

Peut-on conclure à propos des effets du redoublement ?

Marcel Crahay

Depuis le début du XX^e siècle, les chercheurs en éducation s'efforcent d'appréhender avec rigueur les effets du redoublement. Les études menées à ce sujet ont connu une évolution importante, marquée par des améliorations des dispositifs de recherche et une sophistication de plus en plus grande des techniques d'analyse statistique. L'objectif poursuivi est de cerner au plus près l'effet vrai du redoublement, ce qui implique une réflexion méthodologique sur les sources de biais et sur les méthodes les plus appropriées pour les contrôler. Le présent article retrace cette évolution méthodologique, tout en commentant les principaux résultats des recherches.

Mots-clés : redoublement, échec scolaire, promotion automatique, méta-analyse, recherches quasi expérimentales, biais expérimentaux.

LA QUESTION DE REDOUBLEMENT, L'OPINION PUBLIQUE ET LES ÉTUDES COMPARATIVES DU RENDEMENT SCOLAIRE

L'opinion publique a été récemment secouée par les résultats de l'enquête PISA (2000), une enquête à laquelle 31 pays ont participé. Souvent, la presse a monté en épingle le score national, s'enthousiasmant ou se lamentant des compétences en lecture des élèves selon la position du pays au sein du classement international. Mais les études comparatives du rendement ont heureusement d'autres vertus que de flatter ou de flageller l'orgueil des nations. Elles incitent à s'interroger sur les facteurs qui affectent le rendement des systèmes éducatifs.

Or, en l'occurrence, la surprise est venue pour les enseignants mais aussi pour les parents et certains décideurs politiques des bonnes performances d'ensemble de pays qui, à l'instar de la Finlande, ont *de facto* abandonné la pratique du redoublement.

Le fait même que des pays comme le Danemark, la Finlande (1), la Norvège, la Suède et le Royaume-Uni, mais aussi la Corée, la Malaisie, le Japon, la Nouvelle-Guinée, etc. aient aboli la possibilité de faire redoubler, est déjà objet d'étonnement pour certains. Cette information constitue en soi une réfutation de l'affirmation de certains enseignants qui proclament de façon péremptoire que « *le non-redoublement, c'est impossible* » ou d'auteurs qui écrivent que « *l'école a*

besoin de l'échec scolaire » (Milner, 1984). De façon évidente, certains systèmes éducatifs fonctionnent sans échec et, surtout, sans redoublement. Bien évidemment, il reste à savoir s'ils fonctionnent bien. Car, pour bon nombre d'enseignants, de parents et d'hommes politiques, sous l'idée d'interdire le redoublement se cache le spectre de la baisse de niveau.

On voit assez rapidement le parti que l'on peut tirer des études comparatives du rendement scolaire pour la question qui préoccupe ceux qui redoutent les effets pervers de la suppression du redoublement. S'il est vrai que l'abolition du redoublement entraîne automatiquement la baisse de niveau (2), les pays qui ont fait le choix de la promotion automatique ne peuvent se caractériser que par des rendements médiocres. Pour vérifier cette assertion, il suffit de comparer le rendement des pays qui ont conservé la pratique du redoublement et celui de ceux qui l'ont abandonnée.

Les données PISA constituent un matériel de choix pour procéder à ce genre de comparaison. Chaque pays a testé un échantillon représentatif d'élèves de 15 ans reflétant la diversité du paysage éducatif du pays. Dans chaque pays, un minimum de 150 écoles

sélectionnées selon une procédure aléatoire devaient être impliquées. À l'intérieur des écoles, 35 élèves ont été désignés au hasard sur la liste de tous les élèves de 15 ans. Les mêmes règles de sélection ont été appliquées dans l'ensemble des pays et leur application a été contrôlée de façon étroite par un expert en échantillonnage. Le test comporte des questions à choix multiple et une proportion importante (45 %) de questions à réponse ouverte où l'élève doit élaborer sa propre réponse. Tous les élèves de tous les pays participants passent des épreuves de compréhension de textes identiques ; celles-ci ont été traduites dans les différentes langues au départ de deux versions sources, l'une en anglais et l'autre en français. Les résultats en compréhension de l'écrit sont présentés sur une échelle de compétences à cinq niveaux hiérarchisés. À chaque niveau correspondent des tâches de lecture spécifiques (3).

Dans le tableau I, nous avons concentré la présentation des résultats (exprimés sur une échelle normalisée dont la moyenne internationale vaut 500) sur les pays de l'UE pour lesquels nous savons clairement si le redoublement a ou non été aboli.

Tableau I. – **Moyenne nationale et pourcentages d'élèves à chaque niveau de compétence sur l'échelle PISA (échelle combinée de compréhension de l'écrit)**

	Moyenne	Niveau 1 ou en dessous	Niveau 2	Niveau 3	Niveau 4 ou au-dessus
		%	%	%	%
Pays avec redoublement					
Allemagne	484	22,6	22,3	26,8	28,2
Autriche	507	14,6	21,7	29,9	33,7
Belgique	507	19,0	16,8	25,8	38,3
Espagne	493	16,3	25,7	32,8	25,3
France	505	15,2	22,0	30,6	32,2
Italie	487	18,9	25,6	30,6	24,8
Grèce	474	24,4	25,9	28,1	21,7
Luxembourg	444				
Portugal	470	26,3	25,3	27,5	21,0
Pays avec promotion automatique					
Danemark	497	17,9	22,5	29,5	30,0
Finlande	546	6,9	14,3	28,7	50,1
Irlande	527	10,1	17,9	29,7	41,3
Islande	507	14,5	22,0	30,8	32,7
Norvège	506	17,5	19,5	28,1	34,9
Royaume-Uni	523	12,8	19,6	27,5	40,0
Suède	516	12,6	20,3	30,4	36,8
Moyenne Ocdé	500	17,9	21,7	28,7	31,9

Source : Ocdé (2001).

Les tendances sont nettes. De façon précise, on peut dresser les constats suivants :

– Dans les pays nordiques ainsi qu'en Irlande et au Royaume-Uni, systèmes où la promotion automatique a été adoptée, le pourcentage de bons lecteurs (niveau 4 ou 5) est supérieur à la moyenne Ocdé (hormis au Danemark où il est légèrement inférieur), tandis que le pourcentage de mauvais lecteurs (niveau 1 ou en dessous) est inférieur à la moyenne Ocdé (à nouveau, hormis le Danemark où ce pourcentage est égal à la moyenne Ocdé).

– Parmi les pays qui ont conservé la pratique du redoublement, on observe des profils de résultats différents. Le pourcentage d'élèves mauvais lecteurs est supérieur à la moyenne Ocdé en Allemagne, en Belgique, en Italie, en Grèce, au Luxembourg et au Portugal, mais ce n'est pas le cas en Autriche, en Espagne et en France. Quant au pourcentage de bons lecteurs, il est supérieur à la moyenne Ocdé en Autriche, en Belgique et en France, mais ce n'est pas le cas dans les autres pays.

Le résultat de cette comparaison n'étonnera pas les chercheurs coutumiers de ce genre d'études. Les données PISA viennent simplement confirmer avec un surplus de netteté des tendances que dessinaient déjà les études de l'IEA menées en 1970 et en 1990 (voir pour une synthèse assez détaillée Crahay, 2003). Bref, de façon répétée, les comparaisons internationales du rendement en lecture (mais aussi en mathématiques et en sciences) fournissent des données empiriques qui s'opposent à l'affirmation selon laquelle la promotion automatique s'accompagne d'un nivellement par le bas. Les pays qui ont adopté cette façon de gérer le parcours des élèves se caractérisent par des niveaux satisfaisants de performance en lecture (mais aussi en mathématiques et en sciences).

ÉTUDES QUASI EXPÉRIMENTALES DES EFFETS DU REDOUBLEMENT ET LE PROBLÈME DES BIAIS D'ÉCHANTILLONNAGE

Comparaison n'est pas raison, dit avec justesse l'adage populaire. Il se pourrait donc bien que les études comparatives du rendement scolaire biaisent la réflexion en ce qui concerne le redoublement. Plus précisément, on peut légitimement suspecter que la promotion automatique se superpose, dans les pays qui ont adopté cette modalité organisationnelle, à d'autres qualités et que ce soient ces qualités qui

déterminent effectivement le bon rendement scolaire observé dans ces pays. Il est donc sage d'examiner un autre corpus de résultats de recherche.

Indéniablement, la question des effets du redoublement interpelle le monde de l'éducation depuis fort longtemps. Dans sa célèbre revue de question datant de 1975, Jackson fait état d'une recherche menée par Keyes en 1911 (citée par Jackson, 1975). Bref, il semble bien que l'on s'intéresse aux effets du redoublement depuis près d'un siècle et ceci sans que la problématique soit considérée comme définitivement résolue. Or, sans pouvoir établir une comptabilité précise, il est manifeste que le nombre d'études sur le sujet est impressionnant : plus d'une centaine, estimons-nous. Plus surprenant encore, les résultats engrangés au fil de ces multiples efforts de recherche sont substantiellement cohérents. Rien n'y fait : l'incrédulité du monde de la pratique, mais aussi celle de certains décideurs politiques, semble bien irréductible. Pourtant, afin de répondre aux objections qui leur ont été et leur restent opposées, les chercheurs se sont efforcés d'affiner leur méthodologie de recherche. En définitive, on peut considérer les recherches sur les effets du redoublement comme paradigmatiques en ce qui concerne les obstacles méthodologiques que tentent de contourner les chercheurs en éducation pour approcher une saisie fiable des effets d'une pratique d'enseignement. Il est dès lors légitime de retracer l'évolution des recherches dans ce domaine.

Les premières recherches sur les effets du redoublement (telles qu'a pu les recenser Jackson, 1975) ont été menées selon le schéma avant-après ou pré-test – post-test. Il s'agit simplement de mesurer l'état des connaissances et performances d'un échantillon d'élèves en difficulté scolaire avant et après l'année redoublée. Bien évidemment, les instruments de mesure (les tests) doivent être les mêmes avant et après le redoublement. Très vite, les chercheurs se sont rendu compte qu'il n'était pas possible de conclure avec pareil dispositif de recueil de données. Comme le discute avec profondeur Jackson (1975), il est hautement probable d'observer des gains importants entre le pré et le post-test, mais il est impossible de déterminer avec certitude le facteur auquel attribuer ces gains : au redoublement, à la maturation des élèves, à de quelconques changements pédagogiques et/ou psychologiques liés au fait de répéter une année... C'est donc sans surprise que l'auteur américain note que les douze études menées selon ce schéma concluent à la nette évolution cognitive des élèves qui ont redoublé, et ceci quel que soit l'instrument de mesure utilisé.

Logiquement, les chercheurs se sont tournés vers la méthodologie expérimentale dont le principe de base consiste à comparer l'évolution de deux groupes d'élèves, les uns ayant redoublé et les autres pas. Évidemment, la comparaison n'a de sens que si l'on peut garantir l'équivalence initiale des deux groupes d'élèves. Et là réside une difficulté majeure. Certes, il va de soi que les recherches sur le redoublement portent uniquement sur des élèves qui éprouvent des difficultés scolaires, mais comment établir avec certitude que deux groupes d'élèves sont réellement affectés des mêmes caractéristiques (en l'occurrence, des mêmes problèmes). La rigueur expérimentale voudrait qu'à partir d'un échantillon d'élèves chez qui l'on a repéré des difficultés d'apprentissage, on procède à l'affectation aléatoire des individus dans les groupes : autrement dit, l'assignation des élèves au groupe des redoublants ou à celui des promus serait réglée par tirage au sort. Si cette technique est acceptable dans le cadre de nombreuses recherches menées en médecine ou en psychologie, il n'en va pas de même dans le cas du redoublement. Par déontologie, la grande majorité des chercheurs se sont interdit cette procédure que recommandent tous les manuels de méthodologie. Ils se sont placés dans des conditions moins confortables sur le plan de la rigueur. Tirant parti de la variabilité des exigences des enseignants, ils s'attachent à repérer dans un échantillon de classes des élèves en difficulté scolaire pour lesquels la décision de promotion diffère. Ceci est tout à fait possible, en effet : les recherches de Grisay (1984) notamment ont montré qu'avec un même niveau de compétence, des élèves d'un même degré scolaire pouvaient être promus ou, au contraire, être amenés à répéter l'année. Le principe sous-jacent aux études quasi expérimentales consiste à suivre l'évolution sur une ou plusieurs années d'enfants qui ont fait preuve de compétences scolaires équivalentes à un test, mais pour lesquels la décision de promotion est différente.

Considérons à titre d'illustration l'étude menée dans les *MESA Public Schools* par Peterson, Degracie & Ayabe (1987). En 1981, ces trois chercheurs se

sont attachés à repérer, dans les écoles du district concerné, les élèves pour qui une décision de redoublement avait été prise. Parmi les 171 élèves visés par cette mesure, ils ont extrait un échantillon de 106 élèves doublants, dont 65 devaient répéter la première année, 26 la seconde et 15 la troisième. Dans un second temps, les trois chercheurs ont constitué un groupe de contrôle par appariement. L'appariement consiste à assortir des sujets par constitution de paires ; les sujets constituant une paire doivent présenter un maximum de points communs. Dans le cas de l'étude de Peterson *et al.*, les élèves appariés ont même âge, même sexe et même groupe ethnique. L'appariement était encore établi sur la base des résultats obtenus en 1981 par les élèves au Californian Achievement Test (CAT), et ceci en tenant compte du score total ainsi que des scores en lecture, en langage et en mathématique. Bref, tout est fait pour rendre les deux groupes de sujets aussi comparables que possible. Peterson, Degracie & Ayabe (1987) ont ensuite suivi ces deux groupes d'élèves pendant trois années. Plus exactement, ils leur ont présenté le CAT fin 1982, 1983 et 1984. Le plan de recherche peut être schématisé comme dans le Tableau II (4).

Avec pareil dispositif, on est proche d'une situation expérimentale, sans y être tout à fait. Il est de coutume de désigner les recherches menées de la sorte par l'expression d'études quasi expérimentales dans la mesure où on dispose de deux groupes supposés équivalents. Les sujets d'un des groupes, que l'on qualifiera d'*expérimental*, bénéficient d'un traitement : le redoublement. L'état initial des sujets des deux groupes (prétest) ainsi que leur état après traitement (post-test) sont mesurés à l'aide du même instrument ou d'instruments appariés. Notons, toutefois, que si l'on se réfère à l'ouvrage de référence de Cook & Campbell (1979), le qualificatif *quasi expérimental* est usurpé. La quasi expérimentation implique, en effet, que les chercheurs aient la maîtrise du traitement, celui-ci étant appliqué à des groupes naturellement constitués. L'expérimentation pure consistant en un dispositif où les chercheurs contrôlent à la fois

Tableau II. – Schématisation du plan de recherche adopté dans les études sur les effets du redoublement

	Fin 81	Année scolaire 81-82	Fin 82	Fin 83	Fin 84
Groupe « <i>expérimental</i> »	Prétest	Redoublement	Post-test 1	Post-test 2	Post-test 3
Groupe de contrôle	Prétest	Promotion	Post-test 1	Post-test 2	Post-test 3

le traitement, la sélection des sujets et leur affectation (qui est aléatoire) aux différents groupes. Ici, le chercheur n'a ni le contrôle du traitement (le redoublement), ni celui du choix des individus qui en bénéficient. Il est seulement responsable de la composition du groupe de contrôle. Pour le type de travaux d'études que nous discutons ici, Cook & Campbell utilisent l'expression *passive-observational studies* (p. 298). On pourrait aussi utiliser les expressions *observation invoquée* ou mieux *recherche rétrospective avec schéma surimposé* (de Landsheere, 1979, p. 258).

Le dispositif de recherche utilisé dans les études sur les effets du redoublement – que nous continuerons à appeler quasi expérimental comme il est de tradition dans le domaine – n'est donc pas exempt de biais éventuels qui pourraient affecter la validité des conclusions. Comme le signale Jackson (1975), on peut toujours craindre avec pareil plan de recherche que des différences latentes entre élèves redoublants et élèves promus n'aient pas été correctement mises en évidence par le ou les test(s) de départ et, partant, que la comparaison soit biaisée en faveur des élèves promus ou en faveur des doublants.

Ce risque est minimisé dans une procédure expérimentale pure. En une première étape, un échantillon de sujets-cibles est sélectionné soit par recrutement aléatoire lorsqu'on veut faire porter l'expérience sur des sujets tout-venant, soit à l'aide d'une mesure initiale lorsqu'on s'intéresse à un groupe spécifique d'individus. Avec le redoublement, on se trouve dans le second cas : il convient de rassembler un échantillon d'élèves en difficulté scolaire. En une seconde étape, les sujets sont assignés aléatoirement au groupe expérimental ou au groupe de contrôle. Au terme d'une pareille procédure, l'équivalence des deux groupes peut être considérée comme assurée.

Le problème est nettement plus délicat dans une étude de type quasi expérimental (ou plus justement

dans une recherche rétrospective avec schéma surimposé). Le groupe expérimental, qui en l'occurrence va bénéficier du redoublement, est constitué naturellement : ce sont les enseignants qui repèrent les élèves pour lesquels ils jugent la mesure opportune. Le chercheur intervient seulement après pour composer un groupe de contrôle dont il va essayer d'ajuster au mieux les caractéristiques à celles du groupe expérimental. Les tests que passent les élèves avant l'intervention (pour permettre d'établir leur niveau de départ) et ceux qu'ils passent pendant ou après celle-ci (pour mesurer les changements ou les gains induits) sont un des éléments cruciaux du dispositif. L'erreur de mesure qu'ils comportent inévitablement peut faire paraître équivalents au départ des groupes dont les capacités cognitives initiales sont en réalité différentes. Par voie de conséquence, on peut croire, lors du post-test, à des différences de gain alors qu'en fait, il y a un biais de sélection des sujets, biais occulté par l'erreur de mesure du prétest.

La question qui se pose aux chercheurs soucieux de mesurer les effets du redoublement revient à distinguer l'effet vrai de cette mesure de l'effet apparent. Cook & Campbell (1979, pp. 295-301) ont montré que l'erreur d'interprétation des gains prend la même direction que la différence suspectée au départ. Plus précisément, si l'intervention menée au bénéfice des individus du groupe expérimental a un effet positif, celui-ci sera sous-estimé si, en raison de l'erreur de mesure du prétest, on ne s'est pas aperçu de ce que ces sujets sont plus « faibles » que ceux du groupe de contrôle. Il sera surestimé dans le cas opposé, c'est-à-dire si l'on peut suspecter que les élèves doublants sont plus « forts » que les élèves promus. Les risques de sous ou de surestimation vont en sens inverse dans le cas d'un effet négatif. On risque une surestimation dans le cas où il y a des raisons de craindre une erreur de mesure qui occulte une différence initiale en faveur des élèves promus et une sous-estimation dans le cas opposé. L'apport de Cook & Campbell (1979) est résumé dans le tableau qui suit :

Tableau III. – **Risques de sur ou de sous-estimation des effets du redoublement en fonction de l'erreur de mesure redoutée**

On suspecte que le redoublement		est positif	est négatif
On peut craindre que l'erreur de mesure dissimule une différence	en faveur des élèves doublants (groupe expérimental)	Risque de surestimation de l'effet	Risque de sous-estimation de l'effet
	en faveur des élèves promus (groupe de contrôle)	Risque de sous-estimation de l'effet	Risque de surestimation de l'effet

Dans son examen des études quasi expérimentales portant sur les effets du redoublement, Jackson (1975) s'est efforcé de tenir compte du risque de biais d'échantillonnage. Dans un premier temps, il rend compte des résultats de toutes les recherches sans se soucier d'une erreur de mesure éventuelle. Il classe les résultats de ces études en distinguant cinq possibilités : a) Différence significative en faveur des élèves promus, b) Différence non significative en faveur des élèves promus, c) Pas de différence, d) Différence non significative en faveur des élèves doublants, et e) Différence significative en faveur des élèves doublants.

Dans le tableau IV, les chiffres indiquent le nombre de comparaisons relevant de l'une ou l'autre de ces catégories de résultats. Deux critères sont utilisés : résultats à des tests de connaissance et réponse à des échelles d'ajustement social.

Dans ce tableau de résultats, les tendances sont nettes aussi bien en ce qui concerne les connaissances scolaires que l'adaptation socio-affective. Dans une majorité d'analyses portant sur le premier critère, on observe de meilleurs résultats chez les élèves faibles bénéficiant d'une promotion que chez les élèves faibles qui ont redoublé (24 + 45 = 69). La tendance inverse est observée dans 31 analyses. Lorsqu'on se limite aux analyses qui ont abouti à des différences significatives sur le plan statistique, le bilan est plus convaincant encore : 24 recherches concluent contre le redoublement pour 2 en sa faveur. Dans 67 des 104 analyses portant sur le second critère, la tendance est favorable aux élèves promus. Elle est favorable aux doublants dans 37 cas sur 104. Le rapport des résultats significatifs sur le plan statistique est semblable à celui observé pour le premier critère : 27 pour les promus contre 3 pour les doublants.

Tableau IV. – Résultats des études dites quasi expérimentales comparant des élèves présentant des difficultés dont certains doublent leur année tandis que d'autres sont promus

Auteurs	Année de publication	Connaissances scolaires					Adaptation socio-affective				
		a	b	c	d	e	a	b	c	d	e
Mac Elwee	1932	-	-	-	-	-	3	9	0	2	0
Cheyney & Boyer	1933	0	1	0	0	0	-	-	-	-	-
Arthur	1936	0	4	1	2	0	-	-	-	-	-
Anfinson	1941	1	0	2	0	0	-	-	-	-	-
Sandin	1944	-	-	-	-	-	17	3	0	0	0
Goodlab	1954	-	-	-	-	-	1	4	0	0	0
Wallihan	1955	1	0	0	0	0	-	-	-	-	-
Coeffield & Bloomers	1956	3	33	1	22	1	-	-	-	-	-
Anderson	1957	0	0	0	2	0	0	5	0	8	0
Worth	1960	2	3	0	2	1	0	2	0	7	2
Kowitz & Armstrong	1961	1	1	0	0	0	-	-	-	-	-
Kamii & Weikart	1963	4	0	0	0	0	-	-	-	-	-
Chasky	1964	2	2	0	0	0	0	3	0	8	1
Briggs	1966	6	0	0	0	0	2	6	0	6	0
Dobbs & Neville	1967	4	0	0	0	0	-	-	-	-	-
Reinherz & Griffin	1971	0	1	0	1	0	-	-	-	-	-
White & Howard	1973	-	-	-	-	-	4	5	0	0	0
Total		24	45	4	29	2	27	40	0	34	3

Quelle est la part des erreurs de mesure dans ces résultats ? Pour tenter de répondre à cette question cruciale, Jackson (1975) examine dès lors attentivement les propriétés méthodologiques de chacune des études mentionnées dans le tableau précédent et les subdivise en deux catégories selon que l'on peut craindre un biais en faveur des redoublants ou, à l'opposé, un biais en faveur des promus. Comme le montre le tableau V, les résultats observés eu égard au premier critère vont dans le sens du biais redouté.

Soixante analyses sont passibles d'un biais en faveur des élèves promus, mais quarante-quatre sont menacées par le biais inverse. Au total, on peut penser disposer d'un éventail d'analyses relativement équilibré quant aux biais qui auraient pu tronquer la conclusion dans un sens ou dans un autre. Dans le cas présent, le bilan est légèrement déséquilibré en faveur de la promotion et, par voie de conséquence, le caractère préjudiciable du redoublement est probablement quelque peu surestimé. Quoi qu'il en soit, il semble bien qu'on puisse exclure l'idée que la tendance inverse eut pu être observée.

Cette conclusion paraît d'autant plus légitime qu'elle est corroborée par les résultats des trois seules études expérimentales jamais menées sur les effets du redoublement. En effet, à l'inverse de la majorité des chercheurs qui se sont refusés de recourir à l'affectation aléatoire des sujets, Klene & Branson (1929), Farley (1936) et Cook (1941) ont procédé à des études qui se rapprochent très fortement des règles de la pure expérimentation. Pour l'essentiel, ces trois études font état d'absence de différences significatives. Les analyses de Klene & Branson (1929) portent sur 141 élèves répartis entre la seconde et la sixième année de l'enseignement obligatoire. Ils concluent en faveur des élèves promus, mais ne mentionnent pas de différences significatives. La recherche de Farley (1936) porte sur 400 élèves répartis entre la seconde et la cinquième année de l'enseignement obligatoire. Aucune différence signifi-

cative n'apparaît en ce qui concerne les élèves de deuxième et troisième années. En revanche, en ce qui concerne ceux de quatrième et de cinquième années, on observe pour les tests relatifs à la compréhension en lecture une différence significative en faveur des élèves promus. Pour les mêmes élèves, on n'observe pas de différence significative au niveau des tests de mathématique. L'étude de Cook (1941) porte sur 700 élèves distribués dans les sept premières années de la scolarité obligatoire. L'auteur n'observe pas de différence significative entre les deux groupes.

LE CONTRÔLE STATISTIQUE DES BIAIS D'ÉCHANTILLONNAGE

En France aussi, les effets du redoublement ont été objet de recherches. La première à notre connaissance est l'œuvre de Seibel (1984). Profitant d'une évaluation pédagogique mise en œuvre à la demande de la direction des écoles sur un échantillon représentatif de classes et d'élèves du CP (cours préparatoire ou 1^{re} année primaire), ce chercheur a analysé l'évolution des scores de 1 100 élèves aux mêmes épreuves standardisées de connaissances en français et en mathématiques entre juin et décembre de l'année 83. Au sein de cet échantillon, il est possible d'identifier un groupe de redoublants et, complémentairement, de constituer un groupe d'élèves de faiblesse équivalente et qui ne redoublent pas. Dans ce cas précis, on se trouve clairement dans la situation que de Landsheere (1979) désignerait comme une *étude rétrospective avec schéma surimposé*.

De l'examen des données, deux constats importants eu égard à notre propos méritent d'être épinglés :

1. En juin, les résultats moyens des « non-redoublants faibles » et des « redoublants » diffèrent peu, tant au test de français (différence 2,9) qu'au test de mathématiques (3,3).

2. En décembre, les écarts entre ces deux groupes se sont creusés. On observe une différence de 16,3 points au test de français et de 11,7 au test de mathématiques en faveur des élèves faibles qui n'ont pas redoublé.

Le premier de ces constats, c'est-à-dire celui portant sur les résultats à l'épreuve de juin (qui, ici, tient lieu de prétest) peut faire craindre un biais d'échantillonnage en faveur des non-redoublants faibles dans la comparaison avec les redoublants. Ceci peut légitimer des doutes vis-à-vis des résultats de cette étude.

Tableau V. – **Distribution des résultats selon la nature du biais redouté (Adapté de Jackson, 1975)**

Nature du biais redouté	Connaissances scolaires				
	a	b	c	d	e
Biais en faveur des élèves promus	22	28	3	6	1
Biais en faveur des élèves doublants	2	17	1	23	1

Le type de traitement appliqué par Grisay (1993) aux données qu'elle a recueillies, à la demande de la Direction de l'Évaluation et de la Prospective (DEP) du Ministère français de l'Éducation, dans le cadre d'une étude sur les effets et le fonctionnement des collèges, permet de circonscrire les risques de sur ou de sous-estimation des effets du redoublement. Plus précisément, cette chercheuse a appliqué deux types de traitement aux données : le premier, l'analyse de régression, fournit une estimation de l'effet du redoublement entachée d'erreurs de mesure ; le second, mis au point par Heuchenne (1993), permet d'estimer l'effet vrai du traitement. Pour l'essentiel, il s'agit d'une analyse LISREL (5).

L'échantillon est constitué des élèves entrés en sixième (1^{re} année du secondaire) à la rentrée 1990-91 dans 100 collèges de la France métropolitaine. Ces élèves ont passé en septembre 1990, à l'entrée en sixième, les épreuves standardisées de français et de mathématiques. Ils ont été testés à nouveau deux ans plus tard (mai 1992) alors que la majorité d'entre eux se trouvaient en classe de cinquième (2^e année du secondaire). Une partie des élèves s'est cependant vu refuser l'entrée en cinquième. Ces redoublants qui se trouvent encore en fin de sixième au moment du post-test, constituent le groupe-cible de l'analyse opérée par Grisay (1993). Une matrice complète de données est disponible pour 5 137 élèves (4 873 non-redoublants et 264 doublants).

La batterie du prétest comporte huit épreuves (quatre épreuves pour le français, quatre pour les mathématiques), totalisant 213 items. La fidélité globale du prétest est satisfaisante (0,79). Le post-test comporte quatre épreuves. Deux d'entre elles constituent un ancrage, c'est-à-dire qu'elles sont composées d'items déjà présentés lors du prétest. Les deux autres épreuves sont nouvelles. Au total, le post-test repose sur 91 items. Sa fidélité est bonne (0,85).

La corrélation entre la variable *redoublement* et le critère (les résultats aux post-tests) est négative (-0,23). Il en va de même pour le coefficient de régression de cette même variable sur le critère, obtenu sous contrôle des variables de prétest (-0,207). Ceci signifie qu'à résultat égal aux épreuves de français et de mathématique à leur entrée en 6^e, les élèves redoublants obtiennent deux ans plus tard un score au post-test inférieur de plus de deux dixièmes d'écart-type par comparaison avec leurs condisciples non-redoublants. L'application du modèle LISREL aboutit également à un coefficient négatif, mais de moindre amplitude (-0,11). Autrement dit, les redoublants ne sont lésés qu'à concurren-

ce d'un dixième d'écart-type. Si cette valeur correspond à la moitié de celle obtenue par équation de régression, elle reste suffisante pour réfuter l'hypothèse selon laquelle le redoublement agit comme une remédiation permettant aux élèves de récupérer leurs lacunes et de repartir dans le cursus scolaire avec de meilleurs atouts.

L'étude de Grisay (1993, voir aussi 2001) permet de circonscrire l'état actuel du débat scientifique. L'effet négatif du redoublement est établi en ce sens que les élèves faibles qui en sont l'objet progressent moins que des élèves qui, ayant des difficultés comparables, ont été promus. Cet effet négatif, il reste à en estimer exactement l'ampleur. L'application de la technique de la méta-analyse aux données recueillies sur les effets du redoublement est susceptible d'apporter des éléments de réponse à cette question.

MÉTA-ANALYSE ET ESTIMATION DE L'AMPLITUDE DE L'EFFET

La technique de la méta-analyse développée par Glass (1977) et Glass *et al.* (1982) est devenue l'outil indispensable dès lors que les chercheurs s'attachent à établir le bilan des recherches menées sur une question spécifique. Cette technique permet de combiner les résultats des différentes recherches et, partant, leur confère une validité plus grande. Souvent, en effet, on peut douter de la conclusion d'une recherche particulière parce que les analyses ont été réalisées sur de petits nombres. De surcroît, on peut également redouter que cette recherche particulière soit biaisée par une erreur de mesure au point de départ qui, *in fine*, favorise l'un des deux groupes. Avec la méta-analyse, on peut miser d'une part sur un effet d'agrégation des données qui résout le premier problème et, d'autre part, sur un effet de compensation des biais : les recherches grevées d'un biais en faveur des élèves doublants étant compensées par celles menacées par un biais en faveur des élèves promus. Ces bénéfices sont d'autant plus assurés que l'on est en mesure de combiner de nombreuses recherches. Il va de soi, par ailleurs, que les avantages soulignés ci-dessus ne peuvent corriger les défauts graves que comporterait l'une ou l'autre recherche. La qualité d'une méta-analyse dépend donc de la qualité des recherches disponibles, mais aussi de la sévérité des critères méthodologiques adoptés par l'auteur dans la sélection des recherches.

La méta-analyse de Holmes et Matthews (1984) et plus encore celle de Holmes (1990) constitue des

modèles du genre. Ainsi, en 1984, Holmes et Matthews ont examiné 650 recherches portant sur les effets du redoublement. Quarante-quatre ont été retenues car elles présentaient suffisamment de garanties quant à la rigueur méthodologique de l'investigation et, notamment, en ce qui concerne l'équivalence initiale du groupe expérimental et du groupe de contrôle. C'est dire si les chercheurs se sont montrés sévères quant aux critères méthodologiques retenus pour conserver une recherche dans l'échantillon devant donner lieu à méta-analyse. En 1990, Holmes ne se départit pas de sa sévérité méthodologique. En définitive, il conserve 63 recherches, dont 44 faisaient déjà partie de sa précédente publication avec Matthews.

Dans la méta-analyse de 1984, toutes les amplitudes de l'effet calculées (6) sont négatives. Dans celles de 1990, tous les effets sont négatifs, à une exception près. On relève, en effet, un effet positif sur l'image de soi, mais d'amplitude négligeable : + 0,06. Le tableau VI synthétise les résultats obtenus dans le cadre de ces deux méta-analyses.

Les effets négatifs du redoublement sur les gains de connaissance des élèves semblent difficilement discutables. L'amplitude des effets calculés en pre-

nant pour critère l'une ou l'autre mesure de performance varie de - 0,25 (en mathématique) à - 0,37 (en science) lorsque ceux-ci sont mesurés au moyen d'épreuves externes ; elle est de - 0,78 lorsqu'on utilise les épreuves conçues par les enseignants eux-mêmes. Les effets « affectifs » du redoublement sont, en général, négatifs (exception faite de l'image de soi), mais de moindre ampleur : les valeurs varient entre - 0,12 pour le bien-être émotionnel et - 0,23 pour les attitudes comportementales. On reviendra, dans la conclusion, sur cette difficile question des effets affectifs du redoublement ; rappelons dès à présent qu'il s'agit d'une composante psychologique plus difficilement appréhendable par des tests que les acquis cognitifs.

Dans la méta-analyse de 1990, Holmes s'est attaché à voir dans quelle mesure l'effet du redoublement sur les acquis cognitifs dépendait du moment où il survenait. Le tableau VII indique les résultats qu'il a pu obtenir.

Il faut noter à nouveau que toutes les valeurs sont négatives. Contrairement à une opinion largement répandue dans le corps enseignant, le redoublement précoce (avant d'entrer à l'école primaire ou en pre-

Tableau VI. – **Ampleur de l'effet du redoublement (en fractions d'écart type) sur une série de variables-critères (adapté de Holmes & Matthews, 1984 et de Holmes, 1990)**

Critères mesurés	Holmes & Matthews (1984)		Holmes (1990)	
	Nombre d'études	Amplitude de l'effet	Nombre d'études	Amplitude de l'effet
Performances académiques	31	- 0,44	47	- 0,31
Performances en langue maternelle	-		18	- 0,33
Performances en lecture	24	- 0,48	34	- 0,30
Performances en mathématique	20	- 0,33	31	- 0,25
Performances en activités scientifiques	-		3	- 0,37
Résultats aux épreuves de l'enseignant	-		3	- 0,78
Méthodes de travail	1	- 0,41	-	
Réactions affectives générales	-		27	- 0,21
Développement social	13	- 0,27	27	- 0,21
Bien-être émotionnel	5	- 0,37	10	- 0,12
Attitudes comportementales en classe	7	- 0,31	10	- 0,23
Image de soi	9	- 0,19	11	+ 0,06
Attitudes vis-à-vis de l'école	8	- 0,16	10	- 0,18
Fréquentation scolaire	-		5	- 0,22

Tableau VII. – **Ampleur de l'effet du redoublement (en fractions d'écart-type) mesurée par des tests de connaissance en fonction du niveau d'étude auquel survient le redoublement (adapté de Holmes, 1990)**

	En maternelle	P 1	P 2	P 3	P 4	Après la 4 ^e
Nombre d'études	8	12	4	7	6	5
Amplitude de l'effet	- 0,28	- 0,28	- 0,10	- 0,15	- 0,36	- 0,38

mière année) n'a pas d'effet préventif ; les amplitudes de l'effet, mesurées respectivement sur huit études pour le niveau maternel et sur 12 pour la première année, sont de - 0,28.

Mais l'intérêt méthodologique premier de la méta-analyse de 1990 tient à la façon dont Holmes aborde une difficulté majeure des études sur les effets du redoublement : pendant que les redoublants doivent recommencer des apprentissages déjà abordés l'année précédente, les promus sont soumis à de nouveaux apprentissages. Un an plus tard, au moment où les chercheurs prélèvent des mesures pour établir la comparaison, les uns ont une année scolaire d'avance par rapport aux autres, bien que les uns et les autres aient le même âge. Disposant d'épreuves standardisées et étalonnées sur plusieurs années scolaires, il est légitime d'apprécier les progrès des uns et des autres. C'est cette approche que privilégient la plupart des études reprises dans les méta-analyses de Holmes & Matthews (1984) et de Holmes (1990). Il semble pourtant opportun de s'interroger sur ce qu'il en est quand on compare des élèves de faiblesse équivalente, mais dont certains ont répété une année afin de consolider les apprentissages prérequis à ceux prévus dans les années ultérieures. En vérité, le principe du redoublement consiste à sacrifier une année pour permettre à l'enfant de repartir sur de meilleures bases et atteindre – avec une année de retard, certes – des niveaux de compétences auxquels il n'aurait pu prétendre s'il n'avait pas redoublé. Ce raisonnement appelle un autre type de comparaison où l'on met en relation les niveaux de connaissance atteints par des élèves faibles qui ont redoublé et ceux atteints par des élèves faibles qui n'ont jamais redoublé, et ce à un même niveau scolaire (par exemple, à la fin d'un cycle). Dans quelques-unes des études recensées par Holmes en 1990, les élèves doublants et leurs condisciples du groupe de contrôle ont été suivis pendant plusieurs années. Ceci permet de se faire une idée de l'effet à moyen terme du redoublement.

Plus précisément, Holmes (1990) procède à deux types de comparaison.

– Dans un cas, on compare les élèves qui ont redoublé avec des élèves faibles non-redoublants au même âge. Par voie de conséquence, les élèves non-redoublants ont bénéficié des enseignements d'une année supplémentaire. On parlera de **comparaison à âge constant**.

– Dans un second cas, on compare les élèves doublants et non-redoublants au terme de la même année scolaire, en sachant bien que les premiers ont un an de plus par rapport aux seconds. On parlera de **comparaison à degré scolaire constant**.

On peut rendre plus clair le principe de ces deux comparaisons en imaginant une situation concrète. Supposons un groupe d'élèves qui entrent en première année (1P) en 1989. En fin d'année, une partie d'entre eux doublent et une autre partie est promue. En 1990, certains restent en 1P et d'autres sont en 2P. Imaginons que, pour ces deux cohortes d'élèves, la suite du parcours se réalise sans embûche ; cela conduira à la situation suivante :

	1989	1990	1991	1992	1993	1994
Promus	1 P	2 P	3 P	4 P	5 P	6 P
Doublants	1 P	1 P	2 P	3 P	4 P	5 P

Si les élèves des deux groupes se soumettent à des tests de connaissance en fin de chaque année, il sera possible de réaliser les deux comparaisons expliquées ci-dessus :

– En 1991, on compare les performances des promus alors qu'ils sont en 3P avec celles des doublants alors qu'ils sont en 2P. On procède à une comparaison à âge constant.

– On compare les performances des promus obtenues en 1990 en fin de 2P avec celles obtenues par les doublants en 1991 à la fin du même degré. On procède à une comparaison à degré scolaire constant.

Notons que, dans les deux cas, les doublants sont avantagés. Prenons pour la facilité de l'explication le cas des élèves qui ont redoublé en première année :

	1989		1990		1991		1992		1993		1994
Promus	1P	test	2P	test	3P	test	4P	test	5P	test	6P
Doublants	1P	test	1P	test	2P	test	3P	test	4P	test	5P

- Lorsqu'on compare les performances des élèves à âge constant, les scores des redoublants corres-

pondent aux épreuves de 1P lors de la seconde passation (ou lors de leur seconde première année) ; ceux

	1989		1990		1991		1992		1993		1994
Promus	1P	test	2P	test	3P	test	4P	test	5P	test	6P
Doublants	1P	test	1P	test	2P	test	3P	test	4P	test	5P

des promus aux épreuves de 2P (7). Bref, ces derniers sont soumis à un test plus difficile.

- Lorsqu'on compare à degré scolaire constant, ce biais n'existe plus. En revanche, dans cette condition expérimentale, les doublants ont un an de plus, donc un bénéfice en termes de maturation.

Les résultats de ces deux comparaisons sont présentés dans les tableaux VIII et IX.

Ce double constat est éclairant. Lorsqu'on compare le niveau cognitif atteint par les doublants et les non-redoublants à même degré scolaire, les performances des uns et des autres sont équivalentes, et cela reste vrai que les comparaisons soient établies un an après le redoublement ou plus de trois ans après. Ceci signifie donc que, tout en ayant sacrifié une année afin de repartir sur de meilleures bases pour aborder la suite de la scolarité, les élèves redou-

blants ne se montrent pas meilleurs que leurs condisciples qui n'ont pas fait ce sacrifice. C'est, de notre point de vue, cette absence d'effet lors des comparaisons à degré scolaire qui démontre le plus la vanité du redoublement dans sa prétention à aider les élèves en difficulté. Si, de surcroît, on prend en considération les résultats des comparaisons à âge constant (qui mettent en évidence que le retard des élèves redoublants va en s'accroissant avec les années par rapport à ceux qui ont été promus), il est légitime de conclure qu'une année redoublée est bien une année perdue.

EN DÉFINITIVE, QUE PEUT-ON CONCLURE ?

Sauf à nier l'étendue des preuves et l'opiniâtreté des chercheurs à trouver la façon la plus valide d'ap-

Tableau VIII. – Effets à moyen et à long terme du redoublement

(mesuré en fraction d'écart-type).

Résultats des comparaisons à âge constant (adapté de Holmes, 1990)

	1 an après	2 ans après	3 ans après	Plus de 3 ans après
Nombre d'études	28	5	3	3
Amplitude de l'effet	- 0,41	- 0,64	- 0,74	- 0,88

Tableau IX. – Effets à moyen et à long terme du redoublement

(mesuré en fraction d'écart-type).

Résultats des comparaisons à degré scolaire constant (adapté de Holmes, 1990)

	1 an après	2 ans après	3 ans après	Plus de 3 ans après
Nombre d'études	10	7	5	6
Amplitude de l'effet	0	+ 0,02	- 0,12	+ 0,04

préhender les effets du redoublement, l'homme de raison doit admettre que les données de recherche ne plaident pas pour le maintien de cette pratique. Il semble établi désormais que le fait de répéter une année et, partant, de recommencer la totalité d'un programme de cours n'aide pas les élèves en difficulté à surmonter les obstacles qui les empêchent de réussir honorablement à l'école. Étant donné que les recherches dites quasi expérimentales ne portent guère sur le secondaire, on peut comprendre que certains rechignent à étendre aux adolescents ce qui a été observé pour les enfants du primaire.

Du point de vue du chercheur, il semble urgent de dépasser la polémique à propos des effets du redoublement au primaire pour privilégier d'autres questionnements. Car si le redoublement ne constitue pas un moyen pour venir en aide aux élèves en difficulté, il paraît opportun de chercher d'autres moyens pour résoudre cet important problème. Bref, plutôt que de demander de nouvelles preuves concernant les effets du redoublement, il semblerait plus fécond de

demander aux chercheurs de se concentrer sur d'autres objets d'investigation. Car, vu la qualité des efforts mis en œuvre pour conjurer les biais d'échantillonnage et de mesure dans les études sur le redoublement, il semble difficile d'améliorer encore la validité des démonstrations et fort improbable de renverser la tendance des conclusions. Seule la question des effets socio-affectifs du redoublement pourrait encore valoir quelques efforts de recherche. À ce sujet, des travaux de type qualitatif semble apporter un *plus* (Crahay, 2003) qui, de toute façon, se solde par un supplément d'éléments à charge du redoublement.

Reste à savoir si, en matière d'éducation, les gens de terrain et les décideurs politiques sont prêts à se laisser convaincre par un faisceau de recherches qui, tout en améliorant significativement leur contrôle des biais de mesure, débouchent sur des résultats convergents.

Marcel Crahay
Universités de Liège et de Genève

NOTES

- (1) Le redoublement n'est pas légalement interdit en Finlande, mais il n'est pas pratiqué. Dans les pages qui suivent, nous considérons ce pays comme un système à promotion automatique.
- (2) Cette crainte est présente chez les enseignants de la CFBW. Ainsi, dans l'enquête menée par Stegen (1994) auprès de 263 enseignants dans cette communauté, soixante-deux pour cent d'entre eux estiment que la suppression du redoublement en début de secondaire, préconisée par le Ministre de l'Éducation de l'époque, entraînera un nivellement par le bas.
- (3) Grâce à ces échelles, on peut, pour chacun des niveaux, appréhender le type de tâches que les élèves sont capables d'accomplir avec une certaine réussite.
- (4) Soulignons que la plupart des recherches sur les effets du redoublement ne comporte qu'un post-test.
- (5) Il convient de recommander ce type d'analyse chaque fois que l'on ne peut procéder à une assignation aléatoire des sujets aux différents groupes étudiés et chaque fois que l'on ne peut garantir la parfaite (ou quasi parfaite) fidélité du prétest. Le traitement mis au point par Heuchenne (1993) prend en considération :
 - la fidélité globale du post-test ;
 - la fidélité du prétest, dont on se sert pour tenir sous contrôle les compétences initiales des élèves ;
 - l'orientation et l'importance de la corrélation entre traitement et compétence initiale ;

– une composante appelée **tau** qui est une expression complexe, où entrent en compte les fidélités déjà citées, mais aussi une corrélation globale éventuelle des erreurs de mesure des pré et des post-tests ainsi que le coefficient de régression du post-test sur le prétest.

- (6) Rappelons brièvement les principes de base de cette technique. Les recherches portant sur une même variable (ici, le redoublement) sont recensées. Sont conservées les études dans lesquelles les chercheurs ont procédé à la comparaison d'un groupe expérimental avec un groupe de contrôle. Les résultats sont exprimés sous forme standardisée : différence entre groupe expérimental et groupe de contrôle, divisée par l'écart-type du groupe de contrôle.

$$\frac{\text{Moyenne du groupe expérimental} - \text{moyenne du groupe de contrôle}}{\text{écart-type du groupe de contrôle}}$$

= Amplitude de l'effet.

L'importance ou l'amplitude de l'effet de la variable étudiée est exprimée en fractions d'écart-type ; elle peut également être exprimée en centiles.

- (7) Certes, il existe des items d'ancrage (items qui sont présents dans les tests de première et de seconde année) et l'étalement des scores rend les données des deux échantillons comparables, mais ceci ne rend pas le test de deuxième année plus facile.

BIBLIOGRAPHIE

- COOK T.D., CAMPBELL D.T. (1979). – **Quasi experimentation, Design and analysis issues for field settings.** Chicago : Rand Mc Nally.
- CRAHAY M. (2002). – La recherche en éducation : une entreprise d'intelligibilité de faits et de représentations ancrés dans l'histoire sociale. *In* F. Leutenegger & M. Saada-Robert (Ed.), **Expliquer et comprendre en sciences de l'éducation** (Raisons éducatives N° 5, p. 253-275). Bruxelles : De Boeck.
- CRAHAY M. (1996). – Recherche en éducation et réflexion de l'action éducative. *In* C. Hadji & J. Baillé (Ed.), **Recherche en éducation. Vers une nouvelle « alliance »** (Pédagogies en développement, p. 125-160). Bruxelles : De Boeck.
- CRAHAY M. (2003). – **Peut-on lutter contre l'échec scolaire ?** (2^e éd. rev. et augm.). Bruxelles : De Boeck.
- DE LANDSHEERE G. (1992). – **Dictionnaire de l'évaluation et de la recherche en éducation.** Paris : Presses Universitaires de France.
- GLASS G.V. (1977). – Integrating findings : the meta-analysis of research. *In* L.S. Shulman (Ed.), **Review of Research in Education** (Vol. 5, p. 351-379). Itasca, IL : Peacock.
- GRISAY A. (1984). – Les mirages de l'évaluation scolaire (1). Rendement en français, notes et échecs à l'école primaire ? **Revue de la Direction Générale de l'Organisation des Études**, XIX (5), 29-42.
- GRISAY A. (1993). – **Le fonctionnement des collèges et ses effets sur les élèves de sixième et de cinquième.** Paris : Ministère de l'Éducation Nationale, Direction de l'Évaluation et de la Prospective (Les Dossiers Éducation et Formation ; n° 32).
- GRISAY A. (2001). – Évaluer les dispositifs de prise en charge d'élèves faibles (ou forts) : l'utilisation de groupes naturels entraîne des artefacts. **Cahiers du Service de Pédagogie expérimentale.** Liège : Université de Liège.
- HEUCHENNE C. (1993). – Effet des erreurs de cité d'un traitement. **Mathématique, Informatique et Sciences humaines**, 31^e année, 122, 5-19.
- HOLMES C.T. (1990). – Grade Level Retention Effects : A Meta-Analysis of Research Studies. *In* L.A. Shepard & M.L. Smith (Eds.), **Flunking Grades. Research and Policies on Retention** (p. 16-33). Bristol : Falmer Press.
- HOLMES C.T. & MATTHEWS K.M. (1984). – The effects of nonpromotion on elementary and junior high school pupils : a meta-analysis. **Review of Educational Research**, 54 (2), 225-236.
- JACKSON G.B. (1975). – The research evidence on the effects of grade retention. **Review of Educational Research**, 45 (4), 613-635.
- SEIBEL C. (1984). – Genèses et conséquences de l'échec scolaire. **Revue française de pédagogie**, n° 67.
- SLAVIN R.E. (1990). – Achievement effects of ability grouping in secondary schools : a best-evidence synthesis. **Review of Educational Research**, Fall, 60 (3), 471-499.
- PETERSON S. E., DEGRACIE J. S. & AYABE C. R. (1987). – A longitudinal study of the effects of retention/promotion on academic achievement. **American Educational Research Journal**, 24 (1), 107-118.
- POPPER K.R. (1982). – **La connaissance objective.** Paris : Éditions Complexe.
- VAN DER MAREN J.M. (1995). – **Méthodes de recherche pour l'éducation.** Bruxelles : De Boeck.

