui identifient les effets du redoublement suivant le niveau redoublé et combien de temps après les effets sont mesurés. Pour chacun des articles présentés, nous apportons une modeste critique de façon à faire apparaître les limites des différents résultats.

### 1 Les effets du redoublement en maternelle

Les effets du redoublement en maternelle semblent positifs à court terme sur les performances scolaires, lorsqu'on tient compte des caractéristiques inobservables des élèves. Ces dernières jouent en effet un rôle déterminant sur la probabilité de redoubler et le niveau scolaire. Les effets se dissipent toutefois rapidement. Néanmoins, une partie des études montrent que les élèves ayant redoublé en maternelle auraient eu de meilleurs résultats dans la suite de leur scolarité s'ils n'avaient pas redoublé.

À notre connaissance, il n'existe pas d'étude caractérisant les effets du redoublement en maternelle ailleurs qu'aux États-Unis (à l'exception notable du rapport de [Lafontaine et Monseur \(2011\)](#) qui caractérise le redoublement de la maternelle en Belgique francophone et décrit les trajectoires des élèves sur plusieurs années. Ce travail ne s'intéresse toutefois pas à l'impact *causal* du redoublement au sens où il est entendu dans cette section). Ce vide est notamment dû à l'absence de données concernant des élèves de si bas âge et en particulier, à l'absence de données de panel. Aux États-Unis, il existe plusieurs études sur l'impact à court terme du redoublement de la maternelle (*Kindergarten*) telle que [Hong et Raudenbush \(2006\)](#), [Dong \(2010\)](#) et [Fruehwirth et al. \(2011\)](#). Ces trois travaux ont la particularité d'être très innovants dans leurs façon de traiter le biais de sélection et en un sens, chacun contribue au développement de nouvelles méthodes d'estimation.

En effet, [Hong et Raudenbush \(2006\)](#) remarquent que la littérature sur le redoublement suppose en général que l'effet sur les performances scolaires ne dépend pas des pratiques dans l'établissement, notamment si beaucoup d'élèves redoublent ou seulement un nombre très restreint. Or, les auteurs sont convaincus que cette hypothèse n'est pas crédible. Il est en effet possible que les conséquences du redoublement soient différentes si beaucoup d'élèves de l'établissement y sont sujets ou non. Si tel est le cas, un modèle estimant la différence entre redoublants et promus, même tirés au hasard, mesurera un impact "*pollué*" par les effets de pairs associés à la distribution du redoublement. Ils développent alors un modèle théorique permettant d'identifier - entre autres - l'effet du redoublement parmi les écoles à faible redoublement et à fort redoublement ainsi que l'effet d'être dans un établissement à fort redoublement par rapport à un établissement le pratiquant peu. Comme ils ne disposent pas de données expérimentales, les auteurs appariant les élèves et les écoles de chaque groupe de façon à rendre virtuellement identiques les élèves, sur la base de leurs caractéristiques observables.

**Hong et Raudenbush (2006)** montrent alors que les enfants qui redoublent en maternelle auraient obtenu de meilleurs résultats en mathématiques et en lecture 1 an après s'ils n'avaient été retenus.

Ce résultat est valable (mais variable) dans les écoles à faible et fort taux de redoublement. Les auteurs concluent que la politique de rétention en maternelle ne produit pas d'effets bénéfiques aux élèves, même à très court terme. Les résultats de cet article sont donc conformes à la littérature existante que nous avons déjà recensée. Son apport principal est de fournir un cadre conceptuel moins contraint que d'ordinaire. Cependant, le problème de sélection est corrigé par l'appariement, ce qui ne solutionne pas le biais dû aux caractéristiques inobservables.

L'article de **Dong (2010)** en tient compte en développant une méthode économétrique assez sophistiquée. Elle dispose d'un échantillon représentatif d'élèves de maternelle aux États-Unis qu'elle observe plusieurs années successives. Dans ce travail, l'auteur pense que deux éléments inobservables rendent inefficaces les méthodes d'estimation standards. D'une part, les parents peuvent choisir des maternelles dans lesquelles le redoublement n'est pas, ou peu pratiqué. Les élèves des écoles à fort redoublement peuvent donc être différents des élèves des écoles à faible taux de redoublement en raison du choix des parents et de leurs caractéristiques, introduisant un premier biais de sélection. D'autre part, le redoublement peut être déterminé par les *talents* inobservables (ou leur absence) des élèves qui affectent également leurs résultats scolaires futurs. La méthode qu'elle utilise tient compte de ces deux biais en employant notamment des *variables instrumentales*.

Comme les États américains ont des réglementations différentes sur l'âge d'entrée à l'école primaire, à un mois de naissance donnée, un élève pourra être retenu en maternelle alors qu'il ne le serait pas dans un autre État. Aussi, en comparant des enfants de même âge entre plusieurs États, on compare des enfants plus ou moins souvent redoublants (du fait de la législation), mais qui n'ont pas de raison d'être très différents par ailleurs. Cet écart de législation peut permettre d'évaluer les effets du redoublement sans les confondre avec le potentiel scolaire différent des redoublants et des non-redoublants. Une variable de ce type est appelé un *instrument* ou une *variable instrumentale*. Cette stratégie a d'ailleurs été exploitée de nombreuses fois pour répondre à d'autres questions de recherche, et notamment le rendement de l'éducation (**Angrist et Krueger, 1990**).

L'effet du choix de l'école maternelle est plus difficile à neutraliser car il faut trouver au moins un élément indépendant de la famille qui incite les parents à choisir une école plutôt qu'une autre. **Dong (2010)** choisit le mode de transport pour se rendre à l'école d'une part, et la possibilité de redoubler plusieurs fois dans cette école d'autre part. Dans le premier cas, on fait l'hypothèse que les parents n'ont pas choisi leur lieu d'habitation en fonction des pratiques de redoublement dans les écoles maternelles environnantes. Tout se passe donc comme si les pratiques de redoublement des écoles situées à proximité du lieu d'habitation des parents étaient le fruit du hasard. Dès lors, des parents contre le redoublement seront d'autant plus enclins à mettre leur enfant dans une école qui le pratique si elle est beaucoup plus facile d'accès. Le second instrument est la possibilité de redoubler plusieurs fois dans l'école élémentaire.

L'argument pour cette variable est que les écoles qui acceptent de faire redoubler plusieurs fois un élève ont une "*philosophie du redoublement*" qui détermine leurs pratiques en maternelle mais n'influence pas directement la probabilité de redoubler d'un élève.

Ces variables ne sont valables que si elles n'influencent pas le redoublement autrement que par l'auto-sélection des parents dans ces écoles. Cette hypothèse peut être mise en doute (ce qu'admet volontiers l'auteur). Elle défend ses choix en disant que les deux *instruments* sont discutables pour des raisons différentes et que puisque les résultats sont insensibles à l'utilisation de l'un, l'autre ou les deux à la fois, les résultats sont robustes.

**Dong (2010) montre ainsi que les élèves redoublants ont des résultats environ 14 % plus élevés en CP en mathématiques et en lecture que s'ils avaient été promus directement. L'effet sur la lecture tombe à 0 en CE<sub>2</sub> mais reste significatif et positif en mathématiques.** Les conclusions de cet article rejoignent la littérature sur l'éducation intensive en bas âge (Heckman et al., 2013). Ces résultats sont aussi proches de McEwan et Shapiro (2008) qui utilisent une *discontinuité de régression*<sup>17</sup> à partir des dates de naissances des jeunes chiliens pour montrer qu'entrer à l'école primaire avec un an de retard (en ayant redoublé pour des raisons scolaires ou non) a des effets positifs sur les résultats dans des classes équivalentes au CM<sub>1</sub> ainsi qu'en 4<sup>ème</sup> et un effet négatif sur la probabilité de redoubler le CP. On note toutefois que l'utilisation des dates de naissances dans ce contexte est un sujet controversé car le mois de naissance affecte la maturité, qui a un effet direct sur les résultats. Ce qui est mesuré par l'auteur est alors un effet confondu du redoublement et de la maturité. Par ailleurs, les naissances ne sont pas tout à fait aléatoirement distribuées (voir notamment Grenet (2010) pour une discussion sur les données françaises). Ce sujet sera débattu plus en détail dans les sections suivantes.

Ainsi, la prise en compte des différences inobservables, qui influent sur la probabilité de redoubler et sur les performances scolaires, montre que le redoublement de la maternelle peut être bénéfique à court terme aux États-Unis.

En utilisant les mêmes données que Dong (2010), Fruehwirth et al. (2011)<sup>18</sup> développent un modèle structurel visant à mesurer l'effet du redoublement à différents moments de la scolarité. Pour ces auteurs, le redoublement n'est pas un traitement binaire puisqu'il peut intervenir à différents moments de la scolarité. Cela implique une potentielle sélection dynamique des élèves redoublants, où les *moins doués* sont maintenus en premier,... Dès lors, l'effet du redoublement dépend du niveau auquel il survient, et des *talents cachés*<sup>19</sup> des élèves.

**Fruehwirth et al. (2014) montrent alors que les élèves maintenus en maternelle aurait eu des résultats 27 % plus élevés s'ils n'avaient pas redoublé l'année d'après, mais ces effets diminuent avec le temps. A 11 ans, ils auraient des performances 7 % plus élevées s'ils n'avaient pas redoublé.** Les auteurs montrent par ailleurs qu'en raison de différences de caractéristiques inob-

---

17. Nous présentons plus en détail cette méthode dans la section suivante.

18. Il existe une nouvelle version de cette étude retravaillée pour publication : Fruehwirth et al. (2014).

19. Capturés par des effets fixes dans les modèles économétriques.

servables entre redoublants et non-redoublants, ces derniers auraient pu avoir de meilleurs résultats s'ils avaient redoublé. Ce résultat n'est évidemment pas un argument en faveur du redoublement. Les auteurs expliquent que "*l'effet positif est obtenu car les élèves les plus performants, s'ils avaient redoublés, auraient reçu davantage de ressources parentales et scolaires en raison du redoublement que les redoublants moins performants*" (Fruehwirth et al., 2014, p. 4).

Au final, ces quelques articles n'aboutissent pas nécessairement à des résultats convergents, sur tous les aspects. On peut toutefois conclure que l'effet du redoublement en maternelle peut être positif à court terme, mais que ces effets diminuent assez rapidement. Les caractéristiques inobservables jouent un rôle crucial et semblent fortement influencer la probabilité de redoubler et les performances des élèves. Le redoublement est un phénomène dynamique qui ne traite pas les mêmes élèves aux différents niveaux et a donc un effet spécifique à chaque classe.

## 2 Les effets du redoublement au primaire

Les articles étudiant les effets du redoublement au primaire sont très nombreux, mais sont également très spécifiques aux systèmes scolaires qu'ils étudient. Ainsi, une partie importante de la littérature étudie les effets à court et moyen terme du redoublement lorsque celui-ci est décidé sur la base de résultats à un test standardisé. Ces études convergent essentiellement vers les mêmes conclusions : le redoublement a des effets positifs ou nuls à court terme sur les performances scolaires pour les élèves en difficulté. Cependant, ces élèves peuvent bénéficier d'autres modes de remédiation en complément. Ainsi, le redoublement peut-être bénéfique au primaire s'il est accompagné d'autres dispositifs d'aide. Les effets se dissipent toutefois très rapidement (dans la plupart des cas deux ans plus tard) et tendent à devenir négatifs à moyen terme. Plusieurs autres études obtiennent des effets nuls sur les performances scolaires mais généralement négatifs sur les mesures psycho-sociales.

La plupart des études qui présentent les effets du redoublement au primaire se concentrent sur les effets à court terme et à moyen terme d'un redoublement précoce (i.e. CP ou CE<sub>1</sub>). La littérature est nettement plus abondante à ce niveau et les méthodes utilisées sont également plus variées.

Aux États-Unis, la mise en place du redoublement dans les années 1990 comme alternative à la promotion automatique s'est accompagnée de nombreux travaux visant à en mesurer les effets. La spécificité du redoublement dans certains États réside dans l'existence d'évaluation standardisées qui déterminent la promotion ou la rétention des élèves. Le cas de l'État de Chicago est particulièrement étudié dans la littérature (Allensworth, 2004; Nagaoka et Roderick, 2004; Jacob et Lefgren, 2004; Roderick et Nagaoka, 2005; Jacob et Lefgren, 2007).

En effet, les élèves de cet État passent des tests standardisés au mois de juin en lecture et en mathématiques dans des classes équivalentes au CE<sub>2</sub>, à la 6<sup>ème</sup> et à la 4<sup>ème</sup>. Ceux qui obtiennent un

score supérieur à un certain niveau défini à l'avance dans les deux disciplines sont promus. Ceux qui ne répondent pas aux exigences de performances doivent suivre des cours d'été qui durent 6 semaines. A la fin de l'été, ces élèves passent un second test dans les deux matières. Ceux qui échouent à nouveau à l'examen d'août doivent redoubler, les autres sont promus.

Les notes charnières aux examens standardisés créent une frontière qui sépare arbitrairement des élèves très proches. En effet, comme toute forme d'évaluation mesure de façon imparfaite et, avec une part d'aléa, les compétences des enfants, deux élèves ayant des notes très proches ont de bonnes chances de beaucoup se ressembler. Pourtant, dans cette situation, l'un sera juste en dessous du seuil et va redoubler, l'autre sera promu sans condition. Ainsi, tout se passe comme si, du fait d'aléas de notation inévitables, des élèves à la marge des performances attendues avaient été tirés au sort pour redoubler et d'autres promus. Entre ces élèves juste au dessus et juste en dessous, la seule différence est le redoublement. Cette frontière arbitraire permet donc de mesurer l'effet du redoublement pour des élèves similaires ayant un niveau à la marge des exigences de performances attendues. On parle alors d'effet *local* (Angrist et Imbens, 1994; Angrist et Pischke, 2008).

Le contexte des évaluations standardisées de Chicago est un peu plus compliqué puisqu'il existe deux discontinuités, au mois de juin et au mois d'août. La première discontinuité permet d'identifier l'effet du redoublement couplé aux cours d'été par rapport à la promotion. La seconde discontinuité qui compare, parmi les élèves ayant raté la première évaluation, ceux qui ont réussi ou échoué de justesse la seconde, indique l'effet *net* du redoublement sur les élèves ayant suivi les cours d'été.

Jacob et Lefgren (2004) exploitent ces discontinuités et montrent que les élèves qui ont redoublé et suivi des cours d'été à l'issue du niveau équivalent au CE<sub>2</sub>, réussissent mieux à la fois en lecture et en mathématiques que s'ils avaient été promus un an plus tard. L'effet est toujours positif et significatif dans les deux disciplines un an plus tard mais nettement moins élevé. En revanche, l'effet est nul pour les élèves qui redoublent le niveau équivalent à la 6<sup>ème</sup>, même l'année qui suit le redoublement.

En restreignant l'échantillon aux élèves ayant eu des cours d'été et en utilisant la discontinuité induite par le test au mois d'août, les auteurs montrent que les élèves qui ont redoublé réussissent mieux que s'ils avaient été promus à l'issue des cours d'été, mais cet effet disparaît l'année suivante (le coefficient est nul pour la lecture alors que le score de mathématiques est très marginalement significativement différent de 0 (à moins de 10 %)). Il est important de noter que les mesures de performances utilisées par Jacob et Lefgren (2004) sont comparables entre niveaux.

Schwerd et West (2012) font un travail similaire sur les tests standardisés en Floride au CE<sub>2</sub> et obtiennent des résultats comparables : les redoublants progressent davantage en lecture et mathématiques que s'ils avaient été promus, mais cet effet diminue très rapidement. Cette étude est particulièrement intéressante car elle présente des estimations à âge égal et à niveau égal. La comparaison à âge égal produit des résultats positifs en mathématiques et en lecture, mais ils tombent à 0 en 3 ans (et sont négatifs ensuite en mathématiques). La comparaison à niveau égal induit un effet positif qui diminue mais reste positif 6 ans après. L'importance et la persistance des effets s'expliquent en partie par l'existence

de différences importantes de moyens alloués aux redoublants. En effet, la Floride impose aux élèves redoublants des remédiations de façon à leur faire acquérir le niveau requis l'année suivante. En particulier, les élèves ont la possibilité de suivre des cours d'été (comme à Chicago), d'être placés dans une classe à effectif réduit sous la responsabilité d'un enseignant *très performant*<sup>20</sup>, et de recevoir des cours intensifs de lecture pendant l'année redoublée. Même s'ils ne disposent pas de beaucoup d'informations sur le suivi de ces mécanismes de remédiations complémentaires, les auteurs montrent que les élèves ayant redoublé leur CE<sub>2</sub> ont en effet été assignés à des classes plus petites comparées à ce que les élèves promus avaient reçu lorsqu'ils y étaient. De plus, la probabilité qu'ils aient un enseignant ayant moins de deux ans d'expérience est diminué de 8 points.

La limite potentielle de cette recherche réside dans le contexte où elle a été menée. En effet, la Floride dispose de plusieurs motifs d'exemption aux scores standardisés qui peuvent permettre à certains élèves d'être classés comme inaptes aux tests (pour des raisons de handicap notamment). Or, [Figlio et Getzler \(2002\)](#) ont montré que la mise en place des tests standardisés en Floride a conduit les écoles à manipuler la base d'élèves passant le test afin d'améliorer les performances moyenne de l'établissement en augmentant le nombre d'élèves exemptés de tests. Les élèves exemptés sont, de plus, majoritairement issus de minorités et de familles à revenus modestes. Le mécanisme d'exemption est explicité en détail dans [Greene et Winters \(2009\)](#), qui montrent par ailleurs que les élèves issus de minorités redoublent davantage et ce même à performances scolaires données.

En somme, les résultats présentés par [Schwerd et West \(2012\)](#) correspondent à l'effet causal du redoublement *pour une fraction d'élèves similaires ayant un niveau à la marge des exigences de performances attendues*. Pour ces élèves, et ces élèves là seulement, le redoublement accompagné des autres mesures de soutien a des effets positifs qui semblent également perdurer. Leur étude ne permet pas d'inférer quoi que ce soit sur les autres redoublants. Les travaux de [Greene et Winters \(2007\)](#) sont assez similaires sur les mêmes données et obtiennent des résultats proches de [Schwerd et West \(2012\)](#).

[Roderick et Nagaoka \(2005\)](#) étudient l'effet du redoublement en CE<sub>2</sub> et en 6<sup>ème</sup> à Chicago, un et deux ans après (à âge égal) avec deux méthodes différentes. La première est un modèle multi-niveau estimant la progression des élèves, la seconde utilise des variables instrumentales pour *vérifier* la robustesse de leurs résultats.

Les modèles multi-niveaux<sup>21</sup> sont de plus en plus populaires en sciences sociales et de nombreux articles de la littérature sur le redoublement s'en servent ([Silbergitt et al., 2006](#); [Goos et al., 2013](#); [Im et al., 2013](#); [Chen et al., 2014](#)). La qualité des résultats auxquels ils aboutissent est très bien décrite par

20. *assignment to a "high performing teacher"*.

21. Les modèles multi-niveaux, initialement développés en sciences sociales et plus particulièrement dans l'éducation ont été présentés pour prendre en compte une dimension contextuelle dans l'analyse individuelle, par exemple l'appartenance à une classe ou une école dans l'analyse des niveaux scolaires. Ils mesurent ainsi des « effets de grappes ». Ces modèles sont adaptés pour prendre en compte, par exemple, une dimension géographique (classe, établissement, académie,...) dans un modèle estimé au niveau individuel, ce qui permet de repérer des effets spécifiques aux différents niveaux.

Bressoux (2007). Ceux utilisés par Roderick et Nagaoka (2005) cherchent à rendre comparable les élèves redoublants et non redoublants à partir de caractéristiques observables et de composantes inobservables communes<sup>22</sup> à chaque niveau (élèves et établissements principalement). Ils montrent ainsi que les élèves qui ont redoublé auraient obtenu des résultats légèrement plus faibles un an plus tard s'ils n'avaient pas redoublé, mais cet effet est complètement dissipé l'année suivante.

Les auteurs présentent également des estimations à variables instrumentales qui ne montrent aucune progression, même à court terme, causée par le redoublement. Une série d'estimation exploite la discontinuité des tests standardisés tout comme Jacob et Lefgren (2004), La seconde utilise les variations existantes entre les différents districts de la région de Chicago dans la mise en place de la réforme instaurant les tests standardisés. Cette dernière technique n'est pas très convaincante. En effet, l'étude de Corman (2003) avait montré que les différences de politiques de redoublement entre les États n'affectent pas significativement la probabilité d'un élève de redoubler, une fois contrôlé pour d'autres facteurs, mettant en doute la validité des estimations de Roderick et Nagaoka (2005) les exploitant.

De plus, les élèves qui redoublent à Chicago ont également suivi six semaines de cours d'été. Ainsi, les résultats estimés précédemment ne sont pas une mesure de l'effet causal du redoublement seul, mais de l'association du redoublement aux *summer schools*.

**Les résultats des études exploitant les discontinuités vont donc généralement dans le même sens : le redoublement engendre des gains positifs ou nuls à court terme sur les performances en mathématiques et en lecture pour les élèves dits marginaux, c'est-à-dire à la limite des exigences de performances attendues. Cependant, le redoublement n'est généralement pas la seule remédiation dont bénéficient ces élèves. Ces éléments indiquent donc que le redoublement au primaire peut avoir des effets positifs à court terme s'il est accompagné d'autres remédiations et notamment des cours d'été. Les effets s'estompent là aussi très rapidement (trois ans après).**

L'article de Chen et al. (2014) apporte également quelques informations sur l'hétérogénéité de l'effet du redoublement par une méthode conceptuellement différente. En effet, les auteurs ne cherchent pas à mesurer l'effet du redoublement, mais plutôt à identifier les trajectoires de progression des élèves redoublant le CP, tout au long de leur scolarité primaire. En se servant d'un modèle *de croissance à mélange de distributions* (ou GMM : Growth Mixture Modeling (Muthen, 2004)), ils cherchent à détecter des profils latents<sup>23</sup> parmi les redoublants, qui auraient alors des trajectoires de progression différentes. Ce type de modélisation est assez sophistiqué et nous ne la présentons pas ici. On retiendra de ces travaux qu'on peut identifier des profils différents parmi les redoublants : un premier type est caractérisé par des résultats aux tests en lecture et mathématiques initialement plus faibles, mais des progrès plus

---

22. Dans un modèle statistique, il y a toujours une part d'inobservé, une partie de la réalité qui n'est pas explicitée ; les modèles multi-niveaux dissocient les caractéristiques à différents niveaux de façon à percevoir plus finement cette hétérogénéité inobservée et réduire la variabilité.

23. C'est-à-dire cachés que les données permettent de révéler

rapide après le redoublement, un second type présente des performances initialement plus élevées au CP mais un développement moins rapide les années suivant le redoublement. Ces résultats montrent que le redoublement n'affecte pas tous les élèves de la même façon et qu'il existe, selon les auteurs, des types d'élèves aux trajectoires de progression différentes. Ceux qui commencent leur second CP avec un niveau plus faible tendent à rattraper plus vite leur retard, ceux un peu plus fort à leur second CP progressent moins vite.

Plus proche de nous, [Goos \(2013\)](#) a récemment consacré une thèse de doctorat à l'étude des effets du redoublement au CP sur les apprentissages, les aspects psychosociaux et la carrière scolaire ultérieure des élèves en Flandre<sup>24</sup>. Elle a publié un article avec d'autres coauteurs ([Goos et al., 2013](#)) dans lequel ils utilisent un modèle multi-niveaux<sup>25</sup> et appariement les redoublants et non-redoublants sur leurs probabilités de redoubler<sup>26</sup> pour corriger le biais de sélection due aux caractéristiques observables. Leur stratégie consiste à estimer des trajectoires de progressions des élèves et de voir comment le redoublement fait dévier les élèves de ces trajectoires. Ils montrent alors que les élèves redoublants réussissent mieux en mathématiques et en lecture l'année de leur redoublement que leurs pairs à risque qui ont été promus. Cependant, cet effet disparaît lors de la seconde année, c'est-à-dire lorsque les redoublants doivent à nouveau apprendre au même rythme que les autres. En outre, L'estimation de la situation *contre-factuelle* des auteurs indique que les redoublants auraient progressé plus rapidement en mathématiques et en lecture s'ils n'avaient pas été retenus. Les indicateurs de développement psychosociaux révèlent des profils *moins avancés* pour les redoublants (ce qui ne surprendra pas le praticien) mais leurs estimations montrent qu'ils se seraient développés plus rapidement s'ils n'avaient pas redoublé leur CP. [Goos \(2013\)](#)<sup>27</sup> met en relation ces résultats avec les études américaines précédemment citées et suppose que "*beaucoup de redoublants recyclent simplement la matière, au lieu de recevoir une instruction à leur rythme, dans leur groupe de niveau*". Au contraire, les écoles américaines mettent en place des moyens de remédiations complémentaires au redoublement (comme montré dans [Schwerd et West \(2012\)](#); [Greene et Winters \(2009\)](#); [Jacob et Lefgren \(2004\)](#); [Roderick et Nagaoka \(2005\)](#) et également décrit dans [Chenu et al. \(2011\)](#)). Enfin, [Goos et al. \(2013\)](#) montrent que les redoublants ont une probabilité plus faible de redoubler à nouveau mais au prix de chances accrues d'être placés dans des cursus spécialisés.

[Im et al. \(2013\)](#) font un travail dont la méthode est presque identique à celle de [Goos et al. \(2013\)](#) dans des écoles élémentaires au Texas. Ils appariement également les élèves redoublants et non redoublants en fonction de leurs caractéristiques observables de sorte à les rendre *virtuellement* identiques au niveau de leurs caractéristiques observables et estiment des trajectoires de progressions à partir d'un modèle multi-niveaux. Ils montrent alors que les élèves redoublants ont une progression similaire à ceux de leurs pairs non redoublants aux caractéristiques observables identiques en mathématiques et en lecture.

---

24. Belgique flamande

25. Voir note en bas de la page 19

26. Matching sur propensity scores

27. cité dans [Désir \(2014\)](#)

Les mesures de comportement des élèves ne varient pas non plus significativement entre redoublants et leurs pairs promus. En somme, il n'y a aucune différence visible dans l'évolution des performances des redoublants et des non-redoublants une fois qu'on a neutralisé l'écart dû à leurs caractéristiques observables spécifiques.

Au Texas également, [Lorence \(2014\)](#) étudie l'effet du redoublement dans le niveau équivalent au CE<sub>2</sub> en 1994 sur les résultats aux tests de lecture standardisés TAAS<sup>28</sup> jusqu'en classe équivalente à la seconde<sup>29</sup>. Il compare les élèves redoublants avec les élèves du même niveau (i.e. dans la même classe) en les appariant sur leurs caractéristiques observables pour limiter le biais de sélection. Il obtient des résultats positifs et significatifs du redoublement, décroissant dans le temps, mais toujours positifs en seconde. Toutefois, sa stratégie consiste à diviser l'échantillon en fonction de la probabilité estimée de redoubler des individus et à utiliser une variable indicatrice pour le rang. Si cette méthode a déjà été employée par [Hong et Raudenbush \(2006\)](#), le modèle n'était pas du tout le même. Cela revient à supposer une homogénéité parfaite entre les élèves de chaque groupe, ce qui est assez discutable. Compte tenu des données, une méthode de double différence avec des groupes appariés sur la base de la probabilité estimée serait plus appropriée.

[Silbergitt et al. \(2006\)](#) utilisent également un modèle multi-niveaux où les élèves sont appariés sur leurs caractéristiques observables et montrent que le redoublement n'a pas apporté de gain dans les trajectoires de progression en lecture entre le CP et le CM<sub>2</sub>. En particulier, les auteurs ne trouvent ni augmentation ni réduction des résultats des redoublants, l'année redoublée. La comparaison avec les élèves appariés ayant été promus ne révèle pas non plus de différence de progression résultant du redoublement.

En France, [Cosnefroy et Rocher \(2005\)](#) ont réalisé plusieurs études sur les effets du redoublement en employant différentes méthodes. Une première étude compare de façon descriptive la progression des élèves les plus faibles en CE<sub>2</sub> ayant ou non redoublé. À partir des données du panel primaire 1997 (DEPP-MENESR), les auteurs trouvent que l'écart de performances entre les élèves n'ayant pas redoublé et les élèves ayant redoublé le CP n'est pas significativement différent de 0 alors qu'il est positif pour les élèves ayant redoublé le CE<sub>1</sub>. Conscient des biais qu'une si simple comparaison de moyenne contient, les auteurs reproduisent cette comparaison avec une procédure de ré-échantillonnage<sup>30</sup> qui reproduit des tirages aléatoires de non-redoublants dans les strates de niveau correspondants. Il ne s'agit pas réellement d'une méthode de comparaison, mais plutôt une façon de donner de la puissance aux tests de différences.

Les résultats de ces comparaisons sont présentés dans le tableau 1.3. Parmi les 1 000 reproductions de tirages aléatoires, la comparaison des redoublants et non-redoublants n'est significative que dans 57.7 % des cas en français, et 26 % des cas en mathématiques. Parmi ces cas, les comparaisons favorables aux

---

28. Texas assessment of academic skills

29. Ils étudient l'effet du redoublement de la 3<sup>rd</sup> grade jusqu'en 10<sup>th</sup> grade).

30. bootstrapping

### Résultats des tests de comparaisons entre redoublants et non-redoublants avec des tirages simulés.

Sens de la différence	% de T significatifs			% de T non significatifs
	à l'avantage des non-redoublants	à l'avantage des redoublants	Total	Total
CE <sub>2</sub> Français	56.9	0.8	57.7	42.3
CE <sub>2</sub> Mathématiques	22.8	3.9	26.7	73.3

Sources : Tableau 6 P. 28 dans Cosnefroy et Rocher (2005)

**Tableau 1.3 – Résultat du tableau 6 dans Cosnefroy et Rocher (2005)**

élèves promus représentent la très grande majorité. Dit autrement, en comparant les élèves redoublants du CP à un groupe d'élèves non redoublants dont les performances sont similaires (i.e. dans le même décile) au début du CP, les chances d'obtenir une différence de performances en CE<sub>2</sub> à l'avantage des élèves redoublants du CP sont quasi-nulles. Bien-sûr, cette méthode ne vise pas à mesurer l'effet causal tel que nous l'avons décrit jusqu'à présent. Elle vise, en revanche, à établir les chances d'obtenir des différences significatives. Les valeurs obtenues sont assez proches des résultats de la méta-analyse de Jimerson (2001) (voir section 1).

Pour mesurer l'effet causal du redoublement, les auteurs utilisent d'une part, une méthode assez proche de l'appariement que nous avons décrit précédemment, d'autre part un modèle à équation structurelle<sup>31</sup> similaire à Grisay (2001). Dans le premier cas, Cosnefroy et Rocher (2005) identifient des caractéristiques cognitives qui prédisent le redoublement et classent ensuite les redoublants dans différents profils. Ils appariant ensuite des non-redoublants ayant les "mêmes" profils pour pouvoir les comparer. Pour les profils "faibles" et les profils "hétérogènes", il n'y a pas de différence de performances entre les élèves ayant redoublé et leurs pairs similaires ayant toujours été promus. Pour les profils plus "forts", les élèves ayant été promus ont de meilleurs résultats que leurs pairs redoublants.

La seconde méthode a pour objectif de réduire le biais dû aux erreurs de mesure affectant l'évaluation au CP et celle au CE<sub>2</sub>. Ainsi, alors que l'utilisation de régressions classiques présente un effet négatif du redoublement sur les résultats en début de CE<sub>2</sub>, l'utilisation du modèle structurel amène des résultats positifs non significativement différents de 0 (donc de très faible ampleur). Ils font, par ailleurs, un autre modèle similaire afin d'expliquer l'impact du redoublement sur des mesures de la motivation. Leurs conclusions sont cette fois sans appel, le redoublement a un impact négatif sur la croyance de l'élève dans son efficacité personnelle et sa motivation.

Plus précisément, les auteurs montrent que les élèves en retard en CM<sub>2</sub> sous-estiment leur niveau de compétence par rapport à leur niveau réel. "A niveau de performance égal en fin de CM<sub>2</sub>, un élève

31. SEM ou *structural equation modeling*

ayant redoublé au CP ou en CE<sub>1</sub> aura tendance à se percevoir scolairement moins performant que son camarade à l'heure" (Cosnefroy et Rocher, 2005, p. 58). De même, les élèves en retard présentent un niveau élevé d'absence de motivation. Ils présentent davantage de comportements subis et non-choisis, comme une adhésion forte à une phrase comme "je n'arrive pas à voir à quoi ça sert de travailler à l'école".

Les conclusions de Cosnefroy et Rocher (2005) pointent l'inefficacité du redoublement en matière de performance scolaire et sa nocivité en terme de motivation et de sentiment de performance. Par ailleurs, les auteurs comparent les notes donnés par les enseignants aux élèves et leurs résultats à des tests standardisés et constatent que la dispersion au sein des classes est plus importante pour les notes données par les enseignants (les classes sont en réalité plus homogènes que ce que la notation des enseignants fait apparaître). Symétriquement, les différences réelles entre classes sont estompées par les notes des enseignants (les notes moyennes des classes données par les enseignants ne reflètent pas bien leur niveau réel). Ainsi, un élève peut être considéré comme faible dans la classe alors qu'au regard des évaluations standardisées, il n'aura pas ce statut. En l'absence de repère extérieur à la classe, la décision du redoublement se fait presque à l'aveugle, ce qui lui confère un caractère tout à fait arbitraire.

Alet (2010) estime l'effet du redoublement dans les deux premières années du primaire sur les performances scolaires en CE<sub>2</sub> et en 6<sup>ème</sup> à partir du panel primaire de la DEPP de 1997. L'intérêt de cette étude est de considérer que non seulement la décision du redoublement concerne un groupe d'enfant aux caractéristiques (visibles ou non) spécifiques<sup>32</sup>, mais que les résultats initiaux (i.e. à l'entrée du CP) qui servent en général à définir le niveau de départ des élèves, ne reflètent pas les caractéristiques initiales des élèves mais sont le résultat d'événements passés (scolarisation en maternelle, aide des parents,...) qui ont une influence sur la probabilité de redoubler et sur les performances futures. Pour identifier le modèle, il faut donc, comme pour Dong (2010), trouver au moins une variable indépendante des caractéristiques de l'élève et de ses performances futures qui explique ses notes à l'entrée du CP, et au moins une variable indépendante qui n'explique que le redoublement. Elle utilise la durée de la maternelle pour la première et le trimestre de naissance pour la seconde. Elle obtient des résultats positifs importants à court terme et négatif à moyen terme. Les tests sont mesurés en comparant les élèves redoublants à leurs pairs promus dans le même niveau (les redoublants sont donc plus âgés). Les résultats d'Alet (2010) indiquent donc qu'un élève qui redouble le CP réussit mieux en CE<sub>2</sub> que s'il n'avait passé qu'un an en CP, mais moins bien en 6<sup>ème</sup>.

Toutefois, la stratégie utilisée est assez discutable même si la démarche est innovante. En effet, pour que sa première étape soit valable, il faudrait que le nombre d'années passé en maternelle n'ait d'effet sur le redoublement et sur les performances futures qu'en déterminant les résultats d'entrée au CP. Or, la durée de la maternelle a bien un impact sur les notes de CP, mais elle impacte aussi directement la probabilité de redoubler. En effet, un élève peut avoir été retenu en maternelle (notamment s'il est né en fin d'année et qu'il a été jugé trop "jeune" pour rentrer au CP l'année de ses 6 ans), ce qui

---

32. Le redoublement est *endogène*.

diminue sa probabilité de redoubler au CP et CE<sub>1</sub>. En outre, la plupart des modèles d'accumulation du capital humain sont multiplicatifs<sup>33</sup>, ce qui implique qu'une entrée tardive à la maternelle peut avoir des effets délétères qui s'accumulent dans le temps. Ainsi, le nombre d'années de maternelle peut affecter les résultats scolaires bien au delà du test de primaire. L'article a été revu et une argumentation un peu plus fournie de la validité de l'instrument est présente dans la version d'ALET et al. (2013). Lorsque les conditions d'application d'une stratégie de *variable instrumentale* ne sont pas valides, Berkowitz et al. (2008) montrent que les résultats sont distordus au point qu'on peut ne pas rejeter, ou au contraire rejeter à tort, et de façon imprévisible, l'hypothèse nulle d'absence d'effet causal. En outre, l'utilisation des trimestres de naissances repose sur l'hypothèse que d'une part, ils n'affectent pas les performances futures autrement que par l'effet sur le redoublement, et que d'autre part, les caractéristiques familiales ne sont pas liées au trimestre de naissance, mais que ce dernier explique des variables de long terme. Cet instrument est très fréquemment utilisé dans la littérature économique. Buckles et Hungerman (2013) recensent plus de 20 études les utilisant, et notamment l'étude très connue de Angrist et Krueger (1990). Or, le caractère aléatoire de cette variable est en réalité discutable. Buckles et Hungerman (2013) ont trouvé plusieurs relations significatives entre le trimestre de naissance des enfants et des variables de contexte familial. Ainsi, aux États-Unis, les femmes qui ont des enfants en hiver sont significativement plus jeunes, moins éduquées et ont une probabilité plus faible d'être mariées. En France, Grenet (2010) obtient lui aussi des différences de contexte familial significatives entre différents mois de naissance. L'importance du biais induit par cet instrument est toutefois difficile à mesurer.

Enfin, L'article de D'Haultfoeuille (2010) développe une nouvelle méthode d'estimation de l'effet causal lorsque la sélection des *traités*<sup>34</sup> est elle-même liée aux valeurs de la variable d'intérêt et l'illustre par une mesure de l'effet à court terme du redoublement sur les notes. La méthode employée est très sophistiquée et nous ne la présentons pas ici. L'intuition est très bien décrite dans Gary-Bobo et Robin (2014), nous invitons le lecteur intéressé à s'y reporter. Ces travaux n'identifient pas une valeur précise mais un intervalle dans lequel se trouve le *vrai effet*. Une première série de résultats impose des restrictions au modèle et identifie un effet du redoublement du CM<sub>2</sub> positif sur les performances en 6<sup>ème</sup>. En relaxant certaines hypothèses, l'auteur obtient des bornes plus larges, majoritairement positives mais n'excluant pas 0. L'incomplétude des données se reflète dans l'imprécision des bornes. Au final, D'Haultfoeuille (2010) conclut que l'effet du redoublement en fin de primaire est plutôt positif à court terme même si des hypothèses moins contraignantes n'excluent pas 0 de l'effet moyen du traitement sur les traités.

**En résumé, les premières recherches sur les effets du redoublement trouvaient des effets systématiquement négatifs à court et à long terme du redoublement mais souffraient de plusieurs problèmes méthodologiques et de données parfois lacunaires ou difficilement comparables. Les**

33. Les connaissances ne s'additionnent pas d'année en année, mais sont le produit des connaissances accumulées dans le passé (voir notamment Cunha et al. (2010)).

34. La littérature de l'inférence causale s'inspire de l'évaluation de l'effet des traitements dans le domaine médicale et s'est donc approprié une partie de son vocabulaire.

évolutions récentes ont permis de corriger ces problèmes et d'affiner sensiblement les conclusions. Dans la majorité des études récentes, le redoublement semble n'avoir aucun effet sur les variables étudiées à moyen terme. Quelques études obtiennent des effets positifs à court terme dans des contextes très particuliers sans pour autant conclure que le redoublement est une remédiation efficace contre les difficultés d'apprentissage.

### 3 Les effets du redoublement dans le secondaire

Le peu de littérature traitant des effets du redoublement dans le secondaire apporte des conclusions différentes. Dans certains cas, celui-ci a des effets négatifs la plupart du temps sauf pour les plus faibles, pour qui le redoublement pourrait être bénéfique. Dans d'autres cas, le redoublement semble permettre aux élèves de faire des progrès au collège mais leur importance varie selon la méthode utilisée. Ce manque de littérature appelle à de nouvelles recherches afin d'obtenir davantage de preuves quant à l'effet du redoublement dans le secondaire.

Alors que la littérature sur les effets du redoublement au primaire abonde, peu d'études quasi expérimentales portent sur le second degré (Crahay, 2004). Mais deux études récentes concernent des données françaises. Dans l'un de ses articles de thèse, Mahjoub (2008) évalue l'effet du redoublement au collège sur les notes de contrôle continu<sup>35</sup> en 3<sup>ème</sup> à partir de deux méthodes quasi expérimentales : une stratégie de variable instrumentale une stratégie d'appariement. Le modèle étudie la différence de notes (standardisés) entre la 6<sup>ème</sup> et la 3<sup>ème</sup> à partir du panel du secondaire de la DEPP (1995).

Comme d'autres auteurs avant lui, Mahjoub (2008) utilise le trimestre de naissance (dont nous avons déjà largement discuté de la validité dans les sections précédentes) comme élément extérieur influençant la probabilité de redoubler sans autre effet sur les performances qu'à travers le redoublement. Ses données lui permettent de faire une hypothèse un peu moins héroïque. Il admet que le trimestre de naissance affecte la maturité et le *capital humain*<sup>36</sup>, mais il considère que son effet est stable entre la 6<sup>ème</sup> et la 3<sup>ème</sup>. Dès lors, si cette hypothèse est valide, cet effet sera éliminé par l'utilisation de la différence de résultats. Les méthodes d'appariement reposent sur l'estimation de la probabilité de redoubler en fonction des variables observables.

Les résultats sont très sensibles à la méthode et à la spécification du modèle. Mais tous indiquent des effets positifs et généralement significatifs. Cependant, on peut émettre quelques réserves sur ces résultats. D'abord, les estimations préliminaires avec des méthodes plus classiques fournissent déjà des résultats positifs, ce qui est très étonnant par rapport à la littérature existante. Ensuite, le manque de stabilité des estimations montre que le modèle manque de robustesse. Gary-Bobo et al. (2013) répliquent

35. La base de donnée n'a pas d'autres mesures de performances. Cependant, les notes de contrôle continu sont fortement liées au contexte de classe de l'élève et mesurent les performances de façon très imparfaite.

36. Terme utilisé par les économistes pour caractériser le niveau de compétence, de performance des individus. De façon intuitive, l'éducation peut être considérée comme une forme d'investissement dans la connaissance, que Gary Becker appelait capital humain.

ce travail dans une section préliminaire à leurs modèles principaux et montrent que cette approche est effectivement peu robuste et très sensible à l'inclusion d'autres variables de contrôle. Au final, toutes les stratégies quasi-expérimentales estimés par Mahjoub (2008) fournissent des résultats positifs, mais la valeur des coefficients varie grandement suivant la méthode employée.

Gary-Bobo et al. (2013) adoptent une stratégie complètement différente des démarches précédentes en construisant un modèle structurel, c'est-à-dire des équations qui décrivent le comportement et le développement des compétences des élèves, les résultats, les décisions de passage,... et emploient les mêmes données que Mahjoub (2008) pour identifier leur modèle et mesurer ainsi l'effet du redoublement. La démarche et les hypothèses du modèle sont décrites de façon très intuitive dans Gary-Bobo et Robin (2014). Nous nous contentons ici de présenter leurs résultats en gardant à l'esprit que ces derniers sont, comme tous les travaux se basant sur un modèle structurel, sensibles aux hypothèses formulées. Gary-Bobo et al. (2013) trouvent ainsi que le redoublement a un effet moyen négatif et significatif sur les résultats du contrôle continu du brevet en mathématiques et français ainsi que sur la probabilité d'atteindre la 3<sup>ème</sup>. Cependant, comme les redoublants et non-redoublants ont des performances initiales différentes, les auteurs indiquent que pour les plus faibles de ces derniers, le redoublement peut avoir un effet positif sur les scores de contrôle continu au brevet. Fruehwirth et al. (2014) avaient, eux-aussi obtenu des résultats similaires mais de sens opposé pour un redoublement en maternelle. Au final, les auteurs concluent que si le redoublement peut avoir des effets positifs sur les résultats des élèves les plus faibles, cet effet est faible en magnitude. Cependant, le redoublement réduit la probabilité qu'un élève atteigne la 3<sup>ème</sup>. Ils ne recommandent donc pas le maintien du redoublement comme politique de remédiation

En Allemagne, Ehmke et al. (2010) étudient l'effet du redoublement en 3<sup>ème</sup> sur les performances en sciences et en mathématiques et sur la représentation que se font les élèves d'eux même en mathématiques à partir de données longitudinales. Ils corrigent le biais de sélection en appariant sur le score de propension les élèves promus et non promus et effectuent une comparaison à âge égal. Les performances des élèves redoublants ne sont pas différentes de celles de leurs pairs aux caractéristiques similaires qui ont été promus. En revanche, le groupe ayant redoublé a une vision plus positive de ses capacités en mathématiques. On pourra reprocher aux auteurs de ne pas avoir employé d'autres techniques d'appariement pour tester la robustesse de leurs résultats.

Enfin, le très récent travail de Hill (2014) adopte une position originale en ne s'intéressant pas à l'effet du redoublement sur les redoublants mais sur les autres élèves de la classe. Ce courant de littérature sur les effets de pairs et les effets de contexte est très dynamique depuis l'article de Manski (1993) offrant un cadre analytique pour mesurer les effets d'interaction. L'auteur questionne à quel point les redoublants produisent ce que les économistes appellent des *externalités négatives* sur leurs pairs, c'est-à-dire des nuisances indirectes sur leurs camarades de classe. Dans un contexte de classes de lycée américains, l'auteur exploite le suivi longitudinal de plusieurs cohortes d'élèves et élimine l'effet

de caractéristiques inobservables de l'école et des élèves via des effets fixes<sup>37</sup>. Il montre ainsi qu'une augmentation de la proportion de redoublant dans un cours donné augmente modérément mais significativement la probabilité d'échouer la matière pour les élèves qui la prennent pour la première fois. Dit autrement, la part d'élèves redoublant une matière semble nuire aux élèves qui ne l'ont pas redoublé. Cet effet n'apparaît qu'à partir d'un seuil compris entre 5 et 10% de redoublants. Les auteurs donnent aussi quelques éléments qui suggèrent que l'effet des redoublants est distinct de celui d'élèves de faible niveau, même si cette hypothèse ne peut être complètement validée. Les auteurs sont cependant dans l'incapacité d'infirmier la possibilité que les enseignants notent les élèves de façon à obtenir une distribution normale et font donc redoubler les élèves en bas de la distribution, généralement les élèves faibles qui prennent le cours pour la première fois.

Au final, cette recension de la littérature scientifique sur l'impact du redoublement - dans le sens *causal* - sur les performances des élèves, des mesures de motivations de confiance en soi,... conduit à nuancer les conclusions passées d'un effet systématiquement délétère du redoublement. Les recherches plus récentes corrigent davantage les problèmes de biais de sélection et obtiennent des résultats sensiblement différents.

Ainsi, le maintien en maternelle peut avoir des effets positifs sur les performances à court terme mais ces effets se dissipent rapidement et sont généralement négatifs à moyen terme. Par ailleurs, le redoublement est un processus dynamique et tend à concerner les élèves les plus en difficultés d'abord. Les effets sont donc différents à chaque niveau car la population concernée n'est pas la même.

La littérature est très abondante sur les effets du redoublement au primaire. De nombreuses études obtiennent des effets positifs à court terme dans des contextes très particuliers sans pour autant conclure que le redoublement est une remédiation efficace contre les difficultés d'apprentissage.

Enfin, les effets du redoublement dans le secondaire restent aujourd'hui très peu étudiés et les rares travaux consacrés ont des conclusions différentes. Ce vide dans la littérature appelle à de nouvelles recherches afin de pouvoir conclure de façon plus certaine sur les effets du redoublement dans le secondaire.

Il est enfin possible que la présence de plusieurs élèves redoublants dans une classe nuise à la réussite des autres élèves, même si cet élément n'a été analysé que dans Hill (2014) dans les lycées américains. Or, les études s'inscrivent dans un système scolaire donné, parfois même dans un contexte spécifique à un établissement, avec des méthodes d'apprentissage, des institutions, des horaires et un fonctionnement qui lui sont propres. Elles ne nous renseignent pas sur l'effet du redoublement sur les performances du système scolaire. La section suivante s'interroge sur le lien entre redoublement et efficacité du système

---

37. rendu possible par la nature longitudinale des données. S'il existe des différences inobservables constantes entre les élèves, ou entre les établissements, l'économètre qui dispose de plusieurs observations pour chaque individu peut éliminer leur impact par la méthode des *effets fixes*. Si l'influence de ces caractéristiques inobservable peut être considérée comme stable dans le temps de façon crédible, cette méthode permet de donner des interprétations causales aux coefficients obtenus.

scolaire.

### III Performances des systèmes scolaires :

#### Le redoublement à l'épreuve des comparaisons internationales

Dans le premier rapport de la série, nous avons comparé les systèmes scolaires de plusieurs pays européens et identifié quelques caractéristiques spécifiques à la France. Dans la section qui suit, nous présentons une comparaison des performances des systèmes scolaires à partir des tests standardisés de l'étude internationale PISA.

##### 1 Les performances des élèves

La comparaison des résultats aux évaluations de PISA permet de montrer que le redoublement a des effets variables sur les performances scolaires des élèves selon les pays. En effet, les pays à promotion automatique ont globalement de meilleurs résultats quelle que soit la matière (mathématiques, français ou sciences). Concernant la France, son niveau moyen est semblable à la moyenne de l'OCDE. De plus, ces analyses révèlent une importante disparité de performances entre les élèves dans les pays pratiquant le redoublement (différence entre élèves à l'heure et élèves en retard). En France, les élèves à l'heure à 15 ans (c'est-à-dire en 2<sup>nde</sup>) ont des résultats parmi les meilleurs de l'OCDE, alors que ceux en retard d'un ou deux ans, ont de très mauvais résultats, bien inférieurs à la moyenne de l'OCDE.

Les vagues successives d'enquêtes PISA - auxquelles la France a participé - permettent de mettre en exergue quelques résultats importants quant au lien entre redoublement et performances globales du système. [Crahay \(2004\)](#) compare ainsi les résultats à l'enquête PISA 2000 en compréhension de l'écrit et montre que les pays à promotion automatique ont globalement de meilleurs résultats : le pourcentage de bons lecteurs est supérieur à la moyenne de l'OCDE et le pourcentage de mauvais lecteurs est inférieur à cette moyenne. ([Crahay, 2004](#), p.13). Pour les pays pratiquant le redoublement, il semble y avoir plus d'hétérogénéité puisque le pourcentage d'élèves mauvais lecteurs est plus élevé que la moyenne en Allemagne, en Belgique, en Italie, en Grèce et au Luxembourg, mais ce n'est pas le cas en France, en Autriche et Espagne.

[Hugonnier et Quéré \(2012\)](#) indiquent par ailleurs que l'étude PISA de 2009 révèle l'équivalent d'une année de différence entre le niveau des élèves français et finlandais. En 2009, 18,7 % des jeunes de 15 ans de l'ensemble des pays participants n'ont pas atteint un niveau efficace de compréhension de l'écrit pour bien s'intégrer dans la société. A titre de comparaison, le taux d'élèves avec un niveau insuffisant de lecture s'élève à 19,8 % en France. De plus, les élèves sont testés sur leurs capacités à comprendre et à utiliser les mathématiques dans la vie courante. Concernant la France, le niveau des élèves avait légèrement baissé entre 2003 et 2006, se retrouvant en-dessous de la moyenne de l'OCDE, puis a augmenté légèrement en 2009 ([Bourny et al., 2010](#)). Les pays pratiquant le redoublement ont

un score moyen médiocre alors que les pays à promotion automatique comme le Japon et la Finlande révèlent un score moyen élevé en mathématiques. En ce qui concerne les sciences, on observe des résultats semblables aux deux autres domaines : les pays à promotion automatique ont en général de meilleurs résultats alors que les pays à redoublement ont des résultats médiocres.

Dans son analyse, Goldstein (2004) compare la France (pays pratiquant le redoublement) à l'Angleterre (pays pratiquant la promotion automatique) afin de voir si la différence de système scolaire a une influence sur les performances des élèves aux résultats de l'enquête PISA 2000. Il est important de souligner que ces deux systèmes scolaires ne diffèrent pas seulement sur le redoublement, mais sur un ensemble d'éléments cruciaux tels que l'âge d'entrée à l'école primaire. En effet, en France, un enfant rentre au CP l'année de ses 6 ans (peu importe son mois de naissance), alors qu'en Angleterre, les enfants nés entre janvier et août entreront à l'école en septembre l'année de leurs 5 ans, et ceux nés entre septembre et décembre iront à l'école élémentaire l'année suivante. Ainsi, des enfants nés la même année ne seront pas forcément au même niveau d'études. Pourtant, on constate un écart important de performances entre la 3<sup>ème</sup> et la 2<sup>nde</sup> en France, alors que ce n'est pas le cas en Angleterre (grade 9 et grade 10). Par contre, au sein des classes en Angleterre il y a une importante disparité entre les élèves (élèves forts et faibles ensemble), alors qu'en France les classes semblent plus homogènes (élèves faibles en 3<sup>ème</sup> et élèves forts en 2<sup>nde</sup>).

### L'enquête PISA

L'enquête PISA (Program for International Student Assessment) a été mise en place par l'OCDE en 2000 afin de mesurer la performance des systèmes éducatifs de pays participants (Environ une soixantaine de pays pour l'édition 2013). Elle permet de comparer les compétences des élèves à 15 ans dans trois domaines : compréhension de l'écrit, culture mathématiques et culture scientifique. Chaque année, l'évaluation se concentre sur une compétence particulière (2000 : lecture - 2003 : maths - 2006 : sciences et ainsi de suite). On peut ainsi évaluer les systèmes scolaires de ces pays et suivre leur évolution. Chaque pays sélectionne un échantillon d'élèves de 15 ans choisi de manière aléatoire et leur fait passer des tests écrits (questions à choix multiples et questions ouvertes).

Comme l'indiquent Cosnefroy et Rocher (2005), les enquêtes PISA sont parfois les sujets de critiques - comme toutes les évaluations internationales - car elles soulèvent quelques problèmes méthodologiques liés à la fois au contenu de l'enquête, à la traduction des questionnaires, à la correction et à l'interprétation des résultats. Il faut en particulier faire attention aux classements et aux "*palmarès internationaux [qui] donnent une vision réductrice voire biaisée de la réalité*".

L'étude des compétences des élèves met aussi en évidence une disparité des performances des élèves dans les pays pratiquant le redoublement, c'est-à-dire des différences entre les élèves à l'heure et les élèves en retard. En effet, "*quel que soit le moment du cursus scolaire, les élèves en retard ont, en moyenne, des résultats nettement inférieurs aux élèves à l'heure*" (Cosnefroy et Rocher, 2005). En prenant le cas de la France (enquête PISA 2000), les élèves à l'heure à 15 ans (2<sup>nde</sup> générale ou technologique : 48,2 %) ont les meilleurs résultats en compréhension de l'écrit devant la Finlande alors que ceux en retard d'un ou deux ans (3<sup>ème</sup> : 36,5 % ou 4<sup>ème</sup> : 7,11 %) ont de très mauvais résultats, bien inférieurs à la moyenne de

l'OCDE. Des études menées en France, en Espagne et en Belgique (Communauté Française) montrent à chaque fois les mêmes résultats : ces pays qui pratiquent le redoublement révèlent que les élèves en retard ont de moins bons résultats que les élèves à l'heure.

À partir des données de PISA 2012, nous avons reproduit le travail de [Cosnefroy et Rocher \(2005\)](#) pour situer les élèves en retard à 15 ans en France par rapport aux performances moyennes des autres pays, nous le présentons dans la figure 1.1(a).

Si le classement des pays a sensiblement changé par rapport aux données de l'enquête 2000, on peut faire le même constat que [Cosnefroy et Rocher \(2005\)](#). La moyenne française se situe légèrement au dessus de la moyenne des pays de l'OCDE<sup>38</sup> (495 par rapport à 494 points), mais une fois éclatée, on constate que les élèves à l'heure ont un score très élevé (un peu plus bas que le Japon qui pratique la promotion automatique). En revanche, les élèves qui, à 15 ans, ne sont qu'en 3<sup>ème</sup> (et ont donc 1 an de retard) sont au plus bas du classement. La représentativité du score moyen est donc à prendre avec mesure, en tout cas pour la France.

Le lecteur attentif pourrait penser que le résultat serait identique si le retard scolaire n'avait aucun effet mais que la variabilité entre les élèves français était simplement plus grande. En estimant les densités (voir figure 1.2) des notes des élèves français par rapport à la moyenne de l'OCDE, On ne constate pas d'aplatissement plus important de la densité pour la France. Les résultats des élèves français ne sont pas plus dispersés que ceux de tous les pays de l'OCDE ; la variabilité est tout à fait semblable. En revanche, on constate que les distributions des notes des élèves à l'heure et en retard sont particulièrement distinctes. La grande majorité des élèves en retard n'atteint pas la moyenne de l'OCDE (représenté par la ligne verticale) alors que les élèves à l'heure sont globalement très performants.

## 2 Le taux de retard

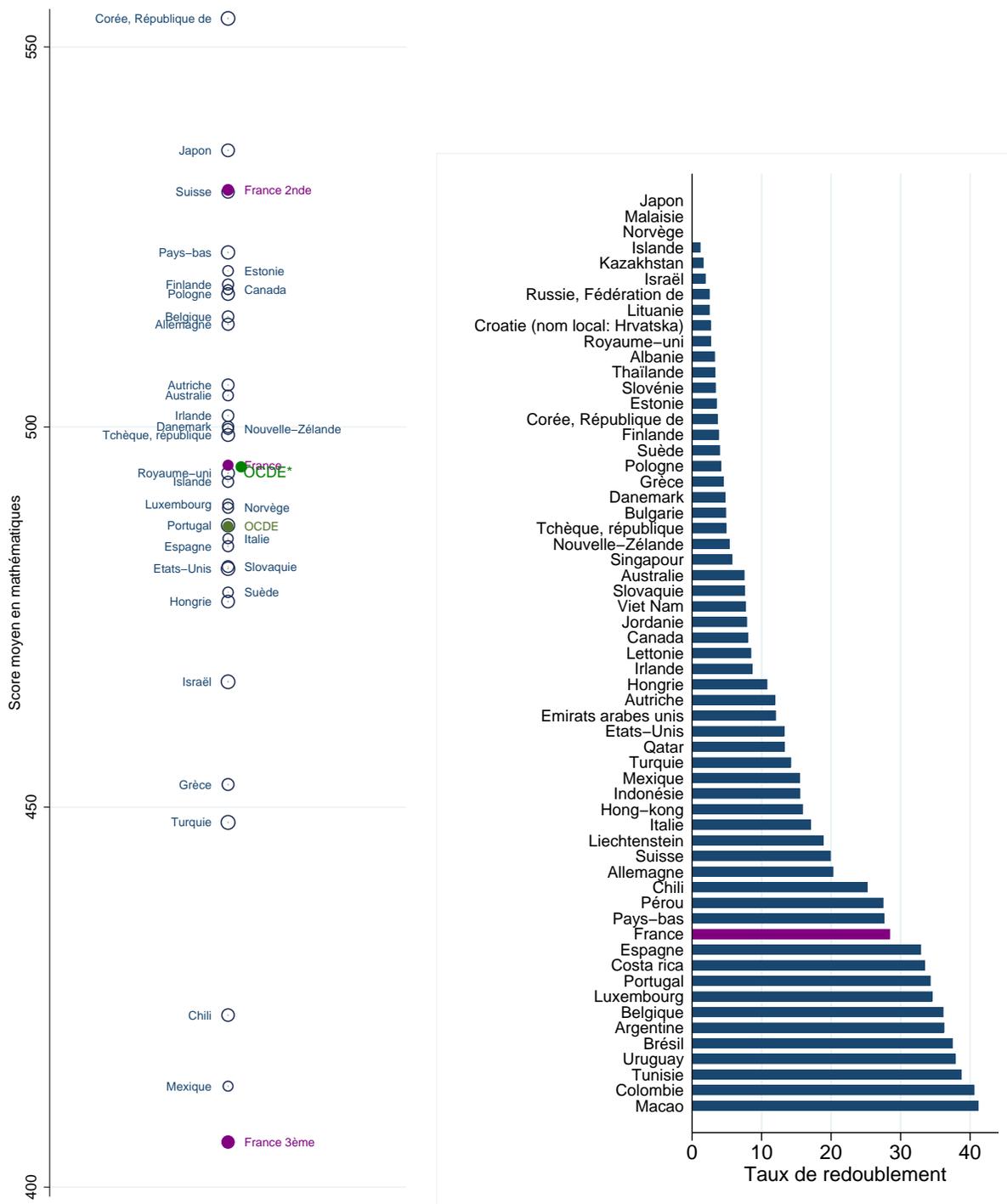
L'enquête PISA 2012 révèle une importante disparité des taux de redoublement au sein de l'OCDE. En effet, 55 % des pays ont un taux de redoublement inférieur à 10 % (Norvège, Finlande, Taïwan, Shanghai, Angleterre, Canada,...). La particularité de ces pays est qu'ils ont pour la plupart limité voire interdit le redoublement. Parmi les pays ayant le plus fort taux de redoublement, la France se trouve au même niveau que les Pays-Bas avec 28 %. De plus, ces analyses mettent en avant une corrélation faible mais négative entre taux de redoublement et performance scolaire.

La dernière enquête PISA révèle l'importante hétérogénéité du taux de redoublement<sup>39</sup> suivant les pays. On illustre ce résultat dans la figure 1.1(b). Parmi les 65 pays participants à l'enquête, 55 %

---

38. L'OCDE, dans ses notes pays, calcule la moyenne de ses pays membres comme une moyenne arithmétique des statistiques par pays. Dans cette étude, ce que nous présentons comme *score moyen des élèves de l'OCDE* est la moyenne pondérée des cinq valeurs plausibles en mathématiques, calculée sur l'ensemble des élèves de l'OCDE.

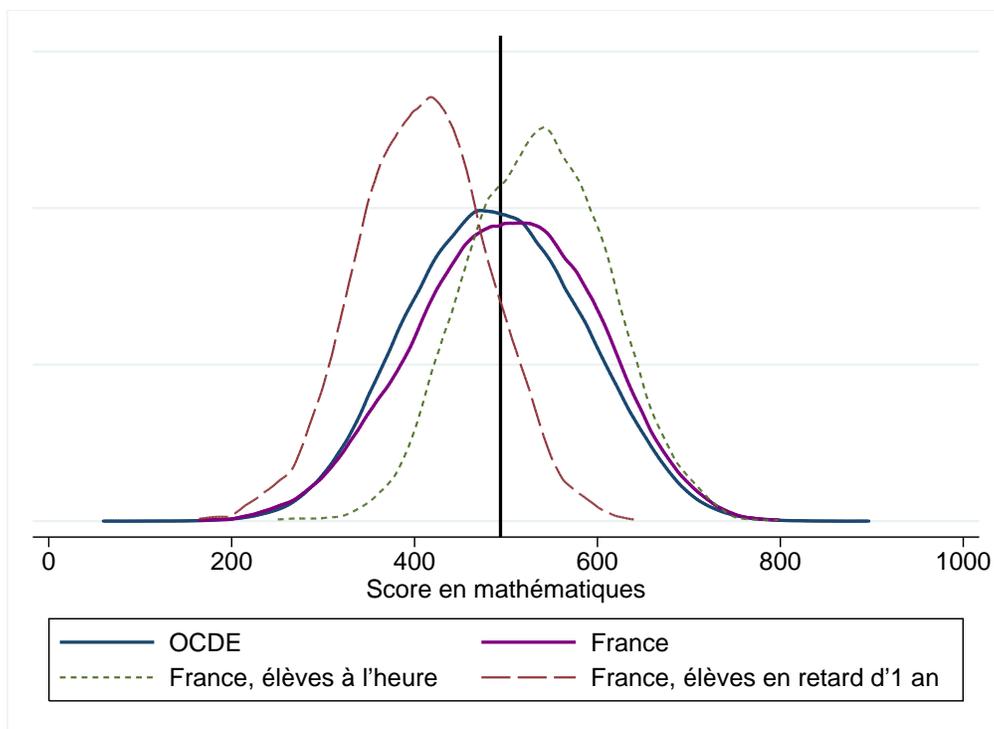
39. Mesuré ici par le taux d'élèves déclarant avoir redoublé au moins une fois pendant leur parcours



(a) Situation des élèves français suivant leur retard scolaire par rapport aux autres pays de l'OCDE 2012 (b) Taux de redoublement dans tous les pays participants à PISA 2012

**Figure 1.1 – Comparaison des taux de redoublement et des performances des élèves Français avec les autres participants à PISA 2012**

Sources : PISA 2012 : Calculs de l'auteur à partir des 5 valeurs plausibles en mathématiques estimées 81 fois comme indiqué dans OECD (2009), chapitre 8. La taille des bulles représente un intervalle de confiance de 95 %, les erreurs standards ont été obtenues après 81 réplifications (Balanced repeated replication). Le graphique contient deux moyennes de l'OCDE, celle étoilée est la moyenne des scores des pays de l'OCDE. Elle est de 494, soit un point de moins que la France; l'autre est calculée sur tous les élèves de l'OCDE avec la même méthode que les moyennes par pays. Elle pondère en quelque sorte la moyenne par le poids relatif de chaque pays de l'OCDE.



**Figure 1.2 – Densité des scores en mathématiques dans PISA 2012**

Densité estimée par kernel de la première valeur plausible en mathématiques (PV1MATH).

ont un taux de redoublants inférieur à 10 %. On trouve notamment dans cette liste les pays nordiques (Norvège, Suède, Finlande, Danemark), les pays baltes et slaves, le Japon et les dragons asiatiques (Taïwan, Corée du Nord, Malaisie, Thaïlande, Vietnam, Shanghai) et les pays du *Common Wealth* Britannique (Angleterre, Irlande, Australie, Nouvelle-Zélande, Canada). La plupart de ces pays pratiquent la promotion automatique mais autorisent le redoublement dans certains cas exceptionnels (maladie, fort absentéisme). En règle générale, ces pays favorisent le suivi et l'aide des élèves en difficulté plutôt que de les faire redoubler.

Les pays d'Europe occidentale font partie des pays ayant le plus fort taux de redoublement. Ainsi, les taux de redoublement en France et aux Pays-Bas est de 28 %, le même niveau que le Pérou. Le Luxembourg et la Belgique ont le même taux de redoublement que le Brésil et l'Argentine, soit 35 % environ.

Si on rapporte ces taux de redoublement aux performances moyennes des élèves à 15 ans dans ces pays, on observe une corrélation faible mais négative (et significative) qui indique que les pays à fort taux de redoublement ont tendance à être moins performants que les pays ayant un taux de redoublement plus faible. Ce résultat est illustré dans la figure 1.4. On retrouve ici les résultats des travaux de Nathalie Mons et ses co-auteurs (Duru-Bellat et al., 2004) même si nous ne contrôlons pas par le niveau de richesse. Les auteurs concluaient alors que *"la pratique du redoublement, non seulement n'améliore pas*

les résultats des élèves, mais tend à les amoindrir, quand on les mesure chez les élèves d'un âge donné" (Duru-Bellat et al. (2004) p. 65 cité dans Paul et Troncin (2004) p28).

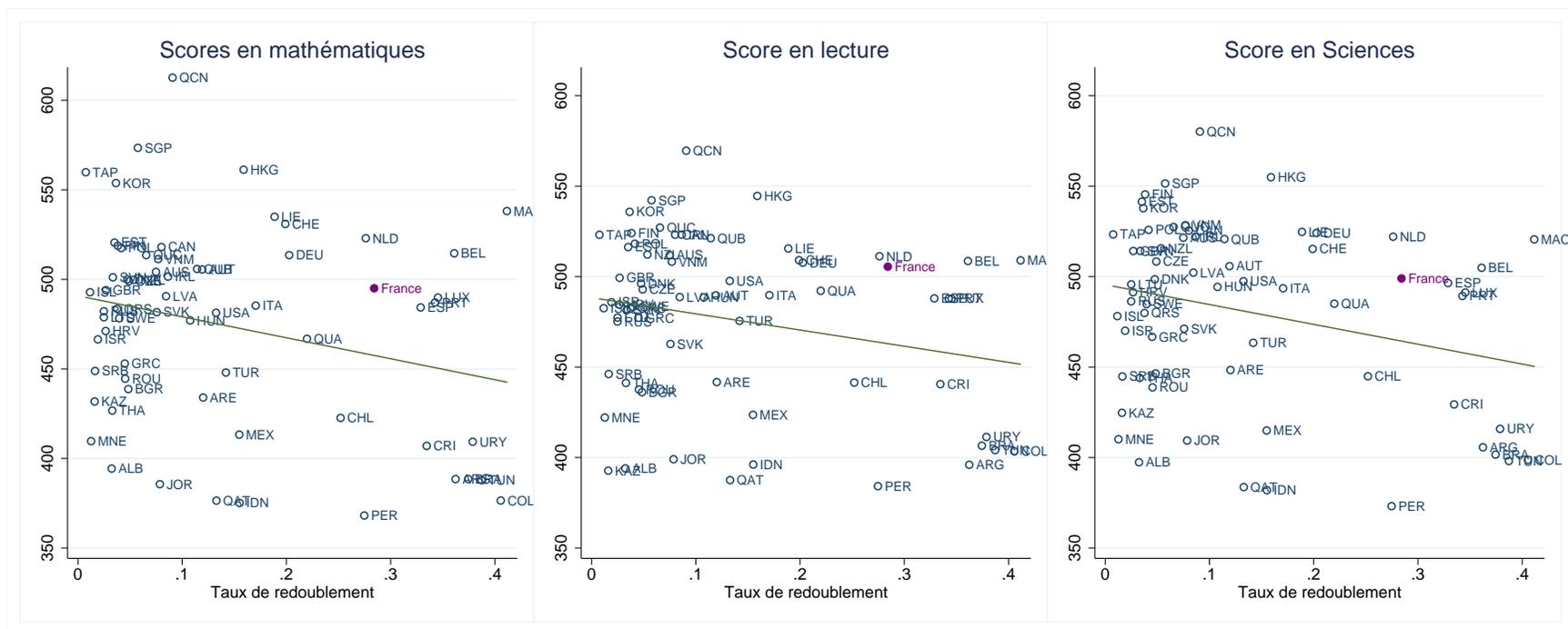
Pour s'assurer de cette corrélation, nous avons régressé les résultats moyens des pays dans les trois disciplines testées dans PISA 2012 sur les taux de redoublement dans un modèle à équations simultanées (Zellner, 1962). Ces modèles autorisent une corrélation entre les erreurs des trois équations, c'est-à-dire, un lien entre *ce qui n'est pas expliqué* par le taux de redoublement dans les scores en mathématiques, en sciences et en lecture.

Estimation simultanée du lien entre performances moyenne et taux de redoublement.			
	Maths	Sciences	Lecture
	$\beta/(SE)$	$\beta/(SE)$	$\beta/(SE)$
Taux de redoublement	-117.018** (52.395)	-89.380** (44.433)	-109.590** (47.634)
Constante	490.793*** (9.744)	488.365*** (8.263)	495.523*** (8.858)
$R^2$	.071	.059	.075
Chi <sup>2</sup>	4.99	4.05	5.293
P-value	.026	.044	.021
Observations	65	65	65

\* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Erreurs standards entre parenthèses  
Régressions simultanées des scores moyens aux tests PISA par pays sur la moyenne de *repeat* par les méthode des *seemingly unrelated regressions*. On autorise ainsi une corrélation entre les résidus des trois équations et donc un lien possible entre ce qui n'est pas expliqué par le redoublement dans les performances moyenne des pays en mathématiques, en sciences et en lecture. Les données ont été calculées avec les pondérations comme recommandé par OECD (2009).

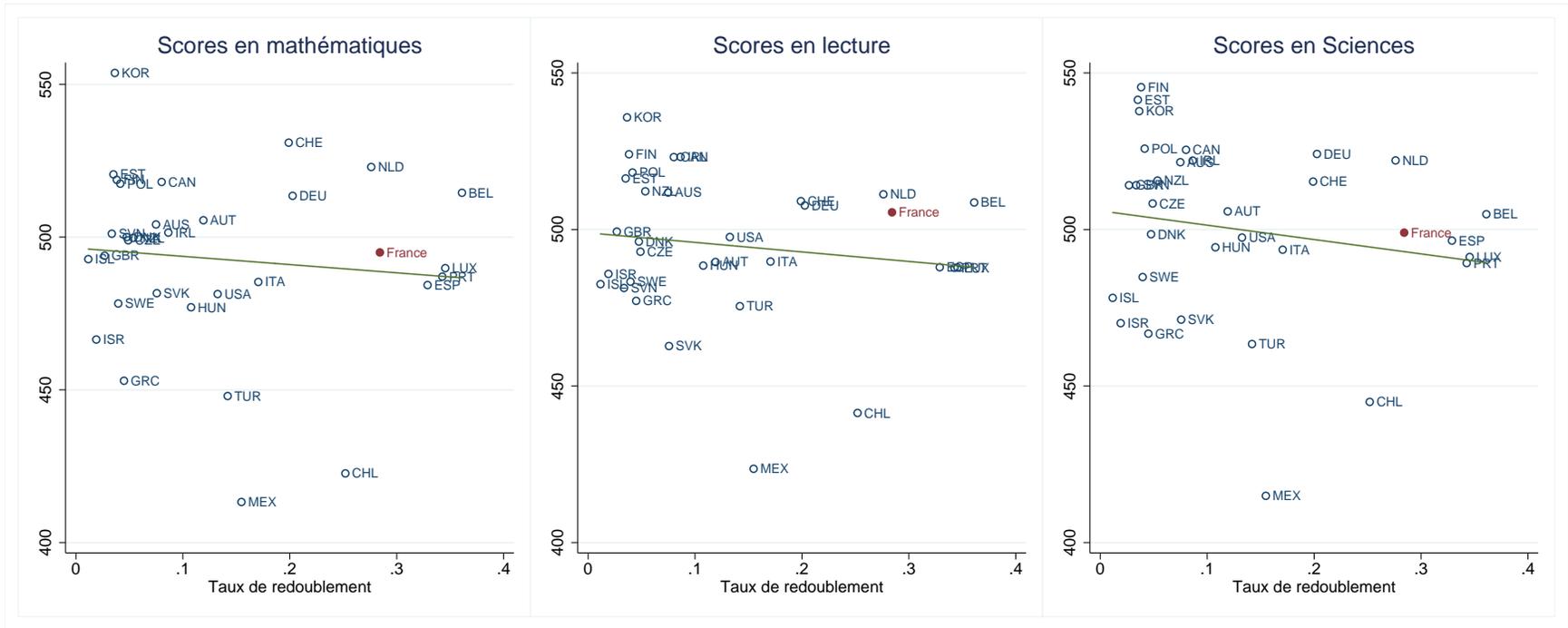
**Tableau 1.4 – Ajustement linéaire des performances moyennes des pays à PISA en fonction du taux de redoublement**

Les résultats sont également négatifs et significatifs, signalant une corrélation négative entre taux de redoublement et performances scolaires. Les coefficients, bien que peu précis en raison du faible nombre d'observations, sont proches d'un écart-type pour chaque variable. Une hausse de 10 point du taux de redoublement est associé à une diminution de 0.1 écart-type de performance environ. Cette corrélation n'est plus significative lorsqu'on restreint l'échantillon aux seuls pays membres de l'OCDE, comme l'illustre la figure 1.4 Les droites de régression sont nettement moins pentues mais toujours décroissantes.



**Figure 1.3 – Corrélation entre taux de redoublement et performance moyenne dans l'enquête PISA 2012**

sources : PISA 2012, calculs de l'auteur à partir de la variable REPEAT dont la fréquence a été calculée par pays et pondérée par la variable W\_FSTUWT. Les scores moyens par pays ont été calculés à partir des valeurs plausibles dans chaque domaine. Les données cognitives de PISA sont en effet basées sur un modèle RASCH et les performances des élèves sont des valeurs plausibles aléatoires. Pour obtenir la moyenne par pays, nous avons calculé la moyenne des 5 valeurs plausibles par discipline, pondérées par la variable W\_FSTUWT et obtenu la moyenne finale par une moyenne arithmétique de ces 5 valeurs. Les erreurs types (non présentées ici) ont été obtenues en répliquant 80 fois l'estimation de chaque moyenne pondérée par les poids W\_FSTR1 à W\_FSTR80.



### 3 Origine sociale

Les enquêtes PISA soulèvent aussi un lien entre performances des élèves et origine sociale qui implique que les élèves de milieu défavorisé ont de moins bons résultats dans l'ensemble et ont plus de chance de redoubler, comme c'est le cas en France. En outre, plus les pays présentent des écarts de performances importants, plus les inégalités sociales sont fortes.

Les enquêtes révèlent aussi la plupart du temps un lien entre performances des élèves et origine sociale. L'indice de statut économique, social et culturel (SESC) permet d'évaluer ce lien. En 2009, il indique qu'en France la performance des élèves est plus corrélée avec leurs origines socio-économiques que dans les autres pays (Fumel et al., 2010). En effet, les élèves provenant de milieu social défavorisé ont tendance à avoir des résultats plus faibles et à redoubler. Ainsi, les pays montrant de faibles résultats aux tests sont souvent ceux avec des inégalités sociales importantes. Les données de PISA permettent de déterminer l'impact de l'origine sociale sur le risque de redoublement en comparant les élèves à niveau égal. De plus, les élèves provenant de milieux défavorisés ont un taux de redoublement plus élevé que les autres, ce qui s'explique par le fait que ces élèves ont un niveau plus faible. De même, dans la majorité des cas, les élèves ayant des parents d'origine étrangère auront plus de chance d'être en retard comparés aux autres enfants (Désir, 2014).

L'analyse de l'OCDE (2013) montre ainsi que 26 % de la variation dans l'effet du statut socio-économique de l'élève sur ses performances en mathématiques peut être expliqué par des différences dans les taux de redoublement entre les pays. Cette corrélation est illustrée dans la figure 1.5 issue de (OCDE, 2013, p.35). L'axe des ordonnées est la part de variance dans les scores en mathématiques des élèves expliquée par l'indice de statut socio-économique<sup>40</sup>, l'axe des abscisse est le taux de redoublement. Les points en gris indiquent la part des scores de mathématiques expliqués par l'indice de statut économique seul, les points en bleu tiennent sous contrôle le niveau de revenu par tête dans le pays.

Dans la note *PISA in focus* du mois de septembre 2014 (OCDE, 2014), les auteurs estiment l'effet du statut social sur la probabilité d'avoir redoublé en contrôlant les performances dans les trois disciplines testées dans PISA. Le graphique 1.6 présente les résultats de ces estimations. Ainsi, à niveaux en mathématiques, sciences et lecture donnés, la probabilité d'avoir redoublé est significativement plus élevée pour les élèves défavorisés par rapport aux élèves favorisés dans 33 pays sur 61. En moyenne, les élèves défavorisés ont 1,5 fois plus de chance que les élèves favorisés d'avoir redoublé, toutes choses égales par ailleurs. En France, le ratio des probabilités vaut 2. On note toutefois que les performances en mathématiques, sciences et lecture sont mesurées après le redoublement, ce qui peut modifier fortement l'intensité et même le sens des corrélations observées (voir annexe II). Il reste que ces résultats apparaissent troublants et indiquent un lien *non-causal* entre redoublement et statut social significatif.

---

40. Il s'agit du  $R^2$  des régressions par pays pour les points en gris et du carré des coefficients de corrélation partiel pour

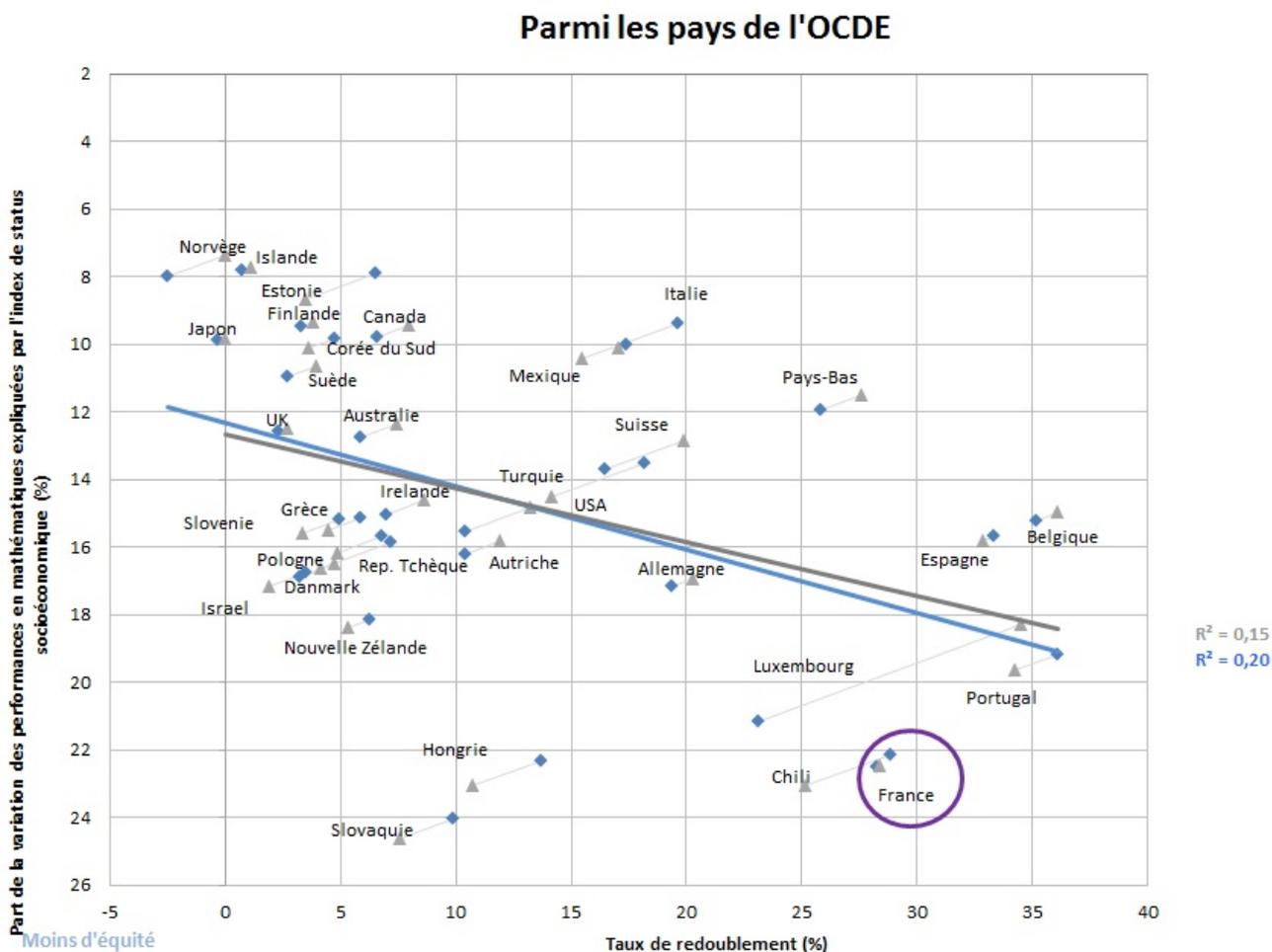
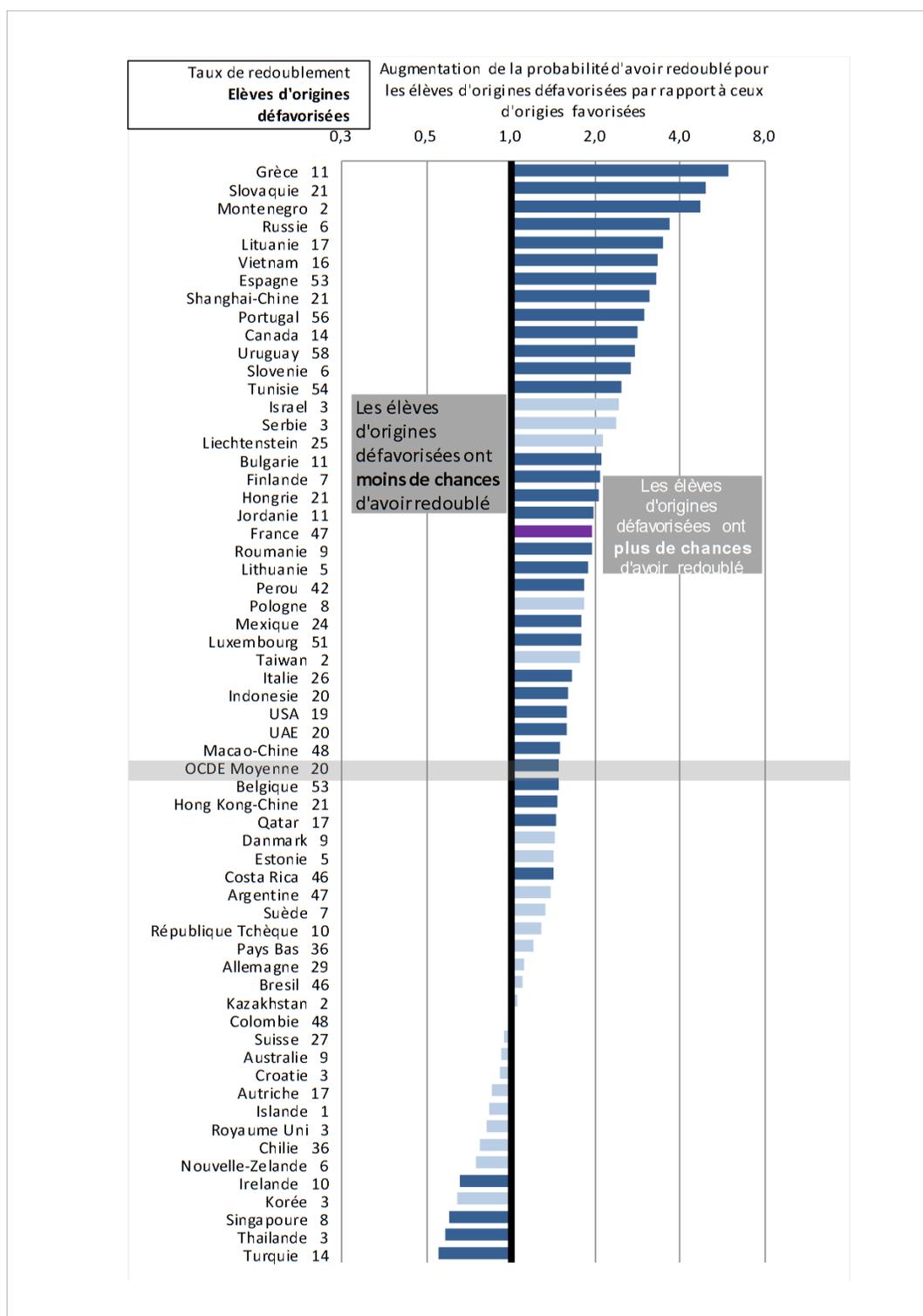


Figure 1.5 – Part de la variance des performances en mathématiques expliquée par l'indice de statut social et le taux de redoublement

Sources : (OCDE, 2013, p;35), figure IV.1.V, traduction du Cnesco



**Figure 1.6 – Impact du statut socio-économique sur la probabilité d'avoir redoublé, toutes choses égales par ailleurs**

Sources : OCDE (2014). Le graphique présente les Odd ratios (voir annexe B) de régressions logistiques estimant la probabilité d'avoir redoublé sur les quartiles de l'indice de composition sociale de la famille de l'élève et des valeurs plausibles en maths, sciences et lecture par pays. Les barres claires ne sont pas significativement différentes de 0. Ainsi, à performance comparable les élèves parmi les 25 % les plus défavorisés ont presque deux fois plus de chance d'avoir déjà redoublé qu'un élève parmi les 25 % les plus favorisés.

Aussi, il existe une relation entre hétérogénéité des performances et inégalités sociales : plus les pays ont des écarts de performances importants, plus les inégalités sociales sont fortes. A l'inverse, Nathalie Mons ([Duru-Bellat et al., 2004](#)) et ses coauteurs montrent que les pays avec des élèves aux performances homogènes révèlent moins d'inégalités sociales.

## IV Les effets du redoublement sur les parcours scolaires et l'insertion dans la vie active

La carrière d'un élève ne se résume pas aux notes qu'il reçoit à la fin de sa scolarité obligatoire. La suite de son parcours est marquée par des choix d'orientation, des décisions de poursuite ou d'abandon des études, et son insertion professionnelle. Alors que l'élève garde un certain contrôle sur les deux premiers, les conditions d'insertion sur le marché du travail dépendent de nombreux facteurs qui ne dépendent ni de ses caractéristiques ni de ses décisions<sup>41</sup>. Cependant, il est légitime de s'interroger sur l'effet d'avoir redoublé sur le niveau final d'éducation de l'élève et ses conditions d'entrée dans le marché du travail.

### 1 L'effet du redoublement sur le décrochage et l'éducation supérieure

Le redoublement a un effet négatif sur les trajectoires scolaires dans le sens où les élèves qui redoublent ont tendance à arrêter l'école plus tôt, sont moins nombreux à obtenir un diplôme du secondaire et ont des ambitions moins fortes.

Le décrochage est entendu ici comme l'action de quitter le système éducatif sans avoir obtenu de diplôme ou d'avoir obtenu le diplôme équivalent à la fin de l'école obligatoire (le Diplôme national du brevet et France par exemple). Dans la plupart des pays pratiquant le redoublement, avoir été retenu une année est associé à un risque de décrochage plus élevé. Cette littérature est recensé dans [Jimerson \(1999\)](#) qui explique également les limites méthodologiques de ces études. Il est en effet très différent d'identifier des décrocheurs et de voir combien parmi eux ont redoublé que de comparer des élèves ayant redoublé avec des élèves faibles ayant été promus. La majorité des études emploient pourtant la première méthode. Par ailleurs, ces comparaisons ne mesurent pas non plus l'effet causal du redoublement, qui reste très largement déterminé par des facteurs inobservables, liés à la probabilité de quitter l'école sans diplôme. Néanmoins, ces recherches indiquent que le redoublement produit peu d'effets positifs et peuvent augmenter le risque de décrocher ([Roderick, 1995](#)).

[Jimerson \(1999\)](#) utilise un panel d'élèves en difficultés suivis pendant 21 ans dans le Minnesota et compare des élèves ayant redoublé une fois entre la dernière année de maternelle (*kindergarten*) et le CE<sub>2</sub> (3<sup>rd</sup> grade.) avec leurs pairs semblables promus aux mêmes années. Les comparaisons sont réalisées à partir de comparaisons statistiques simples<sup>42</sup>. Il montre alors que les élèves ayant redoublé sont significativement plus nombreux (en pourcentage) à abandonner l'école avant 19 ans, sont significativement moins nombreux à obtenir un diplôme au lycée et à être inscrit à 20 ans dans un établissement relevant

---

41. Lire notamment l'article de [Heim \(2014\)](#) qui décrit de façon simple la théorie et la littérature empirique étudiant l'effet d'entrer sur le marché du travail pendant une récession. Parmi les travaux scientifiques recensés, ceux d'[Oreopoulos et al. \(2012\)](#) sont remarquables sur le cas canadien.

42. Analyse de variance et des covariances (*ANOVA et ANCOVA*).

de l'enseignement supérieur.

Plus récemment, [Stearns et al. \(2007\)](#) étudient l'effet du redoublement sur la probabilité de quitter l'école sans diplôme à partir de données de panels américains composés d'élèves entrés en 5<sup>ème</sup> en 1988 qui ont été réinterrogés en 1990, 1992, 1994 et 2000. Ils distinguent les élèves qui décrochent tôt (c'est-à-dire avant la fin de la 3<sup>ème</sup>) de ceux qui décrochent plus tard (jusqu'en 1<sup>ère</sup>). À partir de régressions logistiques, ils montrent que le redoublement prédit significativement le décrochage, pour les "*blancs, noirs et hispaniques*"<sup>43</sup> par rapport aux blancs promus en continu, qu'il ait lieu tôt ou tard, à ressources, caractéristiques socio-démographiques et parcours scolaires donnés. Cependant, l'effet est variable pour les différents groupes ethniques considérés. Les ressources disponibles et les conditions dans lesquels ils évoluent expliquent en partie pourquoi les élèves en retard tendent à abandonner plus souvent l'école que leurs pairs blancs promus. C'est pourquoi les auteurs concluent que les établissements qui souhaitent minimiser le décrochage doivent allouer davantage des ressources identifiées dans ces recherches aux élèves en retard. L'étude met également en lumière le rôle des relations sociales dans l'échec scolaire et montrent ainsi que les élèves ayant redoublés ont moins d'interactions avec leurs enseignants, sont moins engagés dans l'école, ce qui augmente leurs chances de décrocher.

Il est important de noter que ces travaux n'identifient pas l'effet causal du redoublement au sens où nous l'avons détaillé dans la section 1, même si ils tentent de réduire les différences entre élèves promus et retenus par des variables de contrôle. À notre connaissance, seule l'étude de [Eide et Schowalter \(2001\)](#) tente de corriger le problème d'endogénéité du redoublement à l'aide de stratégies quasi-expérimentales. Les auteurs emploient une stratégie de *variable instrumentale* pour expliquer le redoublement et ensuite mesurer son impact sur la probabilité de redoubler et sur le revenu. Les auteurs exploitent eux-aussi les dates de naissances des élèves couplées aux différences de législation entre les États américains en matière d'âge d'entrée à l'école maternelle. En comparant des enfants de même âge entre différents États, on compare des enfants qui sont rentrés plus ou moins tôt à l'école, mais qui n'ont pas de raison d'être très différents par ailleurs. Comme souvent dans la littérature, plus un élève est jeune par rapport à ses pairs, plus il a de chance d'être retenu. C'est sur ces bases que les auteurs fondent leur stratégie empirique. Celle-ci n'est valide que si la date de naissance n'affecte pas les revenus futurs ou la probabilité d'abandonner l'école autrement que par son effet sur les chances de redoubler. Nos connaissances des travaux d'[Angrist et Krueger \(1990\)](#) ou encore [Grenet \(2010\)](#) nous incitent à la prudence quant à la validité de cette hypothèse. Il est probable que la date de naissance ait un impact direct sur les revenus, qui ne passe pas par le redoublement.

[Eide et Schowalter \(2001\)](#) montrent alors que les estimations *naïves* de l'effet du redoublement sur le décrochage sont positives, mais que les estimations par variables instrumentales (i.e. censées mesurer l'impact causal du redoublement sur la variable d'intérêt) changent le signe des estimations sans être significativement différent de 0. Les résultats pour les revenus sont semblables : négatifs d'abord et

43. Les données américaines permettent de caractériser l'origine ethnique des individus selon une norme définie par le *federal Office of Management and Budget (OMB)* et le *United States Census Bureau*

positifs (non significatifs) avec les variables instrumentales. Cependant, leurs estimations sont vraiment peu précises ce qui ne permet pas de conclure sur le sens de l'effet causal.

Jacob et Lefgren (2007) utilisent la même méthode que dans leur étude de 2004 (Jacob et Lefgren, 2004) pour mesurer l'effet causal *local* du redoublement sur la probabilité de terminer le lycée. Ils se servent de la discontinuité induite par les notes seuils aux tests standardisés qui conduisent les étudiants juste en dessous de ce seuil à redoubler, les autres à être promus. Les estimations par les méthodes classiques montrent que le redoublement en 6<sup>ème</sup> et en 4<sup>ème</sup> augmente la probabilité de décrocher respectivement de 6 et 14 %. Cependant, l'ajout de nouvelles variables de contrôle diminue l'effet, indiquant ainsi l'existence de caractéristiques associées à la fois au redoublement et au décrochage. Les discontinuités de régression révèlent alors que le redoublement en 6<sup>ème</sup> n'a d'impact significatif ni sur la probabilité d'abandon, ni sur l'âge auquel l'élève décroche. Au contraire, le redoublement en 4<sup>ème</sup> augmente significativement la probabilité de décrochage (de 8 points de pourcentage). Ces résultats sont locaux, c'est-à-dire, ils correspondent à l'effet du redoublement pour les élèves à la marge de la discontinuité. **Pour ces élèves là, il semble que le redoublement tardif ait un impact négatif durable sur leurs parcours scolaire.**

En France, Afsa (2011) fait plusieurs estimations de l'effet d'avoir redoublé la 6<sup>ème</sup> sur la probabilité d'obtenir un diplôme du secondaire à partir des données du Panel 1995 (DEPP-MENESR). Une première stratégie consiste à estimer l'effet moyen du redoublement à partir d'un modèle de probabilité avec un ensemble de variables de contrôle<sup>44</sup>. Sans surprise, l'effet du redoublement est négatif et significatif, mais bien moindre que sans variables de contrôle. La seconde stratégie employée est proche d'une stratégie de variable instrumentale même si la nature binaire des variables à expliquer (obtention d'un diplôme et redoublement) contraint à l'utilisation de modèle probabiliste à équation simultanée<sup>45</sup> dont l'interprétation n'est pas aussi naturelle. L'identification du modèle tient à l'écart à la moyenne de la classe aux tests de mathématiques et de français à l'entrée en 6<sup>ème</sup>. L'intuition derrière ce choix est que la probabilité de redoubler dépend de la place relative de l'élève par rapport aux autres, indépendamment du niveau de performance absolu. La validité de la méthode est conditionnelle à l'hypothèse que l'écart à la moyenne aux tests de 6<sup>ème</sup> n'a d'impact sur l'obtention d'un diplôme du secondaire qu'à travers la probabilité de redoubler en 6<sup>ème</sup>, ce qui est assez crédible de notre point de vue. L'auteur trouve alors que le redoublement a globalement un effet négatif et significatif avec cette méthode, et donc que le redoublement impacte négativement la probabilité d'obtenir un diplôme du secondaire.

Enfin, Fine et Davis (2003) étudient les trajectoires scolaires des élèves ayant redoublé mais ayant malgré tout obtenu un diplôme du secondaire. Ils mesurent alors la probabilité d'être inscrits dans un cursus de 4 ans ou dans n'importe quel cursus supérieur pour les élèves ayant redoublé ou non à l'aide de régressions logistiques. Ils montrent alors que ces élèves, ayant pourtant déjà "battu les statistiques", ont

---

44. L'auteur estime l'effet d'avoir redoublé sur la probabilité d'avoir un diplôme du secondaire à partir d'un probit avec un ensemble de variables de contrôles.

45. L'auteur a construit une macro SAS pour estimer un probit bivarié.

une probabilité significativement plus faible d'être inscrit dans le supérieur. Les élèves ayant redoublé ont deux fois moins de chance (Odds ratio de 0.47) d'être inscrit dans un cursus universitaire de 4 ans que les élèves qui ont toujours été promus, *toutes choses égales par ailleurs*. Ils ont de même 39 % de probabilité en moins d'être inscrit dans le supérieur, quel que soit le diplôme préparé.

Ces résultats indiquent que même les élèves les plus persistants (dans la poursuite de leurs études) parmi les redoublants ont des trajectoires d'éducation différentes des élèves ayant toujours été promus, à caractéristiques observables données.

## 2 Les effets du redoublement sur l'insertion dans la vie active

Le redoublement semble avoir un impact négatif sur l'insertion dans la vie active en termes de taux d'emploi et de salaire. En France, une étude montre qu'avoir redoublé semble avoir un effet délétère sur le revenu futur du jeune adulte en agissant comme un signal de faible performance envoyé aux employeurs.

Le redoublement semble donc avoir un impact important sur les trajectoires d'éducation des élèves. Aussi, l'étude de l'effet du redoublement sur l'insertion dans la vie active et en particulier, sur le taux d'emploi et le niveau de salaire, requiert une très grande prudence dans la méthode employée. En effet, nous avons déjà montré à quel point le redoublement était endogène, et qu'il était fortement lié aux performances et aux parcours éducatifs des élèves. Or, le taux d'emploi et le niveau de revenu sont conditionnés par les niveaux d'éducation. Il faut donc nécessairement mesurer l'effet à l'aide d'un modèle à plusieurs étages.

À notre connaissance, seul le travail de [Brodaty et al. \(2008\)](#)<sup>46</sup> emploie ce type de stratégie empirique pour mesurer l'effet causal du redoublement sur le taux d'emploi et les rémunérations. Leur stratégie consiste à estimer simultanément trois équations : la première prédit le retard scolaire, la seconde prédit le niveau de fin d'études et enfin la troisième prédit le niveau de revenu (exprimé en logarithme). Le modèle ne peut être identifié que s'il existe au moins une variable spécifique à chacune des deux premières équations qui n'affecte pas le revenu autrement que par son effet sur le redoublement ou le niveau de fin d'études. En se servant des données de *l'enquête génération 92* du CEREQ, les auteurs disposent d'un large panel de jeunes entrés sur le marché du travail en 1992 observés pendant les 5 premières années de leur carrière. Ils construisent une variable pour le niveau d'éducation qui correspond à l'âge de fin d'études théorique pour obtenir un diplôme donné. Par exemple, un individu qui entre sur le marché du travail avec uniquement un Baccalauréat se voit attribué la valeur 18 à sa variable éducation. La variable de revenu est également une mesure synthétique du salaire sur toute la période corrigée de l'effet moyen de l'expérience. Enfin, la mesure du retard scolaire est définie par la différence entre l'âge de fin d'études de l'individu et l'âge moyen de fin d'études pour les jeunes terminant à ce niveau. Par

46. Nous nous basons sur une version révisité daté du 13 février 2013 disponible sur la page de [Gary-Bobo](#)

exemple, un élève ayant obtenu le baccalauréat à 18,5 ans et qui s'est inscrit 2 ans dans le supérieur sans obtenir de diplôme (ayant quitté l'école à 20,5 ans donc) aura un niveau d'éducation de 18 ans et un niveau de retard scolaire de 0,28 car l'âge moyen d'obtention du baccalauréat était de 20,78 années en 1992.

Les auteurs utilisent plusieurs *variables instrumentales* pour identifier leur modèle : la distance à l'université la plus proche lorsque l'élève est entré en 6<sup>ème</sup> et cette distance au carré, le nombre d'établissements d'enseignement professionnel et technologique par habitants dans le département et sa variation sur 10 ans, le taux de chômage moyen dans le département à l'entrée en 6<sup>ème</sup>. Aucune de ces variables ne vérifie les hypothèses nécessaires à l'identification du modèle. Cependant, comme [Dong \(2010\)](#), les auteurs argumentent que les raisons qui rendent ces choix discutables ne sont pas les mêmes et que dès lors, si les résultats sont insensibles à l'utilisation de l'un, l'autre ou l'ensemble, les résultats peuvent être considérés comme valides. On comprend que les auteurs ont fait ces choix faute de mieux et qu'en effet, le peu de variation des résultats entre les différentes spécifications suggère que s'il y a en effet des biais dans ces stratégies, ceux-ci sont sûrement négligeables.

Ainsi, sur la base de ces estimations, une année de retard entraîne une diminution (causale) du revenu moyen pendant les cinq premières années d'environ 9 %. En comparant ces estimations avec le *rendement de l'éducation* calculé dans d'autres études, celui-ci est de même ordre de grandeur ce qui implique qu'un étudiant qui redouble une année pour obtenir un diplôme l'année suivante fait en réalité une opération blanche en termes de perspective de gains salariaux futurs. Les auteurs interprètent ces résultats comme une preuve empirique que le redoublement (ou plutôt le retard) est utilisé par les employeurs comme un signal de productivité des candidats. On peut aussi penser que puisque le redoublement affecte la motivation et la confiance en eux des élèves qui y sont sujets, ceux-ci peuvent être moins dynamiques et ambitieux lors des négociations salariales, que leurs pairs toujours promus, plus sûrs d'eux. Les auteurs ne discutent pas cette possibilité.

Cette revue de la littérature nationale et internationale sur les effets du redoublement a permis de mettre en avant plusieurs conclusions récentes mais aussi de découvrir que certains aspects de la question du redoublement ne sont que très peu couverts. Plusieurs questions restent ainsi peu, ou pas du tout traitées. D'abord, les études existantes ne permettent pas de conclure sur les effets du redoublement dans le second degré. Il n'y a par ailleurs aucune étude traitant de la question du redoublement stratégique, brièvement évoqué dans le premier rapport de la série. Les effets à long terme du redoublement, sur la probabilité d'obtenir un diplôme du supérieur et sur les trajectoires professionnelles sont encore trop rares pour que nous puissions conclure de façon certaines.

---

## Chapitre 2

# Les rôles du redoublement

Le redoublement n'est pas universellement répandu et pratiqué de la même façon. Cependant, ses partisans et détracteurs invoquent en général les mêmes qualités et objections. Meuret (2002) identifie ainsi quatre arguments favorables et cinq arguments contre.

Parmi les arguments pour, on dit que :

- Les élèves ne sont pas capables de suivre dans la classe supérieur par manque de connaissances ;
- Les élèves n'ont pas la maturité nécessaire ;
- Le redoublement joue un rôle incitatif en pesant comme une épée de Damoclès sur le parcours scolaire ;
- Le redoublement permet d'homogénéiser la composition des classes.

À ces arguments, on oppose généralement que

- Le redoublement affecte négativement l'image de soi et donc la motivation ;
- Les élèves revoient inutilement des connaissances déjà acquises ;
- recommencer quelque chose qui a échoué ne garantit pas sa réussite ;
- Le redoublement n'est pas équitable et nuit à l'égalité des chances<sup>1</sup> ;
- Les décisions de redoublement sont prises en fonction du niveau moyen de la classe et sont donc entachées d'arbitraire.

Cette section décrit ce que les partis-prenantes (enseignants, parents, chefs d'établissement,...) attendent du redoublement et comment sont construites leurs croyances. En particulier, nous distinguerons les rôles du redoublement dans l'efficacité pédagogique d'autres fonctions qualifiées de latentes.

---

1. Meuret (2002) rappelle que, à niveau égal, les enfants de pauvres reçoivent de plus basses notes, redoublent plus souvent, ce qui les stigmatise et nuit à leur orientation.

## I Les croyances et attitudes des enseignants et des familles vis-à-vis du redoublement

Des recherches récentes se sont attachées à comprendre l'attachement des enseignants et des parents au redoublement. Elles montrent que les enseignants, les parents et les chefs d'établissement sont effectivement très attachés au redoublement parce qu'ils croient en son efficacité pédagogique. Du point de vue des enseignants en particulier et, à moindre mesure, de celui des chefs d'établissement, un élève qui redouble a, en général, de meilleurs résultats l'année suivante ce qui conforte cette croyance. Le redoublement s'inscrit par ailleurs dans un système de valeurs méritocratiques dans lequel il joue un rôle structurant. Les récents sondages auprès des parents montrent que s'ils pensent en majorité que le redoublement est une bonne chose, ils sont également une majorité à penser qu'il est remplaçable par quelque chose de plus efficace, notamment par l'accompagnement personnalisé. Cette type de réponse indique que l'adhésion des parents au redoublement se *fait faute de mieux*.

Si le redoublement est pratiqué, c'est qu'il existe une forme d'accord social quant à son utilité, où enseignants et parents se retrouvent pour estimer que "*le redoublement est un instrument incontournable de l'acte éducatif*" (Paul et Troncin, 2004). Le récent sondage de OpinionWay (2012) révèle ainsi que 70 % des parents et 64 % des enseignants interrogés<sup>2</sup> sont d'accord avec la phrase "*Le redoublement permet réellement à l'élève de rattraper son retard et d'être mieux préparé pour les classes supérieures*". Les enseignants du privé sont même 80 % à supporter cette assertion. Le redoublement est ainsi largement soutenu par ces deux catégories d'acteurs, qui sont par ailleurs 62 % à penser que le redoublement est "*une bonne chose pour les élèves*".

De façon assez symétrique, 42 % des enseignants et 43 % des parents pensent que "*le redoublement n'aide pas vraiment un élève en difficulté qui continuera à rencontrer les mêmes difficultés*" et respectivement 26 % et 41 % estiment qu'il "*est mauvais pour l'élève qui se retrouve avec des enfants plus jeunes et a déjà étudié le programme*". Ces opinions sont remarquablement homogènes pour les parents selon leurs catégories socio-professionnelles, le niveau de scolarisation de leur enfant et le secteur (public/privé) de l'établissement dans lequel il est inscrit. Il y a des différences plus notables chez les enseignants selon le secteur et le degré dans lequel ils enseignent.

Ainsi, les résultats de ce sondage (avec les limites qui y sont associées) indiquent que les enseignants et les parents français soutiennent largement le redoublement en dépit des nombreuses études qui concluent à son inefficacité (voir chapitre 1). Ils sont toutefois une large majorité à penser que le redoublement pourrait "*être remplacé par d'autres mesures plus efficaces et mieux adaptées, telles que l'accompagnement personnalisé pour les enfants en difficulté*" (77 %). Les enseignants ne sont que

2. Sondage réalisé sur 613 parents et 1100 enseignants selon la méthode des quotas.

56 % à être d'accord avec cette assertion. Cet élément révèle une forme de paradoxe de l'opinion publique : le redoublement est une bonne chose, mais il est remplaçable par quelque chose de plus efficace, notamment par l'accompagnement personnalisé. Ce type de réponse indique que l'adhésion des parents au redoublement se fait faute de mieux.

Dans la littérature scientifique, le point de vue des parents est peu étudié contrairement à celui des enseignants. Paul et Troncin (2004) citent cependant plusieurs études anciennes et étrangères qui révèlent une adhésion massive au redoublement. Les recherches de Troncin (2005) montrent que très peu de parents sont prêts à s'opposer au redoublement de leur enfant en CP, ce que l'auteur interprète comme un signe de confiance des parents envers les enseignants à ce niveau scolaire. Cependant, les auteurs notent que les familles dites « favorisées » sont nettement moins disposées à accepter d'emblée la décision de redoublement, ce qui peut indiquer un plus grand scepticisme à l'égard du redoublement. En particulier, de nombreux redoublements sont aujourd'hui contestés par les parents au primaire, d'après les panels qualitatifs organisés par le Cnesco dont les résultats devront être confortés.

Les résultats d'OpinionWay (2012) sont en adéquation avec la littérature scientifique traitant des croyances des enseignants sur le redoublement (Draelants, 2006; Crahay, 2007; Boraita et Marcoux, 2013; Range et al., 2011; Terry, 2011; Wynn, 2010). Les études plus anciennes que recensent Paul et Troncin (2004) montrent également, et sans ambiguïté, que les enseignants voient dans le redoublement une mesure nécessaire dont les effets positifs surpassent les effets négatifs. La plupart de ces études sont teintées d'une forme de surprise, voire d'exaspération des chercheurs (Draelants, 2008) qui se manifeste très bien dans le titre de l'article de Marcoux et Crahay (2008) : "Mais pourquoi continuent-ils à faire redoubler ?"

Pourtant, les enseignants ont de bonnes raisons de croire aux vertus du redoublement (Xia et Glennie, 2005), notamment parce que "dans la plupart des cas, l'élève redoublant sera un peu meilleur durant son année de redoublement" (Draelants, 2006). Le jugement de l'efficacité pédagogique apparaît alors comme un mélange de bon sens et d'observation qui est très différent de l'approche des chercheurs sur la question. Ces derniers s'interrogent sur ce qui se serait passé pour l'élève redoublant s'il avait été promu. La comparaison ne se fait donc pas sur la base des performances la première et la seconde année, mais sur deux élèves théoriques dont l'un a redoublé et l'autre a été promu (voir chapitre 1).

Les croyances des enseignants se fondent sur plusieurs conceptions des effets attendus du redoublement. À partir d'entretiens menés auprès de 41 enseignants dans le canton de Genève, Marcoux et Crahay (2008) identifient une *vision séquentielle des apprentissages* qui implique qu'un élève ne maîtrisant pas l'une des étapes de cette séquence n'est plus en mesure d'acquérir des compétences aussi robustes. Dès lors, le redoublement apparaît comme une solution adaptée pour solidifier les bases. À cette conception se joint parfois une vision maturationaliste du développement : "L'enseignant arguera alors que la répétition d'une année donne le temps à la maturation de faire son œuvre et que l'élève, retardé d'une année, abordera les apprentissages, objets de difficultés ou de blocages, avec des capacités cognitives améliorées" (Marcoux et Crahay (2008) p512). "Une année supplémentaire leur permettra de

*mûrir et de mieux se préparer à affronter les difficultés de leur scolarité future"* (Paul et Troncin (2004) p. 29). Cette croyance dans l'efficacité du redoublement fonctionne par ailleurs comme un système de référence, protégé par un ensemble d'arguments protecteurs *saturés en attributions externes* (l'élève n'a pas consenti les efforts attendus ; il n'a pas reçu le soutien de ses parents). Il n'est jamais question pour les enseignants d'interroger fondamentalement la pertinence de cette pratique mais davantage l'opportunité de la mettre en œuvre. La question de ce qui se serait passé pour un redoublant s'il avait été promu n'est quasiment jamais soulevée.

Dans la prise de décision concernant le redoublement ou la promotion d'un élève, les résultats scolaires sont, en apparence, importants. Cependant, Marcoux et Crahay (2008) montrent qu'ils ne sont pas les principaux déterminants mais qu'ils jouent davantage un rôle stratégique ou communicationnel à l'égard des élèves et des parents. En effet, animés par la croyance dans les bienfaits potentiels du redoublement, les enseignants débute leur processus de décision par un *"jugement sur l'état psychologique des élèves en difficultés"* (p. 513). Les résultats scolaires jouent alors le rôle d'éléments de validation de ce jugement. Ce dernier est, par ailleurs, fondé sur une vision normative de ce qu'est "le métier d'élève" et "la famille idéale" (ou normale). Ils positionnent alors l'élève sur cette échelle et adoptent un jugement dans lequel la distance à la norme joue un rôle de compensation ou à l'opposé de cumulation.

*"Des enseignants se prononceront pour la réussite d'un élève accumulant les défaillances, mais dont la famille fait preuve d'un potentiel de ressources éducatives importantes. À l'inverse, ils décideront du redoublement d'un élève moyennement faible, mais dont ils estiment que leur famille ne pourra supporter leurs efforts dans le futur"* (Marcoux et Crahay (2008) p 514).

Il en résulte une forme de dilemme moral lorsque, comme dans les situations précédentes, le mérite de l'élève et ses talents ne convergent pas. Dubet (2002) explique que *"l'école démocratique repose sur deux piliers [...] : un principe d'égalité, tous les élèves sont fondamentalement égaux et peuvent prétendre aux mêmes choses ; un principe de mérite, fondant des inégalités justes"*. Or ces deux principes sont antagonistes car ils impliquent de classer les élèves et d'affirmer leur égalité, ce qui oblige à expliquer leurs inégalités de performances comme les conséquences de leur liberté (Draelants, 2008). Au final, cette conception morale de l'enseignement crée un attachement au redoublement qui s'intègre alors comme une institution conforme aux valeurs de l'école démocratique dans lesquels *"tous les élèves sont égaux et les meilleurs sont ceux qui travaillent plus"* (Dubet (2002) p.18).

Le point de vue des chefs d'établissement, bien qu'ayant un rôle fondamental dans la pratique du redoublement, est très peu étudié dans la littérature. A notre connaissance, il n'y a pas d'article s'y intéressant dans les sciences de l'éducation francophone. Quelques articles existent toutefois aux États-Unis. Cannon et Lipscomb (2011) montrent ainsi que la moitié des directeurs d'école interrogés à Los Angeles perçoivent le redoublement comme bénéfique et la majorité estiment qu'un redoublement précoce est plus bénéfique qu'un redoublement tardif. Dans leur étude, Range et al. (2012) montrent

que les enseignants et les principaux d'écoles primaires pensent en majorité que le redoublement aide à prévenir l'échec scolaire future, à maintenir le niveau, permet aux enseignants de fournir d'avantage de soutien aux élèves et motive les élèves à venir à l'école. Ces proportions sont cependant beaucoup plus importantes chez les enseignants que chez les principaux. Cependant, la meilleure façon de réduire le risque de redoubler est, pour ces praticiens, de mobiliser davantage les parents.

c'est aussi le résultat de [Young et Range \(2014\)](#), qui interrogent des chefs d'établissement de deux États des montagnes rocheuses américaines. Pour eux, le travail des enseignants avec les élèves en difficultés est le meilleur moyen de réduire le redoublement mais l'implication des parents est le meilleur moyen de permettre la promotion des élèves.

Ainsi, les enseignants, principaux et directeurs d'écoles semblent attachés au redoublement car cette pratique apparaît bénéfique du point de vue cognitif et moral. Le redoublement joue par ailleurs d'autres fonctions latentes qui lui confèrent davantage encore de légitimité auprès des praticiens.

## II Les fonctions latentes du redoublement

Un autre courant de recherche a adopté une vision *fonctionnaliste* de l'attachement au redoublement et identifié des fonctions latentes, ou cachés. D'abord, le redoublement joue un rôle de menace incitant les élèves à augmenter leurs efforts pour ne pas redoubler. Ensuite, il peut permettre de gérer l'hétérogénéité au sein des classes en plaçant l'élève redoublant dans une situation de léger avantage vis-à-vis de ses pairs nouvellement promus. En outre, le redoublement peut jouer un rôle de signal pour différents acteurs. Pour un enseignant, il indique aux autres enseignants et aux parents qu'il ne transige pas avec les élèves en difficultés. Pour les chefs d'établissement, il permet également de montrer que son établissement est très exigeant et ainsi attirer une population scolaire différente. Enfin, le redoublement offre un pouvoir important aux enseignants sur le parcours scolaire de leurs élèves auquel ils sont attachés.

Au delà de l'attachement des enseignants au redoublement tel que nous l'avons décrit dans la section précédente, [Draelants \(2006, 2008\)](#) a mis en exergue d'autres rôles qui expliquent l'attachement des professeurs à cette pratique, mais également celui des parents, des chefs d'établissement,...

Une première fonction déjà évoquée est celle du rôle incitatif du redoublement sur le travail. Certains voient dans le redoublement une menace incitant les élèves à augmenter leurs efforts pour ne pas redoubler. C'est ce qu'a notamment voulu tester [Koppensteiner \(2014\)](#) en évaluant l'effet de l'interdiction du redoublement au Brésil. Puisque les régions n'ont pas mis en place cette réforme au même moment, la comparaison des États ayant mis en place la promotion automatique avant les autres peut fournir une source de comparaison intéressante. L'auteur se sert précisément de cette *expérimentation naturelle* pour mesurer l'effet de la promotion automatique (par rapport au redoublement) sur les performances des élèves. Il obtient un effet négatif sur les performances des élèves et conclut que le redoublement jouait bien un rôle d'incitation à l'effort. [Draelants \(2008\)](#) constate également en Belgique que "*sans redoublement les notes perdent beaucoup de leur pouvoir régulateur*".

Cet argument est certainement celui qui revient le plus car il s'intègre dans cette conception méritocratique de l'École déjà évoquée ([Dubet, 2002](#)) qui valorise ceux qui travaillent le plus en leur accordant la promotion. Cependant, pour [Gary-Bobo et Robin \(2014\)](#), si le rôle de menace du redoublement est présent, sa portée est assez faible car il existe d'autres sanctions potentielles beaucoup plus fortes que le redoublement, et notamment l'orientation, l'expulsion et même les incitations monétaires.

Par ailleurs, le redoublement peut apparaître comme un moyen de gérer l'hétérogénéité au sein des classes et des établissements. Cet argument repose sur l'idée que l'élève redoublant est moins en difficulté parmi les élèves plus jeunes qui découvrent le programme. Ainsi, le redoublement réduit le différentiel de résultats et contribue à homogénéiser la composition des classes. L'interdiction du redoublement en Belgique en 2001 a donc compliqué le travail des enseignants pour qui la gestion de l'hétérogénéité est

la principale difficulté (Draelants, 2008). Ainsi, l'attachement au redoublement peut aussi refléter un certain attachement à des conditions de travail perçues comme se dégradant. Cependant, ce n'est pas nécessairement la réalité si on en croit Perrenoud (2002) :

*"Chaque enseignant sait bien qu'il accueille de toute façon, au début de chaque année scolaire, un groupe hétérogène et que certains des élèves qui n'ont pas redoublé ne maîtrisent pas pour autant les préalables du programme et qu'ils seront d'emblée en difficulté. Pour assurer une très forte homogénéité des classes, peut-être faudrait-il faire redoubler le tiers d'une cohorte, ce qui serait intolérable."* (p.29)

Il n'est par ailleurs pas certain que l'homogénéité des classes soit plus désirable. Les revues de littérature de Dupriez et Draelants (2004) et Fortin et al. (2011) montrent que de façon générale, les élèves faibles peuvent être pénalisés par des classes homogènes alors que les élèves performants en profitent. Cette question pose également de nombreux problèmes d'ordre éthique que nous ne discutons pas ici.

En outre, le redoublement joue un rôle de signal envoyé par l'enseignant à ses pairs et par l'établissement aux autres établissements, aux enseignants et aux parents. En effet, Gary-Bobo et Robin (2014) expliquent que les enseignants peuvent être favorable au redoublement par peur d'être jugés laxistes par leurs pairs. En ayant laissé passer un élève trop faible dans le niveau supérieur dont un collègue est en charge, l'enseignant s'expose aux critiques, place l'autre enseignant dans une situation délicate, et apparaît complaisant avec les élèves en difficultés.

*"Au moment de décider de la réussite ou de l'échec des élèves, l'enseignant est confronté à un dilemme [...] : faire échouer un élève dont le niveau de performance est à la limite de ce qu'il croit devoir exigé, c'est courir le risque d'interrompre inutilement la scolarité d'un élève, mais cette erreur possible [...], il est fort peu probable qu'on la lui reproche. En revanche, laisser réussir ce même élève, c'est prendre le risque qu'il se montre incapable de suivre l'enseignement du collègue de la classe supérieure ; et là, la probabilité des reproches venant de collègues est bien plus élevée "* (Crahay, 2007).

La décision du redoublement semble donc intervenir dans la construction de la réputation d'un enseignant. Peut-être s'agit-il également d'une réaction d'empathie comme le souligne Pini (1991). Plusieurs études, parmi lesquelles Hutmacher (1993), mettent également en évidence que lorsque l'enseignant "monte avec sa classe" (et donc ne "transmet" pas ses élèves à un collègue), les taux de redoublement sont significativement inférieurs, voire quasi-nuls. Une interprétation possible de ce résultat est l'utilisation du redoublement comme signal envoyé par l'enseignant à ses pairs. Ce n'est toutefois pas la seule (voir Jimerson et al. (2006) et le 3<sup>ème</sup> rapport de la série).

Le redoublement peut également être utilisé par le chef d'établissement pour signaler aux autres établissements, aux parents et aux élèves, à quel point il est exigeant. Dans un contexte où l'établissement est en concurrence dans un environnement institutionnel de quasi-marché (Draelants, 2008), un fort taux

de redoublement permet d'exclure, sans le dire, des élèves d'un niveau insuffisant, car ces derniers iront s'inscrire ailleurs afin d'éviter de redoubler. Cette configuration stratégique peut s'apparenter à une situation de discrimination tarifaire similaire aux modèles de microéconomie des oligopoles. En faisant peser un coût aux élèves et aux parents, l'établissement renforce le phénomène de stratification des élèves par niveaux<sup>3</sup> et par milieu social d'origine (MacLeod et Urquiola, 2012). En l'absence d'incitation à traiter les élèves en difficulté et sans mesure crédible de la valeur ajoutée de l'établissement, "un directeur d'école ou un proviseur préférera typiquement essayer tout à la fois d'embourgeoiser son recrutement et de trier les meilleurs parmi les enfants des classes favorisées" (Gary-Bobo et Robin, 2014). Ce faisant, il signale également au public qu'il ne transige pas avec le niveau des élèves. Les parents privilégiant une stratégie élitiste seront sensibles à cette hiérarchie d'excellence. S'ils jugent un établissement non sur sa capacité à faire réussir un plus grand nombre mais sur sa sélectivité, le redoublement leur apparaît comme un signal crédible de performance de l'établissement.

Cette vision élitiste n'est pas propre aux parents ; Draelants (2008) montre que cette représentation se retrouve également chez les enseignants. "Statistiquement, on observe une association significative entre la perception par les enseignants de la réputation de leur établissement et son degré de sélectivité par rapport à la réussite des élèves. [...] Cela aboutit à perpétuer une situation paradoxale : pour prouver leur qualité et se placer avantageusement sur le marché éducatif, enseignants et écoles sont encouragés à produire des échecs scolaires".

Enfin, le redoublement confère à l'enseignant une influence forte sur le parcours scolaire des élèves auxquels il est attaché. L'estime de certains enseignants pour le redoublement peut alors apparaître comme la revendication par un groupe professionnel de son autonomie et de son influence. "Le redoublement est un des instruments de la sélection méritocratique qui, elle même, symbolise un certain pouvoir enseignant" (Draelants, 2008).

En somme, les enseignants, les parents et les chefs d'établissement semblent d'abord attachés au redoublement parce qu'ils croient en son efficacité pédagogique. Du point de vue des enseignants en particulier, un élève qui redouble a, en général, de meilleurs résultats l'année suivante ce qui conforte cette croyance. Le redoublement s'inscrit par ailleurs dans un système de valeurs méritocratiques dans lequel il joue un rôle structurant. Enfin, il a plusieurs fonctions latentes qui expliquent également l'attachement des parties prenantes à cette institution de l'École.

Dans cette section, nous n'avons pas tenu compte des connaissances des enseignants de la recherche scientifique sur ce sujet, ni comment évoluent leurs avis lorsqu'ils sont informés.

---

3. Plus un élève est mauvais, plus il lui est coûteux d'être dans cet établissement puisque ses chances de redoubler sont plus fortes.

### III La connaissance de la recherche sur le redoublement par les enseignants

Les croyances des enseignants, qui évoluent dans le temps, se fondent principalement sur leur expérience passée et non sur les résultats de la recherche. Malgré les preuves de l'inefficacité du redoublement, la plupart des enseignants reste persuadée que faire redoubler un élève en difficulté est la solution la plus adaptée et la plus efficace. Les connaissances des enseignants sur la recherche sur le redoublement ne permettent pas de les faire changer d'avis quant à l'inefficacité du redoublement. Les recherches montrent un lien important entre les croyances des enseignants à propos du redoublement et la connaissance des résultats de la recherche sur ses effets. De plus, les conceptions psychopédagogiques des enseignants n'ont pas un impact significatif sur ces croyances.

L'évolution des croyances des enseignants est un sujet de recherche très dynamique, notamment dans la littérature francophone où l'équipe du centre de recherche, Développement, Apprentissage et Intervention en Situation Scolaire à l'université de Genève, a réalisé plusieurs travaux très enrichissants sur l'émergence des croyances chez les enseignants et futurs enseignants. Parmi ces publications, [Crahay et al. \(2010\)](#) font une synthèse de la littérature sur la définition des croyances et des connaissances des enseignants et surtout, comment celles-ci évoluent dans le temps.

Cette étude n'est pas spécifique aux connaissances sur le redoublement, mais elle synthétise plusieurs études le concernant. Ainsi, on peut tirer deux enseignements importants de cette étude :

D'abord, il apparaît un lien important entre les croyances à propos du redoublement et la connaissance des recherches sur ses effets. Ensuite, les conceptions psychopédagogiques des enseignants affectent peu ces croyances. [Hausoul \(2005\)](#) a suivi des enseignants qui ont bénéficié d'une formation intensive visant à les sensibiliser aux effets potentiellement négatifs du redoublement lors de leur dernière année de formation initiale. L'auteur constate alors que *"ceux dont les conceptions en matière de redoublement avaient évolué à la suite de cette formation persistent à douter de l'efficacité de cette pratique pédagogique. Toutefois, ils déclarent autant de décisions effectives de redoublement que leurs condisciples de formation qui ne s'étaient pas laissés convaincre et qui, eux-mêmes, présentent des taux d'échec conformes à la norme particulièrement élevée en Communauté française de Belgique. Bref, si l'action de formation semble avoir eu un effet persistant sur les conceptions de certains enseignants en matière de redoublement, celle-ci ne leur a pas donné les armes pour résister à la « loi du milieu ».*" ([Crahay et al., 2010](#), p.104).

[Terry \(2011\)](#) a interrogé des enseignants de l'état de l'Ohio sur leurs croyances concernant le redoublement puis leur a présenté les résultats de la recherche et administré un second questionnaire. Bien que l'étude souffre de plusieurs lacunes méthodologiques importantes, l'auteur met en exergue quelques

résultats intéressants. En particulier, les résultats qualitatifs montrent que les enseignants ne se sont pas laissé convaincre par les résultats de la recherche et continuent de considérer que le redoublement est bénéfique aux élèves. La construction de cette croyance est basée sur leur expérience personnelle et celles des élèves avec lesquels ils ont travaillé dans le passé. Ces résultats sont cohérents avec les résultats de [Marcoux et Crahay \(2008\)](#).

En Belgique, [Crahay et Ory \(2006\)](#) mesure l'évolution des croyances à propos du redoublement de futurs enseignants de six écoles normales<sup>4</sup>. Les informations recueillies montrent que les étudiants de première année se déclarent clairement en faveur du redoublement, tout particulièrement en tout début de formation, alors que l'avis de leurs condisciples de deuxième et de dernière année est, en moyenne, plus mitigé. [Range et al. \(2011\)](#) obtiennent des résultats tout à fait similaires dans deux universités américaines formant des futurs enseignants. Toutefois, [Crahay et Ory \(2006\)](#) observent que l'évolution du positionnement général à l'égard du redoublement ne repose ni sur un développement des conceptions quant aux effets spécifiques (scolaires, cognitifs, affectifs et sociaux) du redoublement, ni sur une transformation de leurs croyances quant aux causes des échecs scolaires.

*"En fin de formation comme en début, les futurs enseignants incriminent les élèves (leur manque d'effort, leur maturité ou encore leurs faibles aptitudes) ainsi que leurs familles et rechignent à mettre en cause l'enseignement reçu ou même le fonctionnement général de l'école. De plus, ils persistent à nier le poids de l'influence de l'origine socioculturelle des élèves dans la fabrication des échecs scolaires. In fine [Crahay et Ory \(2006\)](#) concluent qu'il n'y a pas eu de transformation en profondeur des croyances de la plupart des enseignants en formation" (cité dans [Crahay et al. \(2010\)](#)).*

Les quelques études auxquelles nous avons fait référence sont assez négatives quant à la possibilité de faire évoluer les croyances des enseignants en matière de redoublement. Cependant, [Crahay et Boraita \(2013\)](#) expliquent que la plupart des études sur la question souffrent de problèmes méthodologiques importants. En particulier, la validité psychométrique des questionnaires n'a que très rarement été testée. Ce qui est capturé par les questions n'est donc pas nécessairement le reflet réel des croyances des enseignants. Dans leur récente contribution, [Boraita et Marcoux \(2013\)](#) ont adapté des questionnaires existants, mesuré leur validité psychométrique et construit deux échelles de mesure : l'une concernant les croyances sur le redoublement et l'autre concernant les connaissances de la recherche.

Par ailleurs, les études recensées dans la seconde partie de [Crahay et Boraita \(2013\)](#) ouvrent des pistes nombreuses qui restent à explorer. En particulier, il convient de définir précisément les critères permettant d'identifier ce qu'est une évolution ou un changement positif de croyances.

---

4. Écoles de formation des enseignants en Belgique

---

## Conclusion

Dans ce rapport, nous avons présenté la littérature scientifique s'intéressant aux effets du redoublement sur les performances scolaires, sur des mesures de développement psycho-social et sur les trajectoires scolaires. Nous avons identifié un courant de recherche relativement ancien qui obtenait des effets systématiquement négatifs du redoublement. Cependant, des travaux récents ont mis en lumière certaines faiblesses et mis en oeuvre des techniques d'évaluation plus robustes. Leurs conclusions sont beaucoup plus contrastées quant aux effets délétères du redoublement. Néanmoins, tous s'attachent à dire que le redoublement ne permet pas de lutter efficacement contre les difficultés scolaires. Les comparaisons internationales indiquent par ailleurs que les pays qui pratiquent un fort redoublement ont en moyenne des performances moins élevées et sont souvent plus inégalitaires. Enfin, Le redoublement a systématiquement des effets délétères sur les trajectoires scolaires et semble impacter négativement les conditions d'entrée dans la vie active.

Cependant, cette pratique pédagogique persiste dans de nombreux pays, notamment en France, ce qui révèle un certain attachement des enseignants et des parents. Des recherches récentes se sont attachées à comprendre l'attachement des enseignants et des parents au redoublement. Elles montrent que les enseignants, les parents et les chefs d'établissement sont effectivement très attachés au redoublement parce qu'ils croient en son efficacité pédagogique. Du point de vue des enseignants en particulier et, à moindre mesure, de celui des chefs d'établissement, un élève qui redouble a, en général, de meilleurs résultats l'année suivante ce qui conforte cette croyance. Le redoublement s'inscrit par ailleurs dans un système de valeurs méritocratiques dans lequel il joue un rôle structurant.

Si en France, la communauté éducative reste assez fidèle à la pratique traditionnelle du redoublement, de nombreux pays de l'OCDE ont mis en place des organisations scolaires permettant de le limiter drastiquement, voire de le supprimer, sans dégrader les résultats de leur système scolaire.

Dans le troisième rapport, nous présentons ces différentes alternatives au redoublement.



---

# Annexes

## I La correction du biais de sélection dans l'estimation des effets du redoublement

Le problème fondamental dans l'évaluation d'impact est qu'on n'observe pas ce qui serait arrivé à un individu s'il n'avait pas été affecté par une intervention ou ce qui se serait passé pour une personne si elle en avait bénéficié. Par ailleurs, puisqu'une intervention telle que le redoublement n'arrive pas au hasard aux individus, la différence de performances entre redoublants et promus n'est pas uniquement due au redoublement. Pour mieux mesurer son effet et approcher la relation causale, il faut *minimiser* les différences entre les deux groupes par des méthodes statistiques.

### A Les premières méthodes pour y remédier

Hormis les 4 études expérimentales recensées dans Jackson (1975), les études anciennes (tout particulièrement celles synthétisées dans les premières synthèses de Jackson (1975); Holmes et Matthews (1984); Holmes (1989)) ne corrigent pas ou très peu le problème de sélection du groupe de contrôle. Celles menées depuis le milieu des années 80 ont toutefois été plus sensibles à cette question en contrôlant notamment par les performances passées ou des mesures de QI (Jimerson, 2001).

Ce type de méthodes, pratiquées depuis longtemps et toujours largement utilisées, identifie des caractéristiques observables qui ont un pouvoir prédictif à la fois sur le redoublement et sur les performances et les intègre dans le modèle afin de retirer la part de variation qui leur est imputable. Les chercheurs font alors l'hypothèse qu'une fois ces facteurs contrôlés, le redoublement est comme distribué au hasard, ce qui permet de mesurer l'effet causal du redoublement. Tout se passe comme si le chercheur avait supprimé les différences entre élèves. Bien-sûr, il s'agit d'une hypothèse forte qui est immédiatement invalidée s'il existe des différences inobservables pour les chercheurs qui affectent le redoublement et les performances. Par exemple, un enseignant pourra décider de faire passer un élève aux résultats faibles, mais très motivé et accomplissant de nombreux efforts alors qu'il fera redoubler un élève de niveau comparable mais fournissant peu d'efforts. Le chercheur, en n'observant pas cet élément clé qu'est la motivation, omet une caractéristique importante qui biaise tous ses résultats. Nous verrons dans la section suivante comment remédier à ce problème.

En outre, utiliser les résultats passés ou des mesures de développement cognitifs comme le QI soulève des problèmes de différentes natures et ne permet pas de corriger toutes les différences entre les élèves. Une première difficulté vient du fait que les tests passés, mais aussi les résultats de QI, sont des mesures imparfaites de la caractéristique inobservable que l'on cherche à approcher qu'est la compétence de l'élève. Or, les erreurs de mesure ne se distribuent pas aléatoirement entre les redoublants et les non-redoublants. Plus précisément, lorsqu'un élève jugé comme faible obtient un score bas à un pré-test, il est plus probable que ce score reflète son niveau initial réel qu'une erreur de mesure. Au contraire, si un élève jugé non faible obtient un score bas à un pré-test, il est plus probable que cette discordance

soit due à des erreurs de mesure (ou un "accident" pour cet élève) qu'à la réalité de ses performances. Dans ce cas, si le premier redouble mais pas le second, et qu'on se sert des scores au test standardisé pour comparer ces élèves, alors l'estimation de l'effet du redoublement sera sous-estimé car la prédiction des notes au second test par le premier test sera exagérée par l'erreur de mesure. Cette situation a été décrite et traitée par [Grisay \(1993\)](#) comme un "*artefact des comparaisons de groupes naturels*" ([Grisay, 2001](#)).

D'autre part, les méthodologies qui construisent un groupe de contrôle basé sur des performances passées similaires risquent d'introduire un biais lié au phénomène de *régression vers la moyenne* ([Roderick et Nagaoka, 2005](#); [Mahjoub, 2008](#)). En effet, les élèves qui sont amenés à redoubler ont tendance à l'être après une année particulièrement mauvaise par rapport à leurs résultats moyens (habituels). De même, des élèves d'ordinaire faibles vont être amenés à passer dans la classe supérieure après une année particulièrement brillante. Ainsi, les premiers ont des chances d'être meilleurs l'année suivante alors que les seconds ont de grandes chances de l'être moins, donnant l'illusion d'un effet positif du redoublement lorsque il est mesuré comme une comparaison de ces deux groupes<sup>5</sup> ([Alexander et al., 1994](#); [Shepard et al., 1996](#)).

## B Des méthodes quasi-expérimentales plus robustes mais plus exigeantes

Les méthodes que nous venons de présenter ont pour caractéristique principale d'utiliser quelques variables de contrôle pour réduire le biais de sélection. Elles constituent, certes, une avancée notable mais ne permettent pas d'obtenir des estimateurs sans biais. Cependant, les progrès récents en statistiques et en économétrie ont fait émerger de nouvelles techniques ou familles de techniques qui permettent de mieux estimer l'effet du traitement du redoublement. Au delà des expérimentations contrôlées qui restent la référence mais sont difficilement applicables, il est possible d'utiliser les discontinuités de régression (*regression discontinuity design*), les méthodes d'appariement (*matching*), les méthodes de variables instrumentales et enfin les méthodes d'économétrie structurelle. L'intuition derrière chacune de ces techniques est décrite de façon très compréhensible par [Gary-Bobo et Robin \(2014\)](#).

La discontinuité de régression exploite l'existence d'une règle stricte déterminant le redoublement permettant de mesurer son effet. S'il existe un seuil en dessous duquel tous les élèves redoublent et au dessus duquel tous les élèves passent<sup>6</sup>, on peut s'intéresser aux individus obtenant une note juste sous ce seuil et à ceux ayant eu une note juste au dessus. En effet, puisqu'il existe une part d'aléa dans la notation, il est probable que les élèves se trouvant autour de ce seuil soient si proches qu'on pourrait les intervertir au hasard. Tout se passe "*comme si le fonctionnement normal de l'institution scolaire, en raison d'erreurs aléatoires de notation inévitables, avait de lui même engendré un groupe de traitement et un groupe de contrôle par un choix au hasard*" ([Gary-Bobo et Robin, 2014](#)). Au final, la

5. Le redoublant retourne à son niveau moyen plus élevé et progresse donc alors qu'il se produit exactement l'inverse pour l'élève promu.

6. C'est le cas de plusieurs États américains, du Chili et de l'Uruguay, notamment.

seule différence entre ces élèves juste au dessus et juste en dessous est le redoublement. La comparaison de ces élèves mesure donc l'effet du redoublement. On parle d'estimation *locale* de l'effet du traitement puisqu'on ne peut rien dire a priori de l'effet sur un élève bien meilleur, ou nettement moins bon. On ne peut que mesurer l'effet sur les élèves *tangents* ou *marginiaux*. Par ailleurs, l'emploi de ces méthodes requiert un certain nombre de conditions que nous ne détaillons pas ici<sup>7</sup>.

L'appariement consiste à identifier un ou plusieurs *voisins similaires* afin de constituer un groupe de contrôle ayant un maximum de caractéristiques communes. Cette méthode vise en quelque sorte à "*remédier à l'absence de jumeaux parfaits de chaque élève en appariant chaque redoublant avec un élève qui lui ressemble en apparence le plus possible*" (Gary-Bobo et Robin, 2014). Cette technique ressemble dans son esprit aux méthodes classiques utilisant des variables de contrôles. En ce sens, elle souffre du même problème majeur : elle ne tient pas compte de l'effet des variables inobservables. Mais elle a de nombreux avantages par rapport à ces méthodes qui la rendent nettement plus performante lorsque les bases de données sont suffisamment fournies. En effet, la condition générale de validité de cet estimateur est d'avoir un nombre d'observations suffisamment élevé pour identifier plusieurs "voisins" et de disposer d'un grand nombre de variables pour disposer d'assez d'informations et rapprocher aux mieux les individus du groupe traité avec ses pairs du groupe de contrôle. En ce sens, le matching est un moyen de créer un groupe de contrôle lorsqu'il n'y en a pas et donc d'estimer l'effet du redoublement (ou plus généralement, du traitement).

Cependant, le problème *d'endogénéité* du redoublement demeure avec cette technique en raison de variables inobservables pour le statisticiens, mais qui contribuent à déterminer à la fois les résultats scolaire et la probabilité de redoubler.

*"Si les élèves ont des talents ou "dons", cachés aux yeux du statisticien, mais observés par le professeur, ces talents ou leur absence expliquent à la fois les résultats finaux et la décision du redoublement. une méthode naïve d'analyse de la corrélation entre résultats finaux et traitement [...] aura donc pour effet d'imputer au redoublement une partie des effets de la cause inobservable à ce redoublement"* (Gary-Bobo et Robin, 2014).

Pour résoudre ce problème, on peut se servir de la méthode des variables instrumentales.

Il s'agit d'une méthode économétrique qui exploite l'existence d'un facteur complètement indépendant de l'élève (et donc de ses caractéristiques et de ses performances) qui influence directement l'occurrence du redoublement. Intuitivement, le redoublement dépend des caractéristiques inobservable de l'élèves, mais pas seulement. Il peut être déterminé par des éléments extérieurs à l'élèves qui affectent sa probabilité de redoubler. Si on parvient à identifier un élément qui modifie la probabilité qu'un élève redouble, mais n'a pas d'impact direct sur ses performances, on peut s'en servir pour mesurer l'effet causale du redoublement.

---

7. Il faut notamment que la discontinuité n'en crée pas sur d'autres variables que celles qui intéressent le statisticien. Si par exemple, au voisinage de la discontinuité, les profils des élèves sont différents, alors l'estimation n'est pas valide. le lecteur intéressé pourra se reporter au manuel de Angrist et Pischke (2008) ou à l'article de Lee et Lemieux (2010).

Dans la littérature économique, cette méthode est très employée, en particulier lors d'expérimentations contrôlées où toutes les personnes tirées au hasard ne participent pas au programme. Dans ce cas, le tirage au sort sert d'instrument pour expliquer la participation et permet de mesurer l'effet de se voir proposer le programme. Ce paramètre, appelé en anglais *Intention to treat (ITT)*, est proportionnel à l'effet du traitement. S'il est significatif, le traitement est efficace.

Lorsqu'il n'y a pas de tirage au sort, on peut trouver des instruments dans *la nature*, mais ceux-ci sont rares et souvent peu crédibles. En effet, il est difficile d'isoler un facteur qui a un effet fort sur la probabilité d'être traité (dans notre cas, de redoubler) sans influencer aussi la variable d'intérêt tel que les performances scolaires. Par exemple, il peut être tentant d'utiliser le nombre de jours d'hospitalisation faisant suite à un accident intervenu pendant l'année, n'ayant pas laissé de séquelles, pour *prédire* le redoublement. Le fait que l'élève ait été absent de l'école augmente vraisemblablement sa probabilité de redoubler, et il n'y a *a priori* pas de raison pour que l'accident sans séquelles ait un effet sur les performances scolaires de l'élève quelques années plus tard. L'accident est le résultat d'un aléa indépendant de son talent, il est à priori tout à fait indépendant. Cependant, on peut tout à fait imaginer des raisons pour lesquelles le niveau scolaire d'un élève et sa probabilité d'avoir un accident grave ne seraient pas indépendants. Par exemple, la pratique de sports extrêmes peut être associée à la fois au risque d'avoir un accident et aux performances de l'élève. Ainsi, même dans le cas d'un événement aussi fortuit qu'un accident, la validité de l'instrument peut être mise en cause.

Enfin, les méthodes d'économétrie structurelle peuvent être appliquées à la modélisation du redoublement et l'ont d'ailleurs été (Grisay, 2001; Cosnefroy et Rocher, 2005; Gary-Bobo et al., 2013). Les modèles structurels s'appuient sur la théorie pour imposer des restrictions ou définir le sens des relations entre des variables afin de faciliter l'identification et l'estimation des effets étudiés. Ces méthodes sont contestées par ceux qui se méfient de la théorie car l'interprétation des résultats est bien-sûr conditionnelle à la validité du modèle théorique sous-jacent.

## Interprétation d'un modèle Logit

### Définition

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

Un modèle Logit estime la probabilité de réalisation d'un événement binaire, redoubler ou non, en fonction d'un ensemble de paramètres dont on mesure l'influence. De façon intuitive, on suppose que la réalisation de l'événement (redoubler ou pas) dépend directement d'une variable latente, ou caché, mais dont la distribution de probabilité est connue. Les variables explicatives n'ont alors plus directement d'effet sur le redoublement mais sur cette variable latente.

### Ajustement du modèle

Les paramètres à prendre en compte dans le modèle doivent avoir certaines propriétés importantes pour être valides. Pour introduire une dimension explicative dans le modèle, celle-ci doit être *exogène* au phénomène étudié, c'est-à-dire qui n'est pas déterminée par celui-ci. On ne peut introduire dans le modèle que des variables exogènes. Par exemple, il serait erroné d'introduire une mesure de performance scolaire de PISA comme variable de contrôle car celle-ci est en partie causée par le redoublement.

### Analyse des résultats

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

### Odds ratios ou rapport de côtes

Les Odds ratios sont un indicateur de la probabilité relative d'avoir redoublé plutôt que pas redoublé associée aux modalités d'une variable explicative. Par exemple, si on cherche à mesurer l'effet du genre sur la probabilité d'avoir redoublé, on peut utiliser l'Odd ratio qui sera le rapport entre la probabilité d'avoir redoublé plutôt que pas redoublé pour un garçon (resp. une fille) sur la probabilité d'avoir redoublé plutôt que pas redoublé pour une fille (resp. pour un garçon). L'interprétation de l'Odd ratio est assez intuitive et lisible. S'il est supérieur à 1, le paramètre augmente la probabilité d'avoir redoublé.

S'il est inférieur à 1, il la diminue. La distribution logistique a l'avantage de fournir directement les Odds ratios attachés au paramètre en calculant l'exponentiel du coefficient. Comme pour toute estimation, on estime les erreurs standards et on teste la significativité des coefficients.

## II Commentaire technique sur l'ajout de mauvaises variables de contrôle

La note *PISA in Focus* (OCDE, 2014) présente les Odds ratios de régressions logistiques estimant la probabilité de redoublement en fonction du statut social et des scores aux tests de mathématiques, sciences et lecture de PISA 2012. L'objectif de ce modèle est de mettre en exergue la corrélation entre statut social et redoublement à *niveau donné* dans les discipline testées.

Le problème de cette démarche est qu'elle mobilise des variables dont le but est de contrôler en partie le biais de sélection, mais qui sont elles-mêmes affectées par la variable d'intérêt<sup>8</sup>. On illustre ce problème en présentant une version simplifiée de leur modèle.

Soit  $R_i$  une variable binaire prenant la valeur 1 si l'élève déclare avoir redoublé et 0 sinon. Cette variable est déterminée par un index latent  $R_i^*$ , correspondant à la *propension à redoubler* tel que

$$\begin{aligned} R_i &= 1 & \text{si } R_i^* > 0 \\ R_i &= 0 & \text{si } R_i^* \leq 0 \end{aligned}$$

On définit simplement  $S_i$  le statut social de l'élève tel que  $S_i = \mathbb{1}\{\text{défavorisé}\}$ .

Supposons que la propension à redoubler d'un élève dépende théoriquement de son statut social et de ses performances initiales  $P_i$ , qu'on ne peut observer. Le modèle estimant le redoublement est alors :

$$R_i^* = \alpha + \beta S_i + \gamma P_i + \mu_i \quad (1)$$

où  $\mu_i$  est une perturbation de moyenne nulle et de variance finie. Ici,  $\mu_i$  satisfait  $E(\mu_i S_i) = E(\mu_i P_i) = 0$  par définition. On s'attend à ce que la probabilité augmente lorsque l'élève est défavorisé et donc  $\beta > 0$ . De façon similaire, on s'attend à ce que la probabilité de redoubler diminue lorsque l'élève est plus performant, ce qui implique  $\gamma < 0$ .

Si on pouvait observer  $P_i$ , alors les estimations de ce modèle seraient tout à fait valides et le coefficient qui nous intéresse  $\beta$  serait correctement estimé par un modèle Logit ou Probit. Or  $P_i$  est inconnu. Cependant, nous connaissons les résultats de l'élève au test PISA qu'on note  $p_i$ .

Le modèle que nous estimerons sera dans ce cas :

$$R_i^* = \hat{\alpha} + \hat{\beta} S_i + \hat{\gamma} p_i + \mu_i \quad (2)$$

Comme les tests PISA sont postérieurs au redoublement, ils pourraient être en partie déterminé par le redoublement, ce qui pose problème. Supposons que ce ne soit pas le cas. En revanche, les autres analyses de OCDE (2013) ont largement montré le lien entre les scores aux tests PISA et le statut

8. Voir notamment (Angrist et Pischke, 2008, p47-51)

social. Pour faire simple, supposons qu'il existe un lien parfaitement déterministe entre les scores aux tests PISA et le statut social tel que :

$$p_i = \pi_0 + \pi_1 P_i + \pi_2 S_i \quad (3)$$

Dans le cas présent, on s'attend à ce que  $\pi_1$  soit positif (plus l'élève est performant, meilleurs sont ses résultats à PISA) et à ce que  $\pi_2$  soit négatif (les performances à PISA sont plus faibles pour les élèves défavorisés).

En se servant de l'équation (3) pour remplacer  $P_i$  dans l'équation (1), on obtient :

$$R_i^* = \alpha - \frac{\pi_0}{\pi_1} \gamma + \left( \beta - \frac{\pi_2}{\pi_1} \gamma \right) S_i + \frac{\gamma}{\pi_1} p_i + \mu_i \quad (4)$$

Ainsi, L'estimation de l'équation (2) donne les résultats suivant

$$\begin{cases} \hat{\alpha} = \alpha - \frac{\pi_0}{\pi_1} \gamma \\ \hat{\beta} = \beta - \frac{\pi_2}{\pi_1} \gamma \\ \hat{\gamma} = \frac{\gamma}{\pi_1} \end{cases}$$

Or, puisque  $\pi_2$  et  $\gamma$  sont tous deux théoriquement négatifs, le coefficient estimé par le modèle (2) est fortement diminué par l'effet du statut sur les performances au test PISA au point que même le signe du coefficient peut être faux si la corrélation entre statut social et résultat PISA est suffisamment forte. En effet, dans ce cas,  $\frac{\pi_0}{\pi_1} \gamma > \beta$ , produisant ainsi une estimation de la corrélation entre statut social et redoublement de signe contraire au véritable effet.



---

# Bibliographie

- Afsa, C. (2011). Faut-il supprimer le redoublement ? Working paper, DEPP A.
- Alet, E. (2010, Janvier). Is grade repetition a second chance ? Working paper.
- Alet, E., L. Bonnal, et P. Favard (2013). Repetition : Medicine for a Short-run Remission. *Annales d'Economie et de Statistique* (111-112), 9.
- Alexander, K., D. Entwisle, et S. Dauber (1994). *On the success of failure : a reassessment of the effects of retention in the primary grades*. New-York : Cambridge University Press.
- Allen, C. S., Q. Chen, V. L. Willson, et J. N. Hughes (2009). Quality of research design moderates effects of grade retention on achievement : a meta-analytic, multi-level analysis. *Educ Eval Policy Anal.* 31, 480–499.
- Allensworth, E. (2004). Ending social promotion : dropout rates in chicago after implementation of the eight-grade promotion gate. Technical report, Consortium on Chicago school research.
- Angrist, J. et A. Krueger (1990). The effect of age at school entry on educational attainment : an application of instrumental variables with moments from two samples. Working Papers 654, Princeton University, Department of Economics, Industrial Relations Section. <http://ideas.repec.org/p/pri/indrel/dsp014f16c282x.html>.
- Angrist, J. D. et G. W. Imbens (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Angrist, J. D. et J.-S. Pischke (2008, Décembre). *Mostly Harmless Econometrics : An Empiricist's Companion* (1 ed.). Princeton University Press. <http://www.worldcat.org/isbn/0691120358>.
- Berkowitz, D., M. Caner, et Y. Fang (2008). Are "nearly exogenous instruments" reliable ? *Economics Letters* 101, 20–23.
- Boraita, F. et G. Marcoux (2013). Adaptation et validation d'échelles concernant les croyances des futurs enseignants et leurs connaissances des recherches à propos du redoublement. *Mesure et Évaluation en Éducation* 36(1), 49–81.

- Bourny, G., S. Keskaik, et S. Franck (2010.). L'évolution des acquis des élèves de 15 ans en culture mathématique et en culture scientifique. premiers résultats de l'évaluation internationale pisa 2009. Note d'information, MENESR-DEPP.
- Bressoux, P. (2007, septembre). L'apport des modèles multiniveaux à la recherche en éducation. *Éducation et Didactique* 1(2), 73–88. <http://education-et-didactique.espe-bretagne.fr/IMG/pdf/Bressoux.pdf>.
- Brodaty, T., R. J. Gary-Bobo, et A. Prieto (2008). Does speed signal ability? The impact of grade repetitions on employment and wages. CEPR Discussion Papers 6832, C.E.P.R. Discussion Papers. <http://ideas.repec.org/p/cpr/ceprdp/6832.html>.
- Buckles, K. S. et D. M. Hungerman (2013, Juillet). Season of birth and later outcomes : old questions, new answers. *The Review of Economics and Statistics* 95(3), 711–724. <http://ideas.repec.org/a/tpr/restat/v95y2013i3p711-724.html>.
- Byrnes, D. et K. Yamamoto (1986). Views of grade repetition. *Journal of Research and Development in Education* 10, 14–20.
- Cannon, J. S. et S. Lipscomb (2011). Early grade retention and student success : evidence from los angeles. Technical report.
- Chen, Q., J. N. Hughes, et O.-m. Kwok (2014). Differential growth trajectories for achievement among children retained in first grade. *The Elementary School Journal* 114(3), 327–353.
- Chenu, F., V. Dupont, M. Lejong, V. Staelens, et A. Grisay (2011). Maintien et redoublement en début de scolarité. Technical report, Département Éducation et Formation, Université de Liège.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum. <http://books.google.fr/books?id=T10N21RA09oC>.
- Corman, H. (2003). The effects of state policies, individual characteristics, family characteristics, and neighbourhood characteristics on grade repetition in the united states. *Economics of Education Review* 22, 409–420.
- Cosnefroy, O. et T. Rocher (2005). Le redoublement au cours de la scolarité obligatoire : Nouvelles analyses, même constats. In *Les dossiers*, Number 166. MENESR-DEPP.
- Crahay, M. (2004). Peut-on conclure à propos des effets du redoublement ? *Revue Française de Pédagogie* 148, 11–23.
- Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire?*, Chapter 3 : Que pensent les enseignants du redoublement?, pp. 123–162. De Boeck Supérieur. [www.cairn.info/peut-on-lutter-contre-l-echec-scolaire--9782804154158-page-123.htm](http://www.cairn.info/peut-on-lutter-contre-l-echec-scolaire--9782804154158-page-123.htm).

- 
- Crahay, M. et F. Boraita (2013). Les croyances des futurs enseignants : est-il possible de les faire évoluer en cours de formation initiale et, si oui, comment ? *Revue Française de Pédagogie* 183, 99–158. <http://rfp.revues.org/4186?lang=en>.
- Crahay, M. et P. Ory (2006). Les représentations des normaliens en matière de redoublement évoluent-elles au cours de leur formation à l'école normale ? In *Communication présentée au 4e congrès des chercheurs en éducation, Bruxelles*.
- Crahay, M., P. Wanlin, I. Issaieva, et I. Laduron (2010). Fonctions, structuration et évolution des croyances (et connaissances) des enseignants. *Revue Française de Pédagogie* 172, 85–129. <http://rfp.revues.org/2296?lang=en>.
- Cunha, F., J. Heckman, et S. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. Nber working paper, National Bureau of Economic Research.
- D'Haultfoeuille, X. (2010, Janvier). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics* 154(1), 115. [http://www.crest.fr/ckfinder/userfiles/files/Pageperso/xdhault/idse1\\_Vfin.pdf](http://www.crest.fr/ckfinder/userfiles/files/Pageperso/xdhault/idse1_Vfin.pdf).
- Dong, Y. y. (2010). Kept back to get ahead? kindergarten retention and academic performance. *European Economic Review* 54, 219–236.
- Draelants, H. (2006, Juillet). Le redoublement est moins un problème qu'une solution : comprendre l'attachement social au redoublement en Belgique francophone. *Les cahiers de Recherche en Education et Formation* 52, 1–25. <http://halshs.archives-ouvertes.fr/halshs-00563874>.
- Draelants, H. (2008). Les fonctions latentes du redoublement. *Education et Société* 21, 163–180.
- Désir, C. (2014). Le redoublement : un échec du système scolaire en Fédération Wallonie-Bruxelles ? Technical report, Groupe socialiste du parlement de la Fédération Wallonie-Bruxelles.
- Dubet, F. (2002). Pourquoi ne croit-on pas les sociologues. *Education et Société* 9, 13–25.
- Dupriez, V. et H. Draelants (2004). Classes homogènes versus classes hétérogènes : les apports de la recherche à l'analyse de la problématique. *Revue Française de Pédagogie* 148, 145–165. [http://ife.ens-lyon.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP\\_RF148\\_12.pdf](http://ife.ens-lyon.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP_RF148_12.pdf).
- Duru-Bellat, M., N. Mons, et B. Suchaut (2004). Caractéristiques des systèmes éducatifs et compétences des jeunes à 15 ans : l'éclairage des comparaisons entre pays. Technical Report 66, Cahiers de l'IREDU.
- Ehmke, T., B. Drechsel, et C. H. Cartensen (2010). Effect of grade retention on achievement and self concept in science and mathematics. *Studies in Educational Evaluation* 36, 27–35.

- Eide, E. R. et M. H. Schowalter (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review* 20, 563–576.
- Figlio, D. N. et L. S. Getzler (2002). Accountability, ability and disability : gaming the system. Nber working paper, National Bureau of Economic Research. <http://bear.warrington.ufl.edu/figlio/w9307.pdf>.
- Fine, J. G. et J. M. Davis (2003). Grade retention and enrollment in post-secondary education. *Journal of School Psychology* 41(6), 401 – 411. <http://www.sciencedirect.com/science/article/pii/S0022440503000992>.
- Fortin, L., M. Filiault, A. Plante, et M.-F. Bradley (2011, Avril). Recension des écrits sur le regroupement homogène ou hétérogène des élèves. Technical report, Chaire de recherche de la Commission scolaire de la Région-de-Sherbrooke sur la réussite et la persévérance scolaire. [http://www.csrqs.qc.ca/fileadmin/user\\_upload/Page\\_Accueil/Enseignants/Fenetre\\_pedagogique/PEPS/Regroupementhomogene-heterogene.pdf](http://www.csrqs.qc.ca/fileadmin/user_upload/Page_Accueil/Enseignants/Fenetre_pedagogique/PEPS/Regroupementhomogene-heterogene.pdf).
- Fruehwirth, J. C., S. Navarro, et Y. Takahashi (2011). How the timing of grade retention affects outcomes : identification and estimation of time varying treatment effect. CIBC working paper, Centre for Human Capital and Productivity. <http://ideas.repec.org/p/uwo/hcuwoc/20117.html>.
- Fruehwirth, J. C., S. Navarro, et Y. Takahashi (2014). A framework for the analysis of time-varying treatment effects : how the timing of grade retention affects outcomes. Working Papers 2011-015. [https://sites.google.com/site/janecooleyfruehwirth/TVTE\\_apr\\_4\\_2014\\_FULL.pdf?attredirects=0](https://sites.google.com/site/janecooleyfruehwirth/TVTE_apr_4_2014_FULL.pdf?attredirects=0), revisité et soumis à nouveau pour publication.
- Fumel, S., S. Keskaik, et J. Girard (2010, Décembre). L'évolution des acquis des élèves de 15 ans en compréhension de l'écrit. premiers résultats de l'évaluation internationale pisa 2009. Note d'information, MENESR-DEPP.
- Gary-Bobo, R. et J.-M. Robin (2014, Janvier). La question des redoublements : analyse économique et problèmes statistiques. *Revue Economique* 1(65), 5 – 45. [http://www.cairn.info/resume.php?ID\\_ARTICLE=RECO\\_651\\_0005](http://www.cairn.info/resume.php?ID_ARTICLE=RECO_651_0005).
- Gary-Bobo, R. J., M. Goussé, et J.-M. Robin (2013, Décembre). Grade retention and unobserved heterogeneity. Working paper.
- Goldstein, H. (2004, Novembre). International comparisons of student attainment : some issues arising from the pisa study. *Assessment in Education* 11(3), 319–330.
- Goos, M. (2013). *Grade retention. The role of the national education policy and the effects on students' academic achievement, psychosocial functioning, and school career*. inedited, Katholieke Universiteit Leuven.

- 
- Goos, M., J. V. Damme, P. Onghena, K. Petry, et J. Bilde (2013). First grade retention in the flemish educational context : effects on children's academic growth, psychological growth, and school career thrthrough primary education. *Journal of School Psychology* 51, 323–347.
- Greene, J. P. et M. A. Winters (2007). Revisiting grade retention : an evaluation of florida's test-based promotion policy. *Education Finance and Policy* 2(4), 319–340.
- Greene, J. P. et M. A. Winters (2009). The effects of exemptions to florida's test-based promotion policy : Who is retained ? who benefits academically ? *Economics of Education Review* 28, 135–142.
- Grenet, J. (2010). La date de naissance influence-t-elle les trajectoires scolaires et professionnelles ? une évaluation sur données françaises. *Revue Économique* 61(3), 589–598.
- Grisay, A. (1993). Le fonctionnement du collège et ses effets sur les élèves de 6<sup>ème</sup> et 5<sup>ème</sup>. *Education et Formations* 32, MEN-DEP.
- Grisay, A. (2001). Évaluer les dispositifs de prise en charge d'élèves faibles (ou fort) : l'utilisation de groupes naturels entraine des artefacts. *Cahiers du Service de Pédagogie Expérimentale - Université de Liège* 7-8, 129–151.
- Hausoul, E. (2005). Quels sont les effets de l'insertion professionnelle des enseignants sur leurs représentations sociales et, plus précisément, sur leur représentation de l'échec scolaire ? Master's thesis, Université de Liège. Non publié.
- Heckman, J., R. Pinto, et P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–86.
- Heim, A. (2014, Juin). Au mauvais endroit au mauvais moment ? le devenir des primo-entrants sur un marché du travail en crise. *BSI-Economics*. <http://www.bsi-economics.org/index.php/macroeconomie/item/379-primo-entrants-marche-travail-crise>.
- Hill, A. J. (2014, decembre). The costs of failure : negative externalities in high school course repetition. *Economics of education review* 43.
- Holmes, C. (1989). Grade-level retention effects : a meta-analysis of research studies. In M. Smith (Ed.), *Flunking grades : research and policies on retention*, pp. 16–33. London : The falmer press.
- Holmes, C. et K. Matthews (1984). The effects of nonpromotion on elementary and junior high school pupils : a meta analysis. *Review of Educational Research* 54, 225–236.
- Hong, G. et S. W. Raudenbush (2006). Evaluating kindergarten retention policy : A case study of causal inference for multilevel observational data. *Journal of American Statistical Association* 101(475), 901–910.

- Hugonnier, B. et M. Quéré (2012, 15 Mai). L'efficacité de l'école : enjeu de justice et de compétitivité. In *Compte Rendu*. rapporté par Jean-Pierre Boisivon.
- Hutmacher, W. (1993). *Quand la réalité résiste à la lutte contre l'échec scolaire : analyse du redoublement dans l'enseignement primaire genevois*, Volume 36. Service de la Recherche Sociologique. <http://www.ge.ch/recherche-education/doc/publications/cahiers/36/quand.pdf>.
- Im, M. H., J. N. Hughes, O.-m. Kwok, S. Puckett, et C. A. Cerda (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology* 51, 349–365.
- Jackson, G. (1975). The research evidence on the effect of grade retention. *Review of Educational Research* 45, 613–635.
- Jacob, B. et L. Lefgren (2004). Remedial education and student achievement : a regression-discontinuity analysis. *The review of Economics and Statistics* 86(02), 226–244.
- Jacob, B. et L. Lefgren (2007). The effect of grade retention on high school completion. Nber working paper, National Bureau of Economic Research.
- Jimerson, S. R. (1999). On the failure of failure : examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology* 37(3), 242–272.
- Jimerson, S. R. (2001). Meta analysis of grade retention research : Implications for practice in the 21st century. *school Psychology Review* 30, 420–437.
- Jimerson, S. R., S. M. W. Pletcher, K. Graydon, B. L. Schnurr, A. B. Nickerson, et D. K. Kundert (2006). Beyond grade retention and social promotion : promoting the social and academic competence of students. *Psychology in the Schools* 43(1), 85–97.
- Jost, R. (2007). À la découverte de l'école en finlande. *Activités Mathématiques et Scientifiques* 62, 47–77. [http://www.mlfmonde.org/IMG/pdf/47\\_77\\_ams62.pdf](http://www.mlfmonde.org/IMG/pdf/47_77_ams62.pdf).
- Koppensteiner, M. F. (2014). Automatic grade promotion and student performance : evidence from brazil. *Journal of Development Economics* 107, 277290.
- Lafontaine, D. et C. Monseur (2011). Analyse des causes et conséquences du maintien en 3<sup>ème</sup> maternelle. Technical report, Unité d'analyse des Systèmes et des Pratiques d'enseignement, Université de Liège.
- Lee, D. et T. Lemieux (2010). Regression discontinuity design in economics. *Journal of Economic Literature* 68(2), 281–355.

- 
- Lorence, J. (2014, Novembre). Third-grade retention and reading achievement in texas : a nine year panel study. *Social Science Research* 48, 1–19. <http://www.sciencedirect.com/science/article/pii/S0049089X14000994>.
- MacLeod, W. B. et M. Urquiola (2012, Août). Anti-Lemons : School Reputation, Relative Diversity, and Educational Quality. IZA Discussion Papers 6805, Institute for the Study of Labor. <http://ideas.repec.org/p/iza/izadps/dp6805.html>.
- Mahjoub, M.-B. (2008). *The treatment effect of grade repetition*. Ph. D. thesis, Paris I, Sorbonne.
- Manski, C. F. (1993, Juillet). Identification of endogenous social effects : the reflection problem. *The Review of Economic Studies* 60(3).
- Marcoux, G. et M. Crahay (2008). Mais pourquoi continuent-ils à faire redoubler ? essai de compréhension du jugement des enseignants concernant le redoublement. *Revue Suisse des Sciences de l'Éducation* 30, 501–518.
- McEwan, P. J. et J. S. Shapiro (2008). The benefits of delayed primary school enrollment : discontinuity estimates using exact birth dates. *The Journal of Human Resources* 43(1), 1–29.
- Meuret, D. (2002, 16 Janvier). Le redoublement est il efficace ? les réponses de la recherche en éducation. In *Intervention à l'IUFM de Cergy-Pontoise*.
- Muthen, B. (2004). *Latent variable analysis : Growth mixture modeling and related techniques for longitudinal data*, Chapter 19, pp. 345368. Newbury Park, CA : Sage. : Newbury Park, CA : Sage Publications.
- Nagaoka, J. et M. Roderick (2004). Ending social promotion : the impact of grade repetition. Technical report, Consortium in Chicago School Reform.
- National Commission on Excellence in Education (1983). A nation at risk : the imperative for educational reform. Technical report, US gouvernement.
- OCDE (2013). *PISA 2012 Results : What Makes Schools Successful ? Resources, Policies and Practices*, Volume IV. PISA. <http://dx.doi.org/10.1787/9789264201156-en>.
- OCDE (2014, septembre). Are disadvantaged students more likely to repeat grades ? Number 43. Pisa in Focus.
- OECD (2009). *PISA Data Analysis Manual : SPSS and SAS, Second Edition*, Volume 2nd edition. OECD.
- OpinionWay (2012, Novembre). Le redoublement à l'école, quels ressentis des enseignants et des parents. Sondage. <http://www.apel.fr/images/stories/apel-opinionway-redoublement.pdf>.

- Oreopoulos, P., T. von Wachter, et A. Heisz (2012, Janvier). The short- and long-term career effects of graduating in a recession. *American Economic Journal : Applied Economics* 4(1), 1–29. <http://ideas.repec.org/a/aea/aejapp/v4y2012i1p1-29.html>.
- Paul, J.-j. et T. Troncin (2004, Décembre). Les apports de la recherche sur l'impact du redoublement comme moyen de traité les difficultés scolaire au cours de la scolarité obligatoire. Rapport 14, Haut Conseil de l'évaluation de l'école (HCÉé).
- Perrenoud, P. (2002). *Les cycles d'apprentissage : une autre organisation du travail pour combattre L'échec scolaire*. Presses de l'Université du Québec. <http://books.google.fr/books?id=CJIFFf9PPyWC>.
- Pini, G. (1991). Effets et méfaits du discours pédagogique : échec scolaire et redoublements vus par les enseignants. *Éducation et Recherche* 3, 255–272.
- Range, B. G., C. R. Holt, J. Pijanowski, et S. Young (2012, Printemps). The perceptions of primary grade teachers and elementary principals about the effectiveness of grade-level retention. *The Professional Educator* 36(1).
- Range, B. G., D. A. D. Yonke, et S. Young (2011). Preservice teacher beliefs about retention : how do they know what they don't know? *Journal of Research in Education* 21(2), 77–99. [http://www.eeraonline.org/journal/files/v21/JRE\\_v21n2\\_Article\\_7\\_Range\\_et\\_al.pdf](http://www.eeraonline.org/journal/files/v21/JRE_v21n2_Article_7_Range_et_al.pdf).
- Reynold, A. (1992). Grade retention and school adjustment : an exploratory analysis. *Educational Evaluation and Policy Analysis* 14(2), 101–121.
- Roderick, M. (1995). Grade retention and school dropout : policy debate and research questions. *The reseach bulletin* 15, 88–92.
- Roderick, M. et J. Nagaoka (2005). Retention under chicago's high stakes testing program : helpful, harmful or harmless? *Educational Evaluation and Policy Analysis* 27(4), 309–340.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5), 688–701.
- Schwerd, G. et M. R. West (2012, Septembre). The effects of test-based retention on student outcomes over time : regression discontinuity evidence from florida. Working paper PEPG, Program on Education Policy and Governance.
- Seibel, C. et J. Levasseur (1983). Les apprentissages instrumentaux et le passage en cours préparatoire au cours élémentaire. *Education et Formations* 2, 3–24. [http://ife.ens-lyon.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP\\_RF067\\_1.pdf](http://ife.ens-lyon.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP_RF067_1.pdf).

- 
- Shepard, L., M. Smith, et S. Marion (1996). Failed evidence on grade retention. *Psychology in the Schools* 33(3), 251–406.
- Silbergliitt, B., J. J. Appleton, M. K. Burns, et S. R. Jimerson (2006). Examining the effects of grade retention on student reading performance : A longitudinal study. *Journal of School Psychology* 44, 255-270.
- Stearns, E., S. Moller, B. Judith, et S. Potochnick (2007). Staying back and dropping out : the relationship between grade retention and school dropout. *Sociology of Education* 3, 210–240.
- Terry, S. E. (2011). *Teachers' reliefs towards grade retention in a rural elementary school*. Paper 284., Theses, Dissertations and Capstones.
- Troncin, T. (2005). *Le redoublement : radiographie d'une décision à la recherche de sa légitimité*. Ph. D. thesis, Université de Bourgogne : UFR SCIENCES HUMAINES.
- United States Department of Education (1999). Taking responsibility for ending social promotion. Technical report, United States Department of Education. <http://www.ed.gov/PDFDocs/socialprom.pdf>.
- Wynn, J. L. (2010, Mars). *A study of selected teachers perceptions of grade retention*. Phd thesis for the degree of doctor of education, University of South Florida.
- Xia, N. et E. Glennie (2005). Grade retention : the gap between research and practice. Technical report.
- Young, S. et B. Range (2014). Elementary principals' views about struggling students : Insights about perceived interventions including grade retention. *The Researcher* 26(2), 20–34.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57, 348–368.





**Cnesco**

Carré Suffren

31-35 rue de la Fédération

75 015 Paris

[cnesco.communication@education.gouv.fr](mailto:cnesco.communication@education.gouv.fr)

**École normale supérieure de Lyon**  
**Institut français de l'Éducation**

19 allée de Fontenay

69 007 Lyon

[conf.consensus.ife@ens-lyon.fr](mailto:conf.consensus.ife@ens-lyon.fr)