

Sommaire

Politiques éducatives et évaluation : nouvelles tendances, nouveaux acteurs <i>Georges Felouzis & Siegfried Hanhart</i> <i>Université de Genève</i>	7
Réformes et ordres universitaires locaux <i>Catherine Paradeise* & Jean-Claude Thoenig**</i> <i>*Université Paris-Est Marne-la-Vallée, LATTS et IFRIS</i> <i>**Université Paris-Dauphine, DRM</i>	33
La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone : nouveau paradigme politique et médiation des experts <i>Christian Maroy* & Catherine Mangez**</i> <i>*Universités de Montréal et de Louvain, CRIFPE et GIRSEF</i> <i>**Université de Louvain, GIRSEF</i>	53
L'évaluation des performances scolaires des élèves : un instrument d'évaluation des politiques éducatives ? <i>Nathalie Mons* & Marcel Crahay**</i> <i>*Université Paris-Est Marne-la-Vallée, LATTS</i> <i>**Universités de Genève et de Liège</i>	77
La place de l'évaluation dans les politiques éducatives : le cas de la taille des classes à l'école primaire en France <i>Pascal Bressoux & Laurent Lima</i> <i>Laboratoire des Sciences de l'Éducation</i> <i>Universités Pierre-Mendès-France</i> <i>et Joseph-Fourier, Grenoble</i>	99
Les contradictions organisationnelles de l'évaluation. L'expérience des chefs d'établissement français <i>Anne Barrère</i> <i>Université Paris Descartes, CERLIS</i>	125

Développement de l'évaluation externe et restructuration du métier de direction d'établissement scolaire au Canada	143
<i>Branka Cattonar* & Claude Lessard**</i>	
<i>*Université de Louvain, CIRTES et GIRSEF</i>	
<i>**Université de Montréal, CRIFPE</i>	
Évaluation, concertation, décision: quelle régulation pour le système éducatif?	
Le cas de l'enseignement primaire genevois	165
<i>Georges Solaux*, Jean-Marc Huguenin**, Jean-Paul Payet*** & José V. Ramirez****</i>	
<i>*Université de Bourgogne,</i>	
<i>**IDHEAP, Lausanne</i>	
<i>***Université de Genève</i>	
<i>****HEG et Leading House en économie de l'éducation, Genève</i>	
L'évaluation des élèves: externalisation et hétérogénéité	185
<i>José V. Ramirez</i>	
<i>HEG et Leading House en économie de l'éducation, Genève</i>	
Quelle est l'importance des compétences évaluées dans les avantages liés à la formation sur le marché du travail?	201
<i>Jean-Marc Falter</i>	
<i>Université de Genève, Leading House en économie de l'éducation</i>	
Notices biographiques	219
Table des matières	227

Politiques éducatives et évaluation : nouvelles tendances, nouveaux acteurs

Georges Felouzis & Siegfried Hanhart
Université de Genève

L'évaluation est aujourd'hui au cœur de la question éducative, non pas seulement au sens de la docimologie (Piéron, 1963), mais aussi au sens anglo-saxon de l'*accountability* dès lors que les acteurs de l'éducation doivent « rendre des comptes » pour asseoir la légitimité de leur action. Or l'évaluation ne va pas de soi, tant du point de vue des procédures à mettre en œuvre, que des interprétations possibles des résultats et de ses conséquences sur les acteurs. En témoignent – pour ne considérer que les cas les plus récents – les mouvements sociaux suscités en France par les réformes universitaires renforçant l'évaluation individuelle et ses conséquences sur les carrières des enseignants-chercheurs. On peut aussi relever les publications qui se font l'écho des interrogations grandissantes sur les moyens, les buts et les modalités légitimes de l'évaluation dans le champ éducatif en général (Felouzis, 2005), de l'usage des palmarès (Centre international d'études pédagogiques, 2010) ou plus spécifiquement dans le cadre des réformes universitaires marquées par « la fièvre de l'évaluation » pour reprendre le titre d'un numéro de la *Revue d'Histoire Moderne et Contemporaine* (2008).

L'émergence de l'évaluation comme mode de pilotage des politiques publiques est historiquement associée à une nouvelle conception du rôle de l'État dans la société et au courant dit du *New Public Management*. Dès le début des années 1980, on assiste notamment en Grande-Bretagne et aux États-Unis, à une « révolution bureaucratique » qui instaure de nouveaux critères de gestion basés sur la concurrence des établissements et le pilotage par les résultats. Dans le domaine de la santé, du maintien de l'ordre ou encore de l'éducation, ces nouveaux critères de gestion publique ont eu

pour conséquence de changer de façon radicale les comportements individuels, appliquant au domaine des services publics des critères d'évaluation jusqu'ici réservés au domaine marchand (Le Galès & Scott, 2008). D'un point de vue historique, les conceptions « libérales » de la société et des rapports sociaux ont été déterminantes dans cette évolution. On observe une conjonction historique entre cette idéologie politique et sociale d'une part, et l'émergence de l'évaluation comme mode de gouvernement de l'autre, ainsi qu'une préoccupation croissante pour une plus grande efficacité et efficience¹ de l'action publique, tant dans les pays du Nord (Organisation de coopération et de développements économiques [OCDE], 1997) que dans ceux du Sud (Sall, 1996).

Toutefois, on ne peut se satisfaire, pour penser les nouvelles formes d'évaluation dans le domaine éducatif, d'un tel niveau de généralité. Il nous faut raisonner à d'autres niveaux, celui par exemple des organisations et ainsi souligner, en reprenant partiellement les catégories de la sociologie de Mintzberg (1982), un des aspects fondamentaux de cette évolution. Nous pensons au passage, qu'induit la généralisation des procédures évaluatives, d'un fonctionnement régi par des *bureaucraties professionnelles* à un fonctionnement centré sur l'évaluation et sa diffusion auprès d'un large public. Ce passage, que Maroy désigne sous l'expression de « régulation post-bureaucratique » (Maroy, 2008), relève d'une évolution majeure dans les outils de régulation des comportements collectifs et individuels. Dans le cas de la bureaucratie professionnelle, la forme dominante de la régulation est construite par les professionnels qui décident eux-mêmes de la nature de leur travail et des critères d'excellence. Les normes professionnelles prévalent sur celles de la « ligne hiérarchique ». La régulation des comportements au sein même de l'organisation émane alors de la socialisation au métier et des qualifications, avec un faible contrôle direct sur les activités de travail des individus dont le degré d'engagement peut être très variable. Cet idéal type du fonctionnement de certaines organisations rend assez bien compte des organisations éducatives au sein desquelles les normes et valeurs proprement professionnelles ont un poids très fort². On peut d'ailleurs lire le récent débat genevois sur l'opportunité de créer un corps de chefs d'établissement dans l'enseignement primaire comme le fruit d'une volonté politique de limiter le poids des enseignants en créant une ligne hiérarchique qui limiterait leurs capacités d'action. Ce débat a

1. L'efficacité est un concept de l'économie. Appliqué à un système éducatif, il exprime son efficacité compte tenu du coût économique qu'il engendre.

2. Il est probable que cette adéquation est plus forte dans l'enseignement supérieur que dans l'enseignement primaire ou secondaire. Toutefois, le concept de bureaucratie professionnelle rend assez bien compte du fonctionnement de structures au sein desquelles les professionnels (en l'occurrence les enseignants) ont une marge d'autonomie forte et une relative indépendance vis-à-vis de leur direction et même de leurs collègues. Il est donc très utile pour penser les organisations éducatives.

eu lieu en France lorsque, dans les années 1990, le même type de réforme a été entrepris. Dans la même perspective, la mise en œuvre de la réforme HarmoS³ en Suisse peut être vue comme une façon de limiter l'autonomie des enseignants en créant des standards de compétence et plans d'études communs, comme cela a pu être le cas en Belgique francophone avec la promotion des socles communs de compétences.

Lorsque la régulation des comportements des acteurs passe par l'évaluation, l'autonomie est moindre pour les professionnels soumis à un contrôle externe. Nous n'avons pas encore de terme précis et définitif pour désigner cette nouvelle forme de bureaucratie (bureaucratie « gestionnaire »?). Lorsque Mintzberg (1982) emploie le terme de « bureaucratie professionnelle », il exprime l'idée que le centre du pouvoir décisionnel se situe au niveau des professionnels eux-mêmes. Les universitaires décident des orientations à donner à l'université, des personnes qu'il faut recruter, des critères pertinents pour les recrutements, etc. Si l'on veut forger un nouveau concept « Mintzbergien » qui placerait l'évaluation au centre des processus de régulation des comportements au sein des organisations, il nous faudrait raisonner en termes de pouvoir : qui détient le pouvoir dans le cas d'une bureaucratie régulée par l'évaluation ? C'est l'un des objectifs de cet ouvrage que d'éclaircir ce point et ainsi de forger un nouveau concept descriptif. Nous pouvons décrire la nature de cette nouvelle forme de bureaucratie par l'usage qu'elle fait de l'évaluation : tout se passe comme si sa logique était plus de publiciser les résultats comparatifs que de mesurer avec précision les effets de l'action des individus au travail. Cette publicité de l'évaluation a pour objectif de produire une dynamique « externe » au travers des mécanismes du marché (choix de l'établissement en fonction de sa qualité réelle ou supposée par exemple), et du débat public autour des actions pertinentes à entreprendre pour améliorer le système. Les palmarès des hôpitaux ou des établissements scolaires publiés par la presse en France relèvent du premier cas : la « socialisation » des évaluations – au sens où Moscovici parlait de « socialisation de la psychanalyse » (Moscovici, 1961) – en transforme la nature, le sens, voire le contenu même, au point d'en dévoyer les usages, tant il est vrai que les résultats d'une évaluation ne peuvent le plus souvent se comprendre en dehors des conditions de leur production. Le débat sur l'efficacité comparée des systèmes éducatifs cantonaux en Suisse, consubstantiel à la publication des enquêtes PISA (OFS-CDIP, 2005), relève du second cas : le débat public s'engage alors sur les meilleures options

3. L'accord intercantonal sur l'harmonisation de la scolarité obligatoire (concordat HarmoS) est un nouveau concordat scolaire suisse. Il harmonise au niveau suisse la durée des différents degrés d'enseignement, leurs principaux objectifs et le passage de l'un à l'autre, tout en actualisant les dispositions du concordat scolaire de 1970 qui réglementent déjà uniformément l'âge d'entrée à l'école et la durée de la scolarité obligatoire. Les cantons prennent individuellement la décision d'adhérer au concordat. On trouvera plus d'informations sur le site <http://www.edk.ch/dyn/11737.php>

pour harmoniser les systèmes cantonaux autour d'un cadre commun, sur les solutions envisagées pour améliorer les « performances » des systèmes les moins bien évalués ou encore sur leur efficience (Hanhart, 2007). Dans ce cas, le changement social est impulsé par le politique, mais à partir de la mise en débat des évaluations au sein même de l'espace public.

De ce fait, l'usage de l'évaluation suscite l'émergence de nouveaux acteurs des politiques scolaires. N'est-ce pas le cas des « usagers », qui constituent aujourd'hui des acteurs collectifs de premier plan pour comprendre, par exemple, les marchés scolaires dans l'enseignement secondaire en France (Felouzis & Perroton, 2007a; Van Zanten, 2009), et dans d'autres pays (Chubb & Moe, 1988), ou de façon plus explicite encore des associations de parents d'élèves qui interviennent directement dans la mise en œuvre des politiques scolaires par le biais d'initiatives populaires en Suisse?⁴

On pressent donc, à partir de ces quelques remarques introductives, que l'évaluation est aujourd'hui au centre des politiques éducatives comme un instrument à part entière de gouvernance (Duran, 2010; Lascoumes & Le Galès, 2004). Elle est censée réguler les comportements collectifs et individuels et ainsi faire le lien entre d'un côté les politiques éducatives – au sens des grandes orientations données par l'action publique au nom de l'intérêt général – et de l'autre les dispositifs d'enseignement mis en œuvre par les professionnels. Or ce lien, on le sait, n'est jamais direct. Il est toujours le fruit d'une interprétation qui se conjugue avec les intérêts des acteurs sociaux en situation. D'où la nécessité de rendre compte d'une part des modes de construction et d'autre part des usages de l'évaluation dans le champ éducatif.

Nous proposons pour cela de dessiner les pistes de réflexion qui rendent compte de ces évolutions tant pour raisonner sur la nature de l'évaluation, dans ses rapports avec les outils habituellement utilisés dans les sciences sociales, que pour questionner la façon dont les acteurs sociaux s'approprient – ou rejettent – ces nouvelles formes de régulation des comportements au travail. Ces pistes nous conduiront à aborder une réflexion plus générale sur la légitimité des différentes formes d'évaluation et leurs effets au plan collectif comme individuel.

4. Voir à ce sujet les initiatives populaires intitulées « Libre choix de l'école pour tous », lancées dans plusieurs cantons par l'Association Lobbyparents.ch

QUELS RAPPORTS ENTRETIENNENT LES PROCÉDURES ÉVALUATIVES AVEC LES SCIENCES SOCIALES EN GÉNÉRAL ?

Quels sont les liens qui unissent l'évaluation comme instrument des politiques publiques d'une part et les sciences sociales de l'autre ? Qu'est-ce qui distingue les productions scientifiques en sciences sociales des approches évaluatives ? Qu'est-ce qui sépare par exemple une approche critique des inégalités scolaires d'une approche évaluative du fonctionnement d'un (ou plusieurs) système(s) éducatif(s) ? Poser ces questions revient à raisonner sur les critères de démarcation entre ces deux approches de la réalité scolaire aujourd'hui.

D'un certain point de vue, il faut rappeler que les sciences sociales et les procédures évaluatives ont de nombreux points communs, allant parfois jusqu'à se confondre. En effet, leur but respectif est de connaître des dispositifs, reconstruire des mécanismes, faire émerger des lois sous-jacentes. Lorsque Labov (1993) montre comment la participation des jeunes Noirs des ghettos à la culture de rue et au vernaculaire africain-américain limite leurs apprentissages à l'école, ne donne-t-il pas à voir une évaluation d'un système éducatif qui reproduit les inégalités scolaires en ignorant les inégalités raciales et sociales de la société américaine ? Lorsque Piketty et Valdenaire (2006) décrivent les effets de la taille des classes sur les chances de réussite scolaire, ne proposent-ils pas une évaluation des procédures qui produisent des inégalités liées au contexte éducatif ? On pourrait bien entendu multiplier les exemples pour montrer la proximité entre les méthodes et les objectifs des travaux de sciences sociales et les travaux d'évaluation. Après tout, et si l'on accepte une vision « extensive » de l'évaluation, on pourrait considérer qu'elle recouvre une grande partie des productions de sciences sociales dans le domaine éducatif. *L'évolution pédagogique en France*, ouvrage majeur de Durkheim (1938) sur les structures éducatives et les savoirs qu'elles valorisent et transmettent du Moyen Âge au XIX^e siècle, peut en effet être lu comme un bilan évaluatif du fonctionnement des systèmes éducatifs en France et de leurs liens avec les évolutions sociales et culturelles. De même, on peut voir dans *Les héritiers* de Bourdieu et Passeron (1964) une évaluation du système éducatif français des années 1960 qui insiste sur la distance qui sépare les principes affichés (égalité des chances et mérite) de la réalité des faits observés (inégalités sociales, exclusion scolaire par la culture). Ou encore raisonner sur l'efficacité en éducation à partir de « la loi de Baumol » revient à comprendre les ressorts de la hausse tendancielle des coûts éducatifs dans les économies développées. On est donc ici dans l'idée que l'évaluation et les sciences sociales ne font après tout qu'une seule et même chose : elles dévoilent, par la mise en relation de faits et de concepts, les réalités sous-jacentes de l'école et de l'enseignement. Bien plus, l'évaluation des politiques publiques – en éducation comme dans d'autres domaines

de la réalité sociale – est un champ à part entière des sciences politiques et dans ce cas, évaluation et sciences sociales se confondent pour ne faire qu’une seule et même chose, comme le note fort à propos Duran (1999).

Cette première conception insiste sur la dimension critique de l’évaluation qui confronte les résultats de l’action publique aux objectifs à atteindre et aux valeurs qu’elle défend. Toutefois, d’un autre point de vue, et malgré l’influence déterminante qu’ont pu avoir certains travaux de sciences sociales sur la mise en place des politiques scolaires (Langouët, 2008), les objectifs et les fondements de l’évaluation ne peuvent être mécaniquement associés aux sciences sociales. La connaissance scientifique s’applique à des objets que l’on pourrait qualifier d’« abstraits » dans la perspective d’établir dans le meilleur des cas un *non faux provisoire*, selon les termes de l’épistémologie bachelardienne. Le raisonnement s’applique alors à définir des concepts, y compris lorsqu’on étudie des contextes concrets. Ceux-ci sont toujours appréhendés dans leur dimension générale au sens où l’on étudie les « effets établissement » ou les « effets classe » en eux-mêmes, et non pas l’effet de tel établissement ou de telle classe sur les acquisitions des élèves. Dans le premier cas, on produit une connaissance sur les mécanismes qui expliquent les variations des performances des élèves en fonction de certains facteurs (la pédagogie mise en œuvre, les effets de composition, la nature des programmes, etc.), ce qui revient empiriquement à tester l’effet de « variables aléatoires ». Dans le second, on associe directement la performance collective d’une classe à un objet concret personnalisé par un enseignant ou un groupe d’enseignants. En d’autres termes, dans le premier cas, on produit une connaissance, dans le second une injonction normative, qui n’est jamais éloignée de la recherche de « responsables » voire de « coupables » pour expliquer un dysfonctionnement organisationnel quelconque.

Il n’est néanmoins pas toujours aussi simple de distinguer recherches évaluatives et productions scientifiques tant les procédures et les objectifs peuvent être proches. On pourrait donc penser que, dans ce cas, seuls les usages diffèrent. Il s’agirait alors plus d’une question d’usage des sciences sociales que de différences de « nature » entre sciences et évaluation. On se situerait alors dans une perspective plus relativiste en avançant qu’il n’y a pas d’un côté la science et de l’autre l’évaluation, mais un usage scientifique de l’évaluation et un usage évaluatif de certains résultats des sciences sociales. Dans le premier cas, on peut situer les enquêtes PISA qui répondent à des objectifs d’évaluation (quels systèmes, quels pays, produisent plus ou moins de compétences, plus ou moins d’inégalités entre élèves et selon quels critères) dans une perspective de normalisation des systèmes éducatifs et de gouvernement « par les nombres » (Desrosières, 2008 ; Grek, 2009). Mais en même temps, rien n’empêche de produire des résultats scientifiques à partir de ces enquêtes sur les conséquences des orientations plus ou moins précoces, des filières différenciées, de l’autonomie des établissements ou encore

du redoublement, etc. Il en est de même des indicateurs de performance des lycées en France. Leur visée est évaluative, mais leurs usages peuvent être scientifiques pour raisonner notamment sur les évolutions de la ségrégation scolaire ou les effets d'établissement (Felouzis & Perroton, 2007a) ou sur les conséquences de la récente libéralisation de la carte scolaire (Merle, 2010). Inversement, le deuxième cas – celui des usages évaluatifs des résultats des sciences sociales – renvoie à l'influence, souvent indirecte mais non moins réelle, des sciences sociales sur les réformes éducatives. La mise en place du « collège unique »⁵ en France au milieu des années 1970 a trouvé l'une de ses sources dans la sociologie critique de l'éducation des années 1960 qui insistait sur le caractère inégalitaire et injuste du système éducatif français fortement filiarisé (Baudelot & Establet, 1971 ; Bourdieu & Passeron, 1970).

En définitive, les ambiguïtés de l'évaluation entre pratique scientifique et usages normatifs peuvent être vues non pas comme un problème à résoudre, mais comme une de ses dimensions constitutives. Il s'agit à la fois du « fruit du développement d'une réflexivité des États, des pilotes, des établissements, des équipes enseignantes » et d'une « pratique stratégique et politique » (Demailly, 2006, p. 119). Il devient alors pertinent d'en analyser les usages et les conséquences sur les acteurs de l'éducation et plus généralement sur les modes de régulation des comportements professionnels.

DE LA RÉGULATION PAR LA QUALIFICATION À LA RÉGULATION PAR LES ACCOUNTS : ÉMERGENCE DE NOUVEAUX ACTEURS COLLECTIFS ?

Comment se construit l'articulation entre l'élaboration des politiques éducatives d'une part et leur mise en œuvre de l'autre ? L'évaluation n'est-elle pas devenue aujourd'hui le lien indispensable entre ces deux niveaux de l'action publique en matière éducative ? C'est en tout cas de cette façon qu'elle se présente. Le « pilotage par l'évaluation » peut être vu comme une façon de réguler les comportements professionnels au niveau collectif comme individuel, en donnant une mesure la plus fiable possible de ce qui sépare les acteurs en situation des objectifs à atteindre. Dans ce que Mintzberg (1982) nomme le modèle de la bureaucratie professionnelle, seule l'éthique (ou la « conscience ») professionnelle joue ce rôle. Dans ce cas, l'enseignant s'investit dans son travail simplement parce qu'il a été formé pour ça, il a été socialisé à des normes professionnelles qui lui permettent d'agir de façon « conforme » aux attentes⁶. Outre que cet investissement

5. Le « collège unique » est mis en place en France en 1975 par la loi dite « Haby » du nom de son promoteur, le ministre de l'Éducation René Haby. Tous les élèves accèdent à l'enseignement secondaire après l'école primaire.

6. Nous prenons ici l'exemple de l'enseignement, mais ce modèle s'applique à d'autres domaines comme la santé par exemple.

peut être variable selon les individus, cela implique aussi que les buts et les moyens institutionnels sont définis par la profession elle-même, et donc en fonction de ses valeurs. Dans le cas de la *régulation par l'évaluation*, il en va tout autrement car le « contrôle » (au sens de contrôle social) sur les individus et les collectifs est externe, et se conçoit en fonction du politique plus qu'en fonction de la profession. Dans cette perspective, on considère l'évaluation comme un outil de régulation qui déplace le centre du « pouvoir » (au sens de capacités d'action) des professionnels (les enseignants, les chefs d'établissement qui jouissent par leur statut d'une marge importante d'autonomie) au politique et aux usagers.

Comment qualifier cette évolution ? Pour Maroy (2008) il s'agit, tout au moins dans le contexte européen, d'un passage d'une régulation professionnelle à une régulation « post-bureaucratique » dans laquelle l'évaluation tient une place centrale. Cette place n'est jamais aussi visible que dans les transformations des politiques universitaires de la fin du XX^e siècle en France. Système fortement structuré depuis le début du XIX^e siècle autour des disciplines (Musselin, 2001), l'université française évolue depuis le début des années 1990 vers un système plus centré sur les établissements. Or, cette évolution implique un affaiblissement de la *bureaucratie professionnelle* au bénéfice d'autres critères plus ancrés dans des objectifs politiques que l'on voit à l'œuvre dans la définition même des diplômes et des contenus d'enseignement. Dans une logique professionnelle, il n'est pas souhaitable de construire une offre de formation centrée sur autre chose que les contenus disciplinaires eux-mêmes, alors que dans une logique d'établissement ou de « pôles », d'autres critères deviennent pertinents, comme la professionnalisation et l'insertion des étudiants. L'évaluation des universités sur des critères de taux de réussite et d'insertion des diplômés change la donne et produit une régulation des comportements tout autre, créant des dispositifs d'enseignement dont la logique dépend plus des objectifs politiques et sociaux que des normes professionnelles des universitaires. Dans la même logique, mais au niveau de l'enseignement secondaire, le développement des marchés scolaires et d'une concurrence entre établissements pour « capter » les meilleurs élèves peut être vu comme le résultat d'une logique d'évaluation dont témoigne la publication en France des indicateurs de performance des lycées. Cette logique de l'évaluation a conditionné l'émergence de dispositifs d'enseignement aptes à attirer les meilleurs élèves, tels que les classes à thème, l'enseignement de langues rares ou encore le développement de « filières européennes » dont la logique est plus à chercher dans le développement des marchés scolaires que dans les nécessités de la professionnalité enseignante. Il nous faut ajouter que l'évaluation en elle-même ne produit pas ces phénomènes, mais elle les accompagne de façon constante. Les marchés scolaires ne sont pas le résultat direct des indicateurs de performance des lycées, il s'agit de phénomènes complexes qui impliquent un grand nombre d'acteurs (Felouzis & Perroton, 2007a ; Maroy & Van Zanten,

2007) et ceci dans la plupart des pays européens (Maroy, 2006). Toutefois, on peut observer avec Demailly (2006) que les marchés scolaires tendent à s'accompagner, comme en Angleterre notamment, d'un État évaluateur montrant la forte implication de ces deux dimensions idéal types.

On peut alors s'interroger sur la façon dont ce processus de mise en perspective par l'évaluation redistribue les cartes entre les différents acteurs de l'école, et conduit à concevoir des critères d'évaluation différents. Le poids des « usagers » augmente alors que celui des professionnels diminue dans la définition des buts institutionnels et des moyens à mettre en œuvre pour y parvenir. Les débats sur les savoirs enseignés dans le primaire et le secondaire reflètent cette évolution, tant en Suisse qu'en France. On peut probablement voir dans le retour aux « fondamentaux » dans l'enseignement à l'école primaire (lire, écrire, compter) et au travers des débats sur les méthodes pédagogiques⁷ le résultat d'une définition « efficace » de l'école centrée sur les savoirs de base bien plus que sur une formation plus large des élèves au sein de dispositifs pédagogiques pensés en fonction des normes progressivement constituées de la profession enseignante. Les comparaisons que produit la Direction de l'évaluation, de la prospective et de la performance (DEPP) en France sur la maîtrise de ces fondamentaux à 20 ans d'intervalle (MEN-DEPP, 2008) auprès d'élèves de niveau CM2⁸ peuvent être lues comme des évaluations des pratiques pédagogiques des maîtres qui définissent en fonction de critères plus politiques les objectifs mêmes de l'enseignement. Il ne s'agit pas ici de déplorer l'existence de ces évaluations, mais de simplement noter qu'elles produisent une régulation par les *accounts* plus que par le savoir-faire professionnel des maîtres. On est donc dans un changement de « culture » qui voit la régulation par les *accounts* remplacer la régulation par l'éthique professionnelle et les savoir-faire.

Ajoutons que ces nouveaux modes de gouvernance conduisent les acteurs traditionnels à adopter de nouvelles modalités d'interagir et de négocier avec l'école. Ce fut le cas en Belgique où le décret « mixité » de 2008 a suscité des formes de mobilisation de la part des parents inconnues jusqu'alors pour inscrire leurs enfants dans le « meilleur » établissement. Mais c'est aussi le cas des évaluations internationales PISA dont les effets sur les systèmes éducatifs et les acteurs de l'éducation sont loin d'être négligeables. Le cas de l'Allemagne est, dans cette perspective, exemplaire. Dans son article « Governing by numbers », Grek (2009) montre comment

in response to the Pisa findings, German education authorities organised a conference of ministers in 2002 and proposed reforms of an urgent nature, such

7. Nous pensons au débat sur les méthodes d'apprentissage de la lecture en France (« globale » ou « syllabique ») ou encore à la polémique sur la nature et la place des notes dans l'enseignement et l'évaluation des élèves dans le canton de Genève.

8. Ce qui correspond à la dernière année d'école primaire en France.

as developing standards for measuring students' competences upon completion of secondary schooling and the introduction of large-scale assessment testing at the end of primary and secondary education. (p. 7)

Les conceptions traditionnelles de l'éducation en Allemagne sont fondées politiquement sur une forte décentralisation (l'éducation est de la compétence des *Länder*) et pédagogiquement sur la *Bildung*, qui tend à mettre l'accent plus sur l'épanouissement des élèves que sur la transmission des savoirs. Or, les résultats des premières évaluations PISA, qui plaçaient l'Allemagne dans le bas de l'échelle des nations au plan des compétences de ses jeunes, tendent à bouleverser ces principes pour aller vers une rationalisation plus marquée des politiques scolaires. On est donc ici typiquement, et à l'échelle d'un pays, dans le cas d'un passage d'un mode de régulation par la qualification (des enseignants principalement) à un mode plus centré sur les *accounts*. Cette transformation dessine des tendances plus qu'une casure nette. Mais il reste important de souligner la forte capacité illocutoire des enquêtes PISA qui – malgré les critiques dont elles font régulièrement l'objet – deviennent aujourd'hui un référentiel des politiques éducatives dans beaucoup de pays (Mons, 2007) au point que les résultats PISA 2000 ont pu apparaître en Allemagne comme une *Bildungskatastrophe*.

D'autres exemples dans d'autres contextes nationaux pourraient être donnés. L'important est ici de souligner qu'en s'imposant comme le lien qui unit les politiques publiques d'une part et leur réalisation concrète dans des dispositifs pédagogiques de l'autre, l'évaluation est devenue, bien plus qu'une mesure externe et ex post de l'action éducative, un outil de modelage de sa forme et de son sens. Ce qui pose avec d'autant plus d'acuité la question de la validité et de la légitimité des différentes formes d'évaluation.

QUELLES SONT LES CONDITIONS DE VALIDITÉ ET DE LÉGITIMITÉ DE L'ÉVALUATION ?

Dans leurs modalités de construction, les procédures évaluatives empruntent largement, on l'a dit, aux sciences sociales. Or, on peut se questionner sur leur validité dès lors que leur but n'est pas exclusivement de produire de la connaissance mais d'orienter les comportements des professionnels et de rendre compte de la mise en œuvre des politiques éducatives. En d'autres termes, les questions de méthode sont ici déterminantes au regard des conséquences sociales et individuelles de l'évaluation. Mais il ne suffit pas d'évaluer l'évaluation pour être quitte de sa légitimité. Encore faut-il penser ses usages politiques et sociaux.

Procédons par l'exemple. Les indicateurs de performance de lycées en France existent depuis le milieu des années 1980. Ils ont été créés par la

DEPP⁹ pour présenter sur une base la plus objective possible des résultats par établissement qui prennent en compte la nature des publics scolarisés. Ils sont construits sur les chances de réussite au baccalauréat (la maturité en Suisse) par lycée, compte tenu de l'âge et de l'origine sociale des élèves. Ces indicateurs ont, pendant des années, donné lieu à des palmarès dans la presse, classant chaque établissement en fonction de sa « valeur ajoutée » calculée en comparant le différentiel entre l'attendu¹⁰ et l'observé de la réussite au bac. Pendant des années, ces indicateurs ont défini les « bons » et les moins bons établissements sur la base de ce calcul. Or, une évaluation de ces procédures conduite par Felouzis (2005) en a montré les limites : ne prenant en compte que l'âge et l'origine sociale des élèves pour calculer l'attendu, ces indicateurs présentaient un biais systématique par le simple fait que certains établissements ont dès le départ des élèves scolairement meilleurs que ne le laisse penser leur âge et leur origine sociale, alors que d'autres ont des élèves bien plus faibles que prévu. Nous avons ainsi montré que la prise en compte du niveau de départ des élèves bouleversait de façon très sensible la « valeur ajoutée » des établissements. Ce qui signifie qu'une évaluation des établissements, présentée comme « scientifique » et émanant d'une instance ministérielle officielle, pénalisait en fait les établissements au départ les plus défavorisés. Les équipes pédagogiques, à chaque publication des résultats de leur lycée, se voyaient ainsi attribuer une « valeur ajoutée » négative, sans pour autant que leur action ne soit pour quoi que ce soit dans ce phénomène. Et encore, l'évaluation des établissements en France n'a pas les conséquences qu'elle peut avoir dans d'autres pays. Dans leur analyse critique de la politique d'*accountability* de plusieurs États concernant les écoles, les économistes Kane et Staiger (2002) montrent les limites de cette politique qui consiste à attribuer des moyens en fonction des résultats des élèves à des tests. Les effets de ces politiques peuvent, certes, encourager certaines écoles à accentuer leurs efforts sur la qualité de l'enseignement, mais les résultats globaux ne sont pas à la hauteur de l'investissement économique consenti par les États. D'un point de vue plus individuel, ces procédures ont aussi des effets pervers : elles tendent à pousser les enseignants à centrer leurs activités sur les acquis mesurables par des tests au détriment des autres, ce qui revient à fortement limiter l'ampleur des enseignements.

On pourrait donner d'autres exemples d'indicateurs peu fiables, mais faisant l'objet d'une croyance forte en leur capacité à traduire une certaine réalité. Par exemple dans l'enseignement supérieur le « classement de Shanghai » dont Gingras montre toute la relativité et les écueils dans un article récent (Gingras, 2008). L'important ici est de rappeler que l'évaluation ne produit rien d'autre que des indicateurs et que leur fiabilité et leurs limites

9. Dans les années 1980, il s'agissait de la Direction de l'évaluation et de la prospective (DEP) aujourd'hui DEPP, le deuxième « P » étant pour « Performance ».

10. L'attendu d'un établissement est le taux de réussite des élèves d'un lycée compte tenu de l'âge et de l'origine sociale de ses élèves.

doivent toujours être mises à l'épreuve. La question devient alors celle du « contrôle qualité » au niveau de la production des résultats, et celle de l'insertion de ces résultats dans le débat public qui se fait toujours au prix d'une « traduction » dont la nature est complexe à définir et les modalités difficilement maîtrisables par les acteurs de l'éducation. Mais là n'est pas, bien entendu, la seule interrogation possible. Car il serait naïf de considérer la question de la « fiabilité » des indicateurs comme relevant d'une stricte dimension « technique », et ceci pour deux raisons. La première est qu'il en est de l'évaluation comme des critères d'excellence dans « le champ académique » décrit par Bourdieu (1984) : ils font l'objet de luttes et de controverses pour leur définition en fonction de la position et des intérêts de chacun. Et là, la dimension politique et sociale est pleinement présente dans la définition même des modes d'évaluation. L'exemple des débats liés au classement des revues scientifiques est, parmi tant d'autres, assez parlant (Fridenson, 2009). Il reflète la nature des luttes sociales et professionnelles à l'œuvre dans la définition même de ce qu'est l'excellence en matière de publication scientifique dans les sciences sociales. La deuxième raison est plus liée à l'usage que font les politiques des évaluations. Elle est notamment illustrée dans le présent ouvrage par la contribution de Bressoux et Lima sur l'usage des résultats scientifiques à propos des effets de la taille des classes sur l'efficacité de l'enseignement et l'ampleur des inégalités. Dans ce cas, une interprétation partielle, voire fautive, des résultats de la recherche par les instances ministérielles chargées de l'évaluation vient légitimer une décision politique aboutissant à l'augmentation de la taille moyenne des classes à l'école primaire et donc tendanciellement à une baisse d'efficacité.

Qu'en est-il dans le champ plus proprement lié à l'économie de l'éducation ? La recherche en matière d'efficacité des politiques scolaires s'est mise en place de façon progressive, pour deux motifs principaux : d'une part le manque de données internationales à large échelle sur les performances des élèves, et d'autre part l'absence de données suffisamment désagrégées sur les ressources mises en œuvre dans le domaine éducatif par chaque pays. Remarquons qu'il ne s'agit pas d'une difficulté propre à l'évaluation des politiques éducatives, mais plus généralement à l'analyse des politiques et des dépenses publiques (Mandl, Dierx & Ilzkovitz, 2008). Cependant, les données désormais disponibles sont assez larges et fiables pour donner lieu à des analyses approfondies. Nous en voulons pour preuve l'évolution des analyses d'efficacité telle que nous pouvons l'observer à propos de PISA 2006 et 2009. En 2006 – année où l'accent était mis sur les sciences – l'une des rares analyses d'efficacité portait sur la relation entre les performances des élèves d'une part et la dépense cumulée par élève de l'autre. Les données montraient que la dépense expliquait 19% de la différence de performance moyenne des élèves en culture scientifique (OCDE, 2007). Cependant l'interprétation – et par conséquent la portée – d'un tel résultat était limitée, dans la mesure où la variable représentant les moyens engagés

(la dépense cumulée, toutes matières enseignées confondues) était insuffisamment désagrégée pour expliquer la performance dans une discipline particulière. En 2009 – année où, comme en 2000, l’accent a été mis sur la lecture – l’OCDE a mesuré l’impact de différents facteurs (niveau de rémunération des enseignants, taille moyenne des classes, temps hebdomadaire moyen passé en classe pour l’apprentissage de la langue enseignée, etc.) sur la performance moyenne à l’échelle nationale. C’est ainsi que les résultats de PISA (Organisation for Economic Co-operation and Development [OECD], 2010) font notamment ressortir une corrélation entre la performance des élèves et le niveau de rémunération des enseignants rapporté au revenu national. Cette corrélation est supérieure à celle qui lie la performance des élèves à la taille des classes. Ainsi l’idée que la réduction de la taille des classes est un facteur suffisant pour améliorer les performances des élèves est relativisée dès lors que l’on compare des situations nationales contrastées. Il est clair que l’analyse de l’efficacité des politiques éducatives se développera ces prochaines années et aura de plus en plus d’impact sur l’orientation de ces mêmes politiques et sur la gestion des établissements. Cette tendance participe de ce que nous avons appelé précédemment la régulation par les *accounts*. Cela nous conduit à raisonner sur la définition même de l’efficacité, concept apparemment simple et directement appréhendable, mais qui s’avère, après analyse, multidimensionnel.

QUELLES DÉFINITIONS DE L’EFFICACITÉ POUR L’ÉVALUATION DES POLITIQUES ÉDUCATIVES ?

Nous avons considéré jusqu’ici les buts de l’évaluation comme relevant de l’évidence. Une politique éducative « efficace » peut certes être définie au plan très général comme une politique qui atteint ses objectifs. Pourtant – et c’est là un apport important de l’économie à la question – cette efficacité peut se conjuguer en fonction de critères différenciés si l’on se questionne sur les conséquences économiques de l’action publique. Au plan théorique, l’intérêt des économistes pour l’évaluation des politiques éducatives date d’une cinquantaine d’années, comme en témoignent les travaux fondateurs de la théorie du capital humain de Becker dans les années 1960 (Becker, 1964). Lorsqu’on a commencé à considérer l’éducation comme une source de croissance économique, on s’est efforcé de mettre en évidence les bénéfices économiques de l’éducation en termes de productivité accrue des travailleurs formés (Carnoy, 1995). Cependant, aujourd’hui, l’évaluation économique des politiques éducatives trouve sa justification première dans l’importance des ressources publiques allouées à l’éducation. Et de fait en 2007, les pays de l’OCDE ont en moyenne consacré 13,3 % de leurs dépenses publiques à l’éducation (OCDE, 2010). L’importance de ces dépenses justifie une évaluation des politiques publiques au travers de deux questions principales.

D'abord, les fonds publics sont-ils utilisés de manière optimale? Ensuite, les retombées externes des dépenses publiques correspondent-elles aux attentes? Ces deux questions définissent des modes d'évaluation distincts et des procédures qui ne le sont pas moins. La première relève de la question de l'efficacité interne, la seconde de l'efficacité externe.

Efficacité interne et efficacité des politiques éducatives

La première question revient à s'interroger sur « les bases (et les mesures) d'un bon fonctionnement du système public d'enseignement [...] en tant que système de production particulier » (Jarousse, 1991, p. 94). Il ne s'agit pas seulement de s'intéresser aux performances du système (efficacité), mais aussi à la nature et à l'importance de la relation entre les performances et les moyens engagés, ce que les économistes définissent comme l'efficacité. Revenons à l'enquête internationale PISA 2009 comme exemple d'analyse de l'efficacité et de l'efficacité des politiques d'enseignement primaire et secondaire. Les experts de l'OCDE sont partis de l'hypothèse que la performance des systèmes scolaires nationaux découle notamment de la qualité des enseignants, de la pertinence des curricula, de la mise à disposition d'équipements quantitativement et qualitativement suffisants. Ce n'est pas seulement la disponibilité de ces éléments, mais aussi la manière de les combiner, qui influera sur les performances des élèves et en conséquence sur l'efficacité interne plus ou moins grande de chaque système éducatif (OECD, 2010).

Dans les pays du Nord, depuis une vingtaine d'années, les budgets publics alloués à l'éducation ont tendance à se resserrer, alors que la demande en matière scolaire ne cesse de croître et de se diversifier. Il en découle la nécessité de repenser les politiques publiques, les orientations stratégiques et la gestion des institutions publiques. Dans un tel contexte, l'analyse coût-efficacité pourrait être d'une grande utilité. Celle-ci consiste à comparer les coûts et les retombées prévisibles de différents projets alternatifs; nous utilisons à dessein le conditionnel, car comme le relève Levin (1988), les éducateurs et les évaluateurs dans le domaine éducatif n'ont pas l'habitude d'intégrer ce genre d'analyses dans leur processus de prise de décision. La difficulté à quantifier toutes les retombées d'une action éducative explique ce constat. Par ailleurs, une analyse coût-efficacité en éducation suppose une collaboration multidisciplinaire (économie, sociologie, didactiques, etc.) à laquelle les spécialistes des sciences de l'éducation ne sont pas encore suffisamment préparés.

Au-delà de ce problème, lié à la nature et la multiplicité des champs disciplinaires présents dans le domaine de l'évaluation, il apparaît que l'évaluation de l'efficacité des politiques et des systèmes éducatifs pose

clairement la question de l'utilisation et de la diffusion des résultats. Si cette problématique est encore peu traitée en Europe, elle l'est beaucoup plus aux États-Unis en termes d'*accountability* que l'on traduit par l'obligation politique de rendre compte des résultats de son action – ici pour les directions d'établissements publics. Hanushek et Raymond (2005) ont observé aux États-Unis que les dispositions obligeant les établissements scolaires à rendre compte de leurs résultats ont exercé un effet positif sur les performances des élèves. Cependant cet impact positif n'a été observé que dans les États qui ont assorti l'obligation de rendre compte de sanctions en cas de résultats jugés insuffisants.

Efficacité externe des politiques éducatives : le problème des effets non mesurables

L'évaluation d'une politique éducative ne saurait se limiter à la mesure des effets internes. La théorie du capital humain, à l'origine de l'économie de l'éducation, a mis en lumière l'apport de l'éducation à la croissance économique (Denison, 1962) et l'impact de l'éducation sur le salaire par des mesures de taux de rendement privés (Becker, 1964 ; Mincer, 1958). Si cet impact est statistiquement avéré, il est plus difficile de déterminer la nature précise de cette relation, en particulier les compétences qui sont valorisées sur le marché du travail. C'est précisément cette interrogation que Falter développe dans sa contribution au présent ouvrage en montrant que les compétences évaluées par les enquêtes internationales ne reflètent pas forcément les compétences recherchées dans la formation professionnelle. À ce premier problème structurel s'ajoute le fait que le calcul de la rentabilité privée, publique et sociale des investissements éducatifs présente une limite importante : il ne considère que les retombées quantifiables, omettant les bénéfices privés non marchands tels que « le plaisir intrinsèque à acquérir des connaissances [...] ainsi que la contribution de l'éducation à l'efficacité dans les activités autres que l'activité professionnelle, notamment les activités de loisir » (Grin, 1994, p. 47). Il convient de souligner que les bénéfices non marchands de l'éducation ne concernent pas que les individus qui se sont formés, mais plus généralement la société. Nous nous référons ici aux externalités positives, c'est-à-dire des retombées positives de l'action d'un agent économique sur son environnement. Plus un individu bénéficiera d'un niveau de formation élevé, moins il risquera de commettre des actes délictueux ; il accordera aussi plus d'importance à préserver sa santé et celle de ses enfants. De même, un électeur disposant d'un niveau de formation élevé contribuera à renforcer les processus démocratiques (Moretti, 2005). L'évaluation de l'efficacité des politiques publiques ne tient pas compte des externalités, faute d'un consensus sur la manière d'en estimer l'ampleur (Mueller, 2007) ; de ce fait les taux de rendement sociaux des investissements en éducation sont nécessairement sous-évalués.

Efficacité interne versus efficacité externe ?

Il existe une posture critique face au caractère « interne » de l'évaluation dans les politiques d'évaluation; celle-ci se manifeste notamment par le poids de plus en plus marqué de critères « externes » pour rendre compte de l'efficacité des politiques d'éducation. Ces critères externes ne sont pas, par eux-mêmes nouveaux dans le champ scientifique. On peut toutefois se questionner sur l'émergence du critère d'efficacité économique des diplômes comme critère légitime d'évaluation des politiques d'éducation, notamment pour l'enseignement professionnel et l'enseignement tertiaire. Aujourd'hui, ce critère demeure toujours fortement associé à l'emploi, mais le questionnement a évolué, reflétant de nouvelles interrogations : comment s'opèrent les transitions entre systèmes de formation et marché de l'emploi (Flückiger, Meunier & Falter, 2009) ? Les politiques publiques de formation tendent-elles à faciliter les transitions ? Les mutations constantes dans le monde du travail obligent les individus – quel que soit leur niveau de formation – à réactualiser leurs connaissances professionnelles et à développer de nouvelles compétences, afin d'une part de réduire le risque de chômage et d'autre part de pouvoir profiter des gains salariaux générés par l'évolution technologique. Des études mettent en évidence l'importance de maîtriser certaines compétences, et les facteurs individuels susceptibles d'expliquer le succès sur le marché du travail (par exemple Couppié, Gasquet & Lopez, 2007 ; Falter, Pasche & Hertig, 2007). Dès lors que l'individu ne possède pas ou plus les compétences nécessaires à son insertion (ou à son maintien) sur le marché du travail, il convient de s'interroger : dans quelle mesure les politiques publiques de formation prennent-elles en compte le phénomène de dépréciation du capital humain ? La réponse à cette question permet aussi d'évaluer l'efficacité des politiques publiques de formation.

L'efficacité d'une politique publique dépend non seulement des ressources financières mobilisées, mais aussi des mécanismes d'allocation des fonds publics. Prenons l'exemple d'une évaluation des politiques publiques de formation tenant compte des besoins de formation des adultes. Le succès de telles politiques dépendra dans une large mesure de l'efficacité des modalités d'attribution des fonds publics. L'alternative au financement public des prestataires de services éducatifs consiste à orienter les subventions publiques vers les usagers. Il existe cependant un risque de susciter des effets de substitution (le bénéficiaire profite de la subvention publique pour se faire financer une formation dont il était prêt à assumer le coût) et d'aubaine (le bénéficiaire utilise la subvention publique pour financer une formation dont lui seul retirera un profit, à l'exclusion de la collectivité) (Wolter *et al.*, 2003). Nous pourrions débattre ici de l'efficacité du financement public *versus* le financement privé de l'enseignement, mais cette question mériterait à elle seule que nous y consacrons un ouvrage. Nous nous limiterons à faire remarquer que l'impact d'une politique publique de formation dépendra aussi de l'efficacité des mécanismes de financement.

GOUVERNER L'ÉDUCATION PAR LES NOMBRES: USAGES ET CONSÉQUENCES

Les différentes pistes de réflexion dessinées dans ce chapitre introductif montrent que l'analyse de la place de l'évaluation dans les politiques éducatives constitue un champ de recherche qui touche à l'ensemble des disciplines des sciences sociales. C'est en convoquant la sociologie comme l'économie ou les sciences politiques et de l'éducation qu'il est possible d'objectiver l'évaluation, c'est-à-dire de la prendre pour objet. Mais le caractère volontairement exploratoire de notre propos montre aussi qu'il s'agit d'un champ de recherche récent, dont la capacité critique est un atout majeur dès lors qu'il s'agit de comprendre la nature de l'évaluation et ses fondements, la façon dont elle émerge comme un outil légitime de gouvernance et ses effets sur les organisations et les acteurs de l'éducation. Le propos de cet ouvrage est donc de présenter des résultats de recherches empiriques de façon à refléter la diversité des approches possibles dans la perspective de dépasser les clivages trop simplistes entre « partisans » et « opposants » et ainsi éviter deux écueils. Le premier, que nous avons largement traité ici, consisterait à croire que toute évaluation est le reflet fidèle de la réalité qu'elle est censée mesurer. Cette attitude naïve contrevient de façon évidente à la posture critique que l'on se doit d'adopter devant tout résultat de recherche. Le deuxième écueil est à notre sens tout aussi contre-productif. Il s'agit d'une attitude de rejet de toute évaluation sous prétexte que les résultats de l'éducation échappent en grande partie au mesurable et que seuls les professionnels sont légitimes pour juger d'autres professionnels « en interne ». Cet ouvrage montre la nécessité de dépasser ces deux écueils pour penser l'évaluation dans le champ plus vaste des politiques éducatives et de leur mise en œuvre.

C'est le propos de Paradeise et Thoenig dans le chapitre « Réformes et ordres universitaires locaux ». Il s'agit pour eux de penser les conséquences de nouveaux critères de qualité sur le fonctionnement des organisations d'enseignement supérieur. Les jugements de la qualité universitaire se conjuguent aujourd'hui, notamment avec l'introduction de « palmarès » des universités au plan mondial comme national, en deux catégories, « l'une contextuelle et synthétique que nous nommons *prestige*, l'autre acontextuelle et analytique que nous nommons *excellence* » (p. 34). L'articulation de ces deux critères permet d'identifier quatre types d'établissements en fonction de l'attention qu'ils portent à ces deux catégories de jugement. La thèse de Paradeise et Thoenig est que, loin d'homogénéiser les organisations universitaires, le poids grandissant de l'évaluation (et donc du critère de *l'excellence*) se conjugue avec des « ordres locaux » qui en changent le sens et les usages. Il faut, nous disent les auteurs,

se doter d'une lunette théorique adéquate pour analyser comment les institutions locales se positionnent dans les faits et dans leur quotidien par rapport au global,

comme agents passifs soumis *de facto* aux injonctions d'un principal exogène ou comme acteurs plus ou moins autonomes mobilisant sous contrainte des ressources de leur environnement au service d'un horizon qui leur est propre. (p. 37)

Et la nature des réactions organisationnelles des établissements universitaires dépend étroitement des ressources à la fois scientifiques et sociales mobilisables par les acteurs individuels et collectifs qui s'y inscrivent. C'est ainsi que les établissements qui forment, au niveau mondial, le « gratin », à la fois prestigieux et excellents, s'organisent autour du capital social de leurs membres, alors que les « missionnaires », à l'opposé des premiers, organisent les relations internes autour d'un égalitarisme fortement ancré dans l'institution locale.

En questionnant les changements de paradigme politique en matière d'éducation en Belgique francophone, Maroy et Mangez se situent en amont de la mise en œuvre et des conséquences de l'évaluation. Ils montrent que la construction de la politique de pilotage et d'évaluation dans le secteur de l'enseignement n'est pas le seul fruit du contexte international en matière éducative ou de la crise que traverse l'école en Belgique francophone au début des années 1990. C'est aussi le résultat de l'action d'entrepreneurs institutionnels qui promeuvent une nouvelle conception de l'école. Celle-ci « est de moins en moins définie comme instance de socialisation et comme "institution" et tend à être considérée comme "système de production scolaire", producteur de compétences et savoirs utiles pour les individus qui composent la société » (p. 62). Cette analyse de l'émergence d'un nouveau paradigme d'action publique, et notamment d'un « État évaluateur » qui « pilote » le système éducatif, donne à voir toute la complexité des processus à l'œuvre et la multiplicité des acteurs partie prenante de ces changements qui tendent à définir l'« efficacité » non seulement comme un but à atteindre, mais aussi – et peut-être surtout – comme une valeur en elle-même. Le souhaitable n'est alors plus de former des citoyens libres et conscients d'eux-mêmes. Il est « défini en termes de performances scolaires des élèves [...] plus précisément, ce sont des résultats mesurables qui constituent l'opérationnalisation de ce souhaitable » (p. 64).

Il est pourtant nécessaire, pour comprendre la mise en œuvre et les buts de l'évaluation dans le domaine éducatif, d'en saisir les usages par le politique. C'est le propos de Mons et Crahay qui analysent les modes de mise en œuvre et d'utilisation de l'évaluation des résultats des élèves à des tests standardisés dans différents pays de l'OCDE. La question est alors de savoir comment le politique utilise ces résultats d'évaluation. Ils aboutissent au constat d'un écart très fort entre « d'un côté, la banalisation de la rhétorique de la responsabilisation et de la culture de l'évaluation et, de l'autre, la faiblesse des dispositifs d'évaluation des politiques éducatives

utilisant les résultats quantitatifs de l'évaluation standardisée des élèves » (p. 91). Comment expliquer ce décalage entre volonté affichée et réalité des usages ? Dans beaucoup de pays, nous disent les auteurs, les résultats des évaluations standardisées sont utilisés comme instruments de communication en direction des citoyens et non comme un véritable outil permettant un bilan des politiques scolaires conduites dans une période donnée. Ces résultats servent surtout, « à déporter les responsabilités de l'action publique du niveau central vers le local, qui devient ainsi l'objet et l'acteur privilégié de cette évaluation » (p. 94).

Autant dire que les usages de l'évaluation sont à repenser dans le cadre plus large des contraintes de l'action publique dans les sociétés démocratiques. C'est ce que font, à partir d'une étude de cas sur la taille des classes, Bressoux et Lima. Ils constatent d'abord que le ministère français de l'Éducation promet, pour la rentrée 2010, une augmentation de la taille des classes à l'école primaire en arguant que les « expériences les plus récentes indiquent que la diminution des effectifs dans les classes n'a pas d'effet avéré sur les résultats des élèves » (p. 99). Or, les auteurs montrent à partir d'une revue de la littérature et d'une analyse d'une enquête conduite en première année de primaire par le ministère lui-même, que la taille des classes a un effet significatif sur les acquis des élèves et que cet effet est loin d'être négligeable : un élève en moins par classe entraîne un gain de 2,2 % d'écart type. De plus, ce gain est d'autant plus fort que les élèves sont de milieu défavorisé. Dans ce cas donc, les résultats « scientifiques » de l'évaluation ne sont pas utilisés pour eux-mêmes et pour améliorer l'efficacité de l'enseignement, mais détournés, voire inversés, pour légitimer une politique d'économie de moyens. Cela pousse les auteurs à conclure qu'« il n'est pas interdit de penser [...] que les gains économiques engrangés maintenant [par l'augmentation de la taille des classes] ne se traduiront pas par des pertes ultérieures, dans ce [...] domaine économique qu'on pensait préserver » (p. 122).

L'usage de l'évaluation par le politique relève, on le voit, de visées plus complexes et diversifiées que ne le laisse penser le discours « officiel » sur les nécessités du pilotage et « l'efficacité » des systèmes éducatifs. Mais l'on peut aussi se centrer sur des usages plus localisés de l'évaluation, notamment au niveau des établissements. C'est le propos de Barrère qui se questionne sur les effets de la généralisation des pratiques d'évaluation sur le travail de chefs d'établissement. Comment ces évaluations sont-elles vécues et utilisées dans un cadre local ? Car l'évaluation, avant d'être une pratique de gouvernance, est d'abord une activité quotidienne qui relève du fonctionnement des établissements et constitue « un ensemble de tâches et de problèmes de plus en plus présents » (p. 138) pour les chefs d'établissement. Il ressort de l'enquête ethnographique de l'auteur que l'évaluation est, au quotidien, vécue en fonction de tensions très fortes par les acteurs de l'éducation. À la fois « postbureaucratique » car valorisant

la mise en œuvre de moyens propres, elle renforce simultanément le pôle proprement administratif du travail des chefs d'établissement. Elle oriente aussi les pratiques et les politiques au bénéfice de ce qui est « chiffrable » par rapport à ce qui ne l'est pas. « Si l'évaluation invite à rendre visible des résultats de l'action, cela veut-il dire que ce qui n'est pas chiffrable n'a pas de valeur ? » (p. 138). En ce sens, elle constitue un instrument de régulation des politiques d'établissement que les acteurs locaux s'approprient en vue de poursuivre leurs buts propres.

L'évaluation a donc des effets notables sur le métier de chef d'établissement que Cattonar et Lessard étudient à partir d'une vaste enquête portant sur un échantillon représentatif de chefs d'établissement au Canada. Ils proposent d'étudier comment se restructure ce métier au regard des récentes évolutions du mode de régulation scolaire centré sur un pilotage du système éducatif par les résultats. Dans un tel contexte, les chefs d'établissement deviennent des personnages clés de ces politiques dès lors que « l'établissement est [...] devenu une "unité de reddition de comptes", tenue publiquement responsable de la performance des élèves » (p. 145). En devenant des « leaders pédagogiques » plus que des leaders administratifs, les chefs d'établissement voient leur rôle et leur métier se transformer au plan quantitatif (leur charge de travail augmente) mais aussi au plan qualitatif (leurs activités changent) et au plan subjectif (leur satisfaction au travail augmente). « La régulation par les résultats semble ainsi renforcer le rôle des directions en tant qu'acteurs intermédiaires entre les différents paliers et acteurs scolaires (enseignants, parents, communauté, autorités) » (p. 159).

Les contributions à cet ouvrage qui traitent des chefs d'établissement montrent donc que les modes d'appropriation de l'évaluation par les acteurs sont déterminants pour comprendre ses effets sur les systèmes éducatifs. Cette appropriation peut se faire, comme dans le cas de l'Observatoire du fonctionnement de l'enseignement primaire étudié par Solaux, Huguenin, Payet & Ramirez, par l'ensemble des acteurs de l'éducation, des parents aux enseignants et aux chefs d'établissement. Pour les auteurs, partie prenante de l'Observatoire, il s'agit de montrer comment articuler entre elles les logiques politique, scientifique et démocratique. Et cette articulation ne va pas de soi, tant il est vrai que la figure de l'expert domine trop souvent les procédures évaluatives, au point d'en constituer la figure idéale typique. Dans le cas de l'Observatoire, « il a été possible [...] d'apporter des informations relevant du champ scientifique aux acteurs sociaux en vue de les aider à déterminer des choix ou des pistes d'action pour mieux vivre l'école » (p. 166). Notons que ce lien entre résultats d'analyses scientifiques et délibération démocratique n'est pas le fruit du hasard ou du bon vouloir des « experts ». Ce processus est inscrit dans le mandat même de l'Observatoire qui a suscité un long processus de concertation dans la perspective d'améliorer l'efficacité et l'équité de l'enseignement primaire genevois. Il a permis notamment la « révision

de la mécanique de l'allocation des ressources en personnel administratif pour qu'elle soit fondée sur le nombre d'élèves et leurs caractéristiques socio-économiques » (p. 180).

Les usages de l'évaluation ne se limitent pas à l'objectivation du fonctionnement des institutions éducatives. On peut aussi utiliser les évaluations standardisées des élèves pour identifier, comme le fait Ramirez, les biais liés à la façon dont les enseignants évaluent leurs élèves. Dans cette perspective, l'auteur se questionne sur les apports d'une externalisation de l'évaluation des élèves dans le secondaire en comparant les notes moyennes obtenues par les élèves du secondaire obligatoire à Genève et leurs résultats aux épreuves communes du canton. « La question principale est de savoir si l'usage des évaluations externes [...] [pour évaluer les élèves] peut renforcer l'équité sociale » (p. 186) car ces évaluations – anonymes et corrigées « à l'aveugle » – sont par nature non affectées par les effets d'attentes des enseignants et les stéréotypes. Les résultats donnent à voir les avantages qu'il y aurait à externaliser l'évaluation des élèves car cela éviterait le poids très lourd « de l'interaction enseignant-élève sur les inégalités observées en termes de performance scolaire, au même titre que les effets de contexte » (p. 198). Autrement dit, une externalisation de l'évaluation des élèves aurait pour conséquence de limiter les inégalités scolaires liées à la filière et aux caractéristiques sociodémographiques des élèves; en un mot, cela renforcerait l'équité du système éducatif à Genève.

Enfin, la contribution de Falter traite de la relation entre les compétences mesurées par PISA et les transitions scolaires vers le secondaire supérieur¹¹ et l'enseignement tertiaire. Le suivi longitudinal des jeunes Suisses ayant passé les tests PISA en 2000 est de ce point de vue très informatif. L'auteur constate notamment que « la formation professionnelle de haut niveau attire des individus relativement similaires à ceux choisissant la formation gymnasiale. À l'inverse, la formation professionnelle de niveau inférieur est clairement distincte » et Falter s'interroge : « Pourquoi des individus aux aptitudes similaires suivent-ils des voies différentes ? » (p. 205). Falter s'intéresse également à la manière dont les compétences de base sont valorisées ultérieurement sur le marché du travail, à partir des données suisses de l'enquête ALL (*Adult Literacy and Lifeskills Survey*). L'auteur montre que si les compétences de base sont explicatives de l'insertion professionnelle et du niveau de salaire, elles sont loin de tout expliquer. Et ceci pour au moins une raison, liée à l'importance des compétences dites « non cognitives »

11. En Suisse, le secondaire supérieur comporte deux orientations : une voie de formation générale dite gymnasiale qui mène à la maturité (baccalauréat) et à des diplômes de culture générale, et une voie professionnelle initiale qui conduit à l'obtention de diplômes professionnels (certificats fédéraux de capacité). En 2008, 69% des jeunes en formation secondaire supérieure fréquentaient la voie professionnelle initiale (source : <http://www.bfs.admin.ch/bfs/portal/fr/index/themen/15/04/00/blank/uebersicht.html>)

(estime de soi, contrôle personnel, etc.) qui ne sont pas mesurées dans les évaluations internationales. De plus, l'une des limites des évaluations actuelles est de ne considérer que des compétences acquises par le biais d'une formation générale. « Il existe donc une nécessité de fournir un instrument en adéquation avec la formation professionnelle » (p. 215).

En définitive, l'existence même de cet ouvrage montre que nous possédons aujourd'hui assez de recul pour comprendre les usages et les conséquences de l'évaluation non pas seulement *in abstracto*, mais de façon concrète et empirique. Et chaque contribution donne à voir tout ce qui sépare d'un côté le modèle idéal des politiques scolaires dont la régulation et la transparence seraient assurées par des comptes rendus objectifs et réguliers sur leur fonctionnement, et d'un autre côté la réalité concrète des usages contextualisés de l'évaluation qui en changent nettement le visage et révèlent des situations tout autres. C'est donc à travailler à ce nécessaire désenchantement du « gouvernement par les nombres » que nous consacrons les pages qui suivent.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Baudelot, C. & Establet, R. (1971). *L'école capitaliste en France*. Paris: Maspero.
- Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago.
- Bourdieu, P. (1984). *Homo Academicus*. Paris: Minit.
- Bourdieu, P. & Passeron, J.-C. (1964). *Les héritiers. Les étudiants et la culture*. Paris: Minit.
- Bourdieu, P. & Passeron, J.-C. (1970). *La Reproduction : éléments pour une théorie du système d'enseignement*. Paris: Minit.
- Carnoy, M. (1995). The Benefits of Education. Introduction. In M. Carnoy (Ed.), *International Encyclopedia of Economics of Education* (pp. 113-125). Cambridge, UK: Pergamon.
- Centre international d'études pédagogiques. (2010). Palmarès et classements en éducation. *Revue internationale d'éducation de Sèvres*, 54, 29-164.
- Chubb, J. & Moe, T. (1988). Politics, Markets, and the Organization of Schools. *The American Political Science Review*, 82(4), 1065-1087.
- Couppié, T., Gasquet, C. & Lopez, A. (2007). *Quand la carrière commence... les sept premières années de vie active de la génération 98*. Marseille: Cereq.
- Demailly, L. (2006). En Europe: l'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. *Éducation et Sociétés*, 17(1), 105-118.
- Denison, E. F. (1962). *The sources of economic growth in the United States and the alternatives before us* (Committee for Economic

- Development. Supplementary Paper 13). New York : Committee for Economic Development.
- Desrosières, A. (2008). *Gouverner par les nombres*. Paris : Presses de l'École des Mines.
- Duran, P. (1999). *Penser l'action publique*. Paris : L.G.D.J.
- Duran, P. (2010). L'évaluation des politiques publiques : une résistible obligation. *Revue française des affaires sociales*, 1-2, 5-24.
- Durkheim, E. (1938). *L'évolution pédagogique en France*. Paris : PUF.
- Falter, J.-M., Pasche, C. & Hertig, P. (2007). *Compétences, formation et marché du travail en Suisse. Une exploitation des résultats de l'enquête internationale sur les compétences des adultes (ALL)*. Neuchâtel : Office fédéral de la statistique.
- Felouzis, G. (2005). Les indicateurs de performance des lycées : le marché scolaire fait des différences. *Revue française de sociologie*, 46(1), 3-36.
- Felouzis, G. & Perroton, J. (2007a). Les « marchés scolaires ». Une analyse en termes d'économie de la qualité. *Revue française de sociologie*, (48)4, 693-722.
- Felouzis, G. & Perroton, J. (2007b). Repenser les effets d'établissement. Marchés scolaires et mobilisation. *Revue française de pédagogie*, 159, 103-118.
- Flückiger, Y., Meunier, M. & Falter, J.-M. (2009). Capital humain, de l'école au marché du travail. In M. Oris, E. Widmer, A. de Ribaupierre, D. Joye, D. Spini, G. Labouvie-Vief & J.-M. Falter (Éd.), *Transitions dans le parcours de vie et construction des inégalités* (pp. 169-184). Lausanne : Presses polytechniques et universitaires romandes.
- Fridenson, P. (2009). La multiplication des classements de revues de sciences sociales. *Le mouvement social*, 226(1), 3-14.
- Gingras, Y. (2008). Du mauvais usage de faux indicateurs. *Revue d'Histoire Moderne et Contemporaine*, 5(55-4bis), 67-79.
- Grek, S. (2009). Governing by numbers : the Pisa « effect » in Europe. *Journal of Education Policy*, 24(11), 23-37.
- Grin, F. (1994). *L'économie de l'éducation et l'évaluation des systèmes de formation. Rapport de tendances*. Berne : FNRS, Programme national de recherche 33.
- Hanhart, S. (2007). Les indicateurs de dépenses des systèmes d'enseignement : de la comparaison des ressources engagées à la mesure de l'efficacité. *Éducateur SER, numéro spécial*, 29-32.
- Hanushek, E. A. & Raymond, M. E. (2005). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Jarousse, J.-P. (1991). L'économie de l'éducation du « capital humain » à l'évaluation des processus et des systèmes éducatifs. *Perspectives documentaires en éducation*, 1991(23), 79-105.

- Kane, T. J. & Staiger, D. O. (2002). The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives*, 16(4), 91-114.
- Labov, W. (1993). Peut-on combattre l'illettrisme? *Actes de la recherche en sciences sociales*, 100, 37-50.
- La fièvre de l'évaluation. Présentation. (2008). *Revue d'Histoire Moderne et Contemporaine*, 55-4bis, 5-6.
- Langouët, G. (2008). *50 ans d'école, et demain?* Paris: Fabert.
- Lascoumes, P. & Le Galès, P. (Éd.). (2004). *Gouverner par les instruments*. Paris: Presses de Sciences Po.
- Le Galès, P. & Scott, A. (2008). Une révolution bureaucratique britannique? Autonomie sans contrôle ou « freer markets, more rules ». *Revue française de sociologie*, 49(2), 301-330.
- Levin, H. M. (1988). Cost-Effectiveness and Educational Policy. *Educational Evaluation and Policy Analysis*, 10(1), 51-69.
- Mandl, U., Dierx, A. & Ilzkovitz, F. (2008). *The effectiveness and efficiency of public spending* (No. 301). Brussels: European Commission.
- Maroy, C. (2006). *École, régulation et marché. Une analyse de six espaces scolaires locaux en Europe*. Paris: PUF.
- Maroy, C. (2008). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? *Sociologie et sociétés*, 40(1), 31-55.
- Maroy, C. & Van Zanten, A. (2007). Régulation et compétition entre établissements scolaires dans six espaces locaux en Europe. *Sociologie du travail*, 49, 464-478.
- Merle, P. (2010). Structure et dynamique de la ségrégation scolaire dans les collèges parisiens. *Revue française de pédagogie*, 170, 73-85.
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *The Journal of Political Economy*, 66(4), 281-302.
- Ministère de l'Éducation nationale, Direction de l'évaluation, de la prospective et de la performance (MEN-DEPP). (2008). Lire, écrire, compter: les performances des élèves de CM2 à vingt ans d'intervalle 1987-1997. *Note d'information 08.38*.
- Mintzberg, H. (1982). *Structure et dynamique des organisations*. Paris: Éditions d'Organisation.
- Mons, N. (2007). *Les nouvelles politiques éducatives. La France fait-elle les bons choix?* Paris: PUF.
- Moretti, E. (2005). *Social Returns to Human Capital*. (Research Summary) NBER Reporter. Consulté le 15 février 2011 dans <http://www.nber.org/reporter/spring05/moretti.html>
- Moscovici, S. (1961). *La psychanalyse, son image, son public*. Paris: PUF.
- Mueller, N. (2007). *(Mis-)Understanding Education Externalities* (MPRA Paper No. 5331). Muenchen. Consulté le 9 février 2011 dans <http://mpra.ub.uni-muenchen.de/5331/>
- Musselin, C. (2001). *La longue marche des universités françaises*. Paris: PUF.

- Office fédéral de la statistique, Conférence suisse des directeurs cantonaux de l'instruction publique (OFS-CDIP). (2005). *Pisa 2003 : Compétences pour l'avenir*. Berne.
- Organisation de coopération et de développement économiques. (1997). *Regards sur l'éducation. Les indicateurs de l'OCDE 1997*. Paris.
- Organisation de coopération et de développement économiques. (2007). *Pisa 2006. Les compétences en sciences, un atout pour réussir* (vol. 1). Paris.
- Organisation de coopération et de développement économiques. (2010). *Regards sur l'éducation. Les indicateurs de l'OCDE 2010*. Paris.
- Organisation for Economic Co-operation and Development. (2010). *PISA 2009 Results : What Makes a School Successful ? Resources, Policies and Practices* (vol. 4). Paris.
- Piéron, H. (1963). *Examens et docimologie*. Paris : PUF.
- Piketty, T. & Valdenaire, M. (2006). L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. *Les dossiers évaluations et statistiques, MEN-DEPP, 173*.
- Sall, H. M. (1996). *Efficacité et équité de l'enseignement supérieur. Quels étudiants réussissent à l'Université de Dakar? Tome 1. Cadre théorique et méthodologique*. Dakar : Université Cheikh Anta Diop.
- Van Zanten, A. (2009). *Choisir son école. Stratégies familiales et médiations locales*. Paris : PUF.
- Wolter, C. S., Denzler, S., Evéquo, G., Hanhart, S., Nussbaum, O., Ragni, T., Schläfli, A. & Weber, B. (2003). *Nachfrageorientierte Finanzierung in der Weiterbildung* (Trendbericht Nr 7). Aarau : Schweizerische Koordinationsstelle für Bildungsfragen.

Réformes et ordres universitaires locaux¹

Catherine Paradeise* & Jean-Claude Thoenig**

***Université Paris-Est Marne-la-Vallée, LATTs et IFRIS²**

****Université Paris-Dauphine, DRM³**

La formation et la recherche universitaires sont-elles aujourd'hui soumises à une dynamique irréversible de globalisation? Les spécificités locales seront-elles bientôt écrasées par le rouleau compresseur des standards uniformisants? Les institutions locales et nationales d'enseignement supérieur, en premier lieu universitaires, sont-elles contraintes de s'aligner sur une stratégie unique sinon une vision uniforme de leur devenir, et condamnées à abdiquer toute capacité propre de définir un projet alternatif?

Un vaste mouvement de rationalisation s'est mis en route, où l'international tire le national, que ce soit à travers des évaluations et classements privés ou publics des universités, des revues, des laboratoires ou des diplômes. Les outils de mesure émanent d'institutions locales (classement de Shanghai, *Centrum für Hochschulentwicklung*), d'organes de presse écrite (*Business Week*, *Financial Times*, etc.), d'agences nationales d'évaluation (AERES, Agence d'évaluation de la recherche et de l'enseignement supérieur, en France), d'associations internationales d'accréditation (EQUIS, *European Quality Improvement System*), d'organisations internationales (OCDE). Ils produisent une *soft law* sous la forme de dispositifs de mesure diffusant largement par imitation. Le caractère effectivement obligatoire de ces dispositifs à adhésion volontaire, qui échappent de plus en plus souvent au contrôle

1. Cet article a bénéficié du soutien de l'Agence nationale de la recherche (ANR) dans le cadre du projet en cours PrestEnce (du Prestige à l'Excellence. La fabrication de la qualité académique).

2. Respectivement « Laboratoire techniques, territoires et sociétés » et « Institut francilien recherche innovation et société ».

3. Dauphine recherches en management.

d'opportunité et de qualité des autorités publiques, dépend du degré de partage de la croyance en leur valeur comme signal de qualité.

Une véritable onde de choc saisit les enseignants et les chercheurs. Pour autant, est-il certain que cette standardisation relative des critères de jugement conduise inexorablement à la convergence organisationnelle? De nombreuses recherches récentes entreprises sur le monde académique montrent que le doute est permis (Bleiklie, 2008; Musselin, 2003, 2005; Paradeise, Reale, Goastellec & Bleiklie, 2009). Même si l'économie globale des réformes de l'enseignement supérieur en Europe occidentale telles qu'on peut les lire ex post leur est commune, la précocité, la rapidité, les modalités de leur introduction comme leurs formes d'appropriation sont très variables d'un pays à l'autre. Quoi de commun entre la mise en place précoce, systématique, brutale de dispositifs liant incitation, mesure de la performance et rétribution tels qu'ils sont élaborés en Grande-Bretagne dès le milieu des années 1980, et l'incrémentalisme disjoint qui préside en France ou en Suisse aux évolutions jusqu'au milieu des années 2000? Dans chaque pays, les institutions académiques ont produit et assimilé les réformes récentes à leur manière, en fonction des rapports de force consolidés par l'histoire entre centres et périphéries, acteurs professionnels et managériaux, disciplines, en se saisissant de l'offre de réforme comme de ressources d'action nouvelles ou au contraire les subissant comme des contraintes mutilantes. Au final, si partout le paysage connaît d'indéniables transformations, il donne l'image d'une forte diversité intra et internationale, bien éloignée de la vision homogénéisante affichée par la thèse de la convergence.

Nous tenterons ici d'élaborer une typologie de cette diversité des formes locales d'appropriation des nouvelles normes d'action, en nous appuyant sur la prémisse de Merton (1973) selon laquelle la qualité académique repose sur l'appui réciproque de dimensions instrumentales et honorifiques. Depuis les années où ce dernier écrivait ces mots, les dimensions honorifiques se sont démultipliées en deux formes très différenciées de jugement, l'une contextuelle et synthétique que nous nommons *prestige*, l'autre acontextuelle et analytique que nous nommons *excellence*.

Nous dégageons quatre types idéaux d'établissements, tous percutés par les mesures d'excellence, en fonction de la manière dont ils combinent l'attention qu'ils portent en valeur au prestige et à l'excellence. Nous cherchons ensuite à appairer chacun de ces types avec les dimensions instrumentales qui le caractérisent.

LES JUGEMENTS DE QUALITÉ PERCUTÉS PAR « L'EXCELLENCE »

Le développement de la mesure de « l'excellence »⁴ à l'aide de standards d'évaluation de la production de recherche et de formation pourrait faire oublier qu'il ne s'agit là que de l'un des modes possibles de qualification de la valeur. Les jugements de qualité qu'il produit ne sont pas superposables à ceux qu'engendrent les mécanismes producteurs de la réputation et du prestige (Eymard-Duverney & Marchal, 1997 ; Karpik, 1989, 1996 ; Musselin & Paradeise, 2002).

L'excellence s'équipe d'outils analytiques qui appuient une économie du jugement fondée sur des quasi-prix. Les jugements de prestige reposent sur des évaluations diffuses, synthétiques, souvent implicites. Contrairement aux jugements d'excellence, les jugements de prestige ne pénètrent pas la matière de ces boîtes noires que sont les institutions de formation et de recherche. Ils s'alimentent aux idées véhiculées au sein de mondes sociaux (et pour partie peu vérifiées), quant à la sélectivité du recrutement, la qualité du personnel académique et de la formation, le placement des étudiants sur le marché du travail, etc. Ils sont fortement influencés par l'histoire et la géographie qui ont fait « la cote » des établissements. Leur ancrage dans des représentations situées explique une bonne part de l'inertie de ces jugements réputationnels au niveau local. C'est ainsi qu'en France, le prestige comparé des grandes écoles et des universités est à peu près insensible à leurs évolutions respectives, et fait fonction de représentation autoréalisatrice. Il explique aussi la difficulté à comparer les institutions ou les personnes entre elles puisque le jugement est cardinal, non seulement contextuel, et porte aussi sur des caractéristiques plus ascriptives qu'acquisitives. C'est tout le contraire avec les jugements d'excellence, qui disposent les entités évaluées sur des échelles ordinales, voire numériques qui, offrant un horizon acontextuel de référence, affranchissent l'évaluation de la qualité de la médiation des réseaux sociaux locaux ou nationaux. Le résultat de la mesure fait foi et loi, en offrant en particulier une solution instrumentalement rationnelle et impersonnelle au problème de l'arbitrage sur les choix de formation des étudiants, les choix de carrière des universitaires, les choix d'allocation des ressources par les tutelles, etc. Le caractère analytique des jugements d'excellence en fait des outils de prédilection des nouvelles rationalisations de la gouvernance et du pilotage des établissements (Porter, 1995 ; Whitley & Gläser, 2007).

Le contraste et l'éventuelle complémentarité entre les deux formes d'évaluation, par l'excellence et par le prestige, méritent donc attention puisque le recours généralisé à l'étalonnage sur divers indicateurs oriente de plus en plus le destin des vainqueurs et des vaincus de la compétition pour les

4. Que nous ne considérons pas comme une propriété intrinsèque mais plutôt comme un « symbole condensateur » de l'esprit des réformes.

ressources. En effet, les champions des dispositifs d'excellence partagent une conception de l'enseignement supérieur et de la recherche qui recouvre une théorie de l'action bien typée, qui véhicule une façon radicalement nouvelle de penser et de gouverner les politiques d'enseignement supérieur et de recherche, de récompenser le succès et de sanctionner l'échec des établissements, de leurs départements et centres de recherche. Ils entendent, pour le meilleur et pour le pire, abolir les images de marque et les rentes que ne corrobore pas la mesure de l'excellence, pour leur substituer des jugements bien calés sur des performances strictement identifiées et ouvertes à la comparaison internationale. Les outils de mesure de l'excellence visent ainsi une rupture fondamentale au regard des schèmes encore dominants de la qualité dans la plupart des pays d'Europe occidentale.

UN STATUT THÉORIQUE DU LOCAL

Rompre avec le sens commun n'est pas chose aisée, a fortiori quand le milieu qui le véhicule est celui des enseignants et des chercheurs. À écouter leurs craintes et critiques, la partie serait déjà jouée sinon perdue. Le monde académique se trouverait pris dans ce mouvement de globalisation par la standardisation qui affecte déjà puissamment d'autres domaines de l'activité économique et publique. Le démontreraient amplement les travaux de sciences sociales qui rendent compte du recours massif à des instruments et indicateurs de jugement, de pilotage et de certification par les entreprises comme par les agences publiques (Brunsson & Jacobsson, 2000; Le Galès & Lascoumes, 2004; Power, 1997).

Le choix interprétatif macrodéterministe qui sous-tend cette approche veut que la globalisation se nourrisse de dispositifs d'incitation et de contrôle à distance, définis de façon hétéronome par des organismes tiers et surplombant les acteurs locaux, conduisant toutes les parties prenantes d'un même domaine d'activité à uniformément se référer aux mêmes standards normatifs et cognitifs, engendrant ainsi à terme plus ou moins court des pratiques identiques sur tous les terrains locaux. Cette véritable cage de fer imposerait donc un seul modèle à qui voudrait conquérir ou conserver une qualité qui le distingue des autres acteurs de son secteur. Pour espérer être reconnue et accéder aux ressources qui en découlent, chaque université devrait copier « la meilleure », celle qui est classée en tête de tel ou tel classement.

Une telle approche s'inscrit largement dans la posture de la théorie institutionnelle en sociologie des organisations, par exemple lorsqu'elle interprète la proximité des formes et cadres d'action adoptés par des entreprises ou agences publiques opérant dans un même secteur ou marché et concurrentes sur les mêmes sources de financement par des effets d'isomorphisme organisationnel (DiMaggio & Powell, 1983). Pour espérer survivre, il

faut savoir s'adapter aux exigences implicites formulées par les incitations à suivre le modèle des « meilleurs ». La différenciation locale serait mortelle. Le rouleau compresseur de la globalisation serait à l'œuvre.

Pour rompre avec la vision du tout global, il faut réhabiliter le local et faire de la diversité des comportements des organisations universitaires une question de recherche, une hypothèse à tester empiriquement sans préjuger des trajectoires institutionnelles. En d'autres termes, il faut se doter d'une lunette théorique adéquate pour analyser comment les institutions locales se positionnent dans les faits et dans leur quotidien par rapport au global, comme agents passifs soumis *de facto* aux injonctions d'un principal exogène ou comme acteurs plus ou moins autonomes mobilisant sous contrainte des ressources de leur environnement au service d'un horizon qui leur est propre. Dans cette perspective, deux apports théoriques majeurs s'imposent.

Le premier est fourni par le fondateur de la sociologie des sciences, lorsqu'il s'intéresse aux universités et à leurs composantes comme autant d'acteurs poursuivant des stratégies et des politiques. Dans un célèbre article resté programmatique, Merton (1960/1973) invite à distinguer deux dimensions de la qualité académique des institutions universitaires : une dimension instrumentale et une dimension honorifique.

Par dimension instrumentale, il désigne la fabrique des sources de la qualité, c'est-à-dire la façon dont les établissements académiques produisent et gèrent les occasions qui permettent l'éclosion et la concrétisation de leurs potentialités. Ce qui conduit à plusieurs questions. Quels types d'environnement humain et de fonctionnement organisationnel permettent-ils une créativité socialement valorisée, notamment en termes d'inventivité scientifique et de formation ? Quels ensembles de qualités, quelles ressources, quels positionnements en termes d'activités, quelles fonctions de production, quels dispositifs organisationnels adoptent-ils et valorisent-ils dans leur gestion en tant qu'établissement, y compris dans celles de leurs composantes ?

Par dimension honorifique, Merton désigne les processus de valorisation de l'activité dans les contextes institutionnels et marchands où s'enchaînent les institutions universitaires. Il fait du *prestige* le critère honorifique de référence, celui qui permet que le talent d'une institution soit honoré, récompensé matériellement et symboliquement, et la compétence de ses membres identifiée. On préférera parler ici de *qualité*, afin d'éviter de réduire la reconnaissance honorifique au seul jugement de prestige qui occupe tout l'espace des jugements de valeur sur les établissements au moment où écrit Merton. À côté des jugements de *prestige* proprement dit, mode dominant sinon hégémonique jusqu'à la fin du XX^e siècle, le début du XXI^e marque en effet l'émergence des jugements d'*excellence* évoqués plus haut. L'identification de la dimension honorifique de la qualité nourrit des questions importantes. Quelles parties prenantes, notamment externes, définissent les sources de valeur et de qualité pour telle université et ses composantes ?

Quels types de performances y font référence et justifient les politiques et les actes concrets qu'adoptent leurs responsables et leurs membres ?

Le second apport provient de la sociologie des organisations. Une université est une entité pluraliste composée de composantes différenciées et dont l'action concrète reflète des dynamiques de pouvoir. Pour comprendre la performance productive des départements, March (1962) développe le concept d'ordre local, qui lui permet d'élucider le pilotage et les fonctionnements des collectifs de production et de diffusion du savoir dans les universités américaines. Il s'attache ainsi à cette dimension instrumentale qui caractérise le travail d'organisation des ressources (De Terssac, 1992) plutôt qu'aux incitations en tant que telles. Il regarde les composantes des universités comme des espaces spécifiques, gérant au quotidien des tensions entre les différentes logiques d'action auxquelles elles sont soumises et construisant leurs ressources en tirant variablement parti d'environnements eux-mêmes diversifiés (Serow, 2000).

L'hypothèse centrale d'une telle approche est que les arrangements organisationnels concrets construisent les ressources qui affectent les modalités et les niveaux de la performance. Réciproquement, la transformation cognitive majeure (Michaud & Thoenig, 2009) engendrée par les nouveaux dispositifs de jugement de qualité impersonnels et désencastrés affecte le travail et les arrangements organisationnels (Cret, 2007 ; Djelic, 2006).

Traiter le local comme un espace majeur pour comprendre la construction de l'ordre social permet de rompre avec une variante de la théorie du tout global ou du prêt-à-porter des standards. Celle-ci laisse entendre que, lorsque le local ne s'aligne pas sur les injonctions du global, c'est qu'il est victime de contraintes issues de son passé, qu'il reste plombé par les pesanteurs de son environnement immédiat, bref qu'il résiste de façon irrationnelle parce que son comportement est contraint par sa dépendance de sentier. Au contraire, nous formulons l'hypothèse selon laquelle le local peut, dans des circonstances données, intégrer des changements découlant de la standardisation globale tout en les accommodant à ses arrangements organisationnels propres, aux cognitions et aux valeurs qu'il mobilise pour agir et décider, aux critères de succès dont ses membres sont les porteurs et les interprètes.

Les ordres locaux sont donc construits par l'action, et lisibles à travers l'articulation des décisions à divers niveaux des systèmes d'enseignement supérieur, les uns ne s'alignant pas simplement et mécaniquement sur les autres. Ils s'expriment diversement selon les combinaisons d'horizons que valorisent les acteurs : servir la communauté locale, alimenter le marché du travail national, se classer comme une référence scientifique internationale, persévérer simplement dans leur être, etc. Ils mobilisent les deux facettes complémentaires et distinctes de la qualité académique, sans que l'une entretienne nécessairement des rapports de corrélation simples avec l'autre.

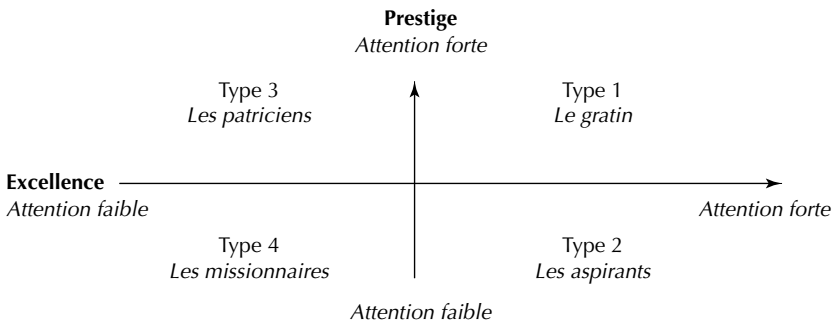
UNE TYPOLOGIE DE LA PRODUCTION LOCALE DE LA QUALITÉ : QUATRE POSITIONNEMENTS FACE AU PRESTIGE ET À L'EXCELLENCE

De cette double conceptualisation en termes d'ordres locaux et de dispositifs de jugement de la qualité résultent des questions importantes pour la pratique, qui n'ont pourtant pas fait à notre connaissance l'objet de programmes empiriques.

Quel degré de recouvrement rencontre-t-on aujourd'hui entre évaluation de prestige et évaluation d'excellence, et comment ce recouvrement varie-t-il à travers pays et disciplines ? Certaines qualités instrumentales sont-elles plus importantes que d'autres pour assurer la valorisation des départements ? Comment se concilient les configurations des dimensions instrumentales et des dimensions honorifiques de la qualité, sur des registres aussi divers que la scène internationale de la reconnaissance par la recherche, les marchés du travail nationaux ou les contributions à la communauté locale ?

En croisant ces deux dimensions de la dimension honorifique de la qualité, on identifie quatre types.

Figure 1 : Typologie des établissements en fonction de l'attention qu'ils portent aux deux dimensions honorifiques de la qualité



On examinera d'abord comment les institutions universitaires se distribuent aujourd'hui au carrefour du prestige et de l'excellence, comment s'y articulent la qualité qu'ils reconnaissent aux contenus qu'elles produisent de façon endogène et la qualité que leur reconnaissent des publics tiers. Nous chercherons ensuite à comprendre comment les dispositifs de jugement de prestige et d'excellence agissent sur les conditions instrumentales de la qualité, et réciproquement.

Le gratin

Il est fait d'établissements de grand prestige international, qui se classent aussi au plus haut niveau sur les indicateurs d'excellence. Ils portent une grande attention à leur prestige comme à leur excellence. Leurs membres font référence comme élite académique. Ils servent de modèle à leurs concurrents.

Reconnus parmi les meilleurs dans leur genre, faisant référence nationalement et internationalement dans leur domaine, ils s'adaptent rapidement aux évolutions nationales et internationales des critères de jugement de qualité et aux transformations des publics qui comptent, en jouant sur les tableaux de la recherche fondamentale et appliquée, de la formation à divers niveaux, de la valorisation désintéressée et marchande de leurs produits. Leur positionnement en tête varie peu d'un classement international d'excellence à l'autre, quels que soient les échelles ou les critères de jugement employés. Leur leadership paraît protégé par une sorte de rente pérenne de situation. Tout à la fois prestigieux et excellents, ils représentent les parangons de la vertu académique.

Ils semblent ne pas avoir à fournir d'efforts exceptionnels pour se maintenir au sommet en termes honorifiques quand apparaissent de nouvelles formes de valorisation en la matière. Pourtant, s'ils valorisent leurs performances auprès de parties prenantes externes qui comptent, ils n'en sont pas moins attentifs à la façon dont ils fabriquent et entretiennent de façon endogène les sources de qualité qui sont au fondement de leur prestige comme de leur excellence. S'ils le peuvent, c'est qu'ils bénéficient de la durée de diverses ressources matérielles et institutionnelles nécessaires à la consolidation de leur qualité instrumentale interne. Ils peuvent se consacrer largement à cultiver et à améliorer leurs arrangements instrumentaux internes déjà excellents par ailleurs. Ils construisent ainsi un cercle vertueux qui voit leur qualité instrumentale renforcer leur qualité honorifique. Pourtant, s'ils répondent aux doubles exigences du prestige et de l'excellence standardisée en matière de produits, qu'il s'agisse de leurs publications, de leurs formations, de leur environnement, ils n'ont pas de caractéristiques uniformes. Chacun possède ses singularités.

Les aspirants

Sous le label d'*aspirants*, nous désignons des établissements qui jouissent d'un réel prestige local ou national au moment où ils sont percutés par l'excellence. Ils découvrent avec effroi que ce prestige se dilue lorsqu'on sort des cadres sociaux nationaux, car ses modes d'expression ne sont pas congruents avec les outils d'identification internationale de l'excellence. Par exemple, ils accordent un grand prix à la formation et au lien avec les

professionnels, mais pas à la publication. Ou bien ils valorisent les manuels, essais et ouvrages de réflexion, mais pas particulièrement les articles publiés dans des revues dites internationales. La comparaison sur les standards classiques les fait tout simplement disparaître du tableau des élus pour diverses raisons : une trop petite taille, une insuffisante concentration sur la publication, la faible lisibilité de leur offre, un corps enseignant trop peu cosmopolite, une faible attractivité internationale, etc.

Ils ambitionnent de transmuier rapidement leur capital de prestige en excellence internationale. Pour entrer dans la cour des grands, qui doit les sauver d'être les perdants d'une concurrence globalisée, les aspirants cherchent à aligner leurs stratégies sur les exigences d'une bonne évaluation analytique en termes d'excellence. Ils concentrent ambitions et ressources sur leurs performances sur la base des standards établis par des classements en vue : proportion de non locaux ou non nationaux dans les rangs du personnel académique et des étudiants, parité des genres, publications dans les revues de rang alpha, taux et niveaux de placement des anciens étudiants sur le marché du travail, etc. Cet objectif les conduit par exemple à préférer les publiants sur le *mercato* académique international et à développer leurs incitations internes par des primes à la publication labellisée de haut niveau.

Contrairement aux établissements installés dans le gratin, les aspirants s'engagent dans un effort radical de recomposition. Leur stratégie rompt assez brutalement avec leur passé. En centrant leur attention sur l'excellence, ils s'intéressent plus à la labellisation externe de leurs produits qu'à leurs orientations en termes de contenus. Ils sous-traitent ainsi leur politique de recherche à des instances tierces comme les revues les mieux classées ou les institutions de financement les plus influentes dont ils adoptent les orientations thématiques et les critères de choix. Ils développent en quelque sorte une économie de cueillette. Quant à la formation qu'ils offrent, ils s'alignent sur les établissements du gratin.

Cette quête à marche forcée, qui requiert un leadership autoritaire, est développée sans trop de considération de ses effets sur l'établissement comme institution. Elle peut en ébranler gravement l'*affectio societatis*, en déroutant un corps académique ancien dont les états de service se trouvent déclassés par les priorités nouvelles de l'excellence : la multiplicité des tâches qu'ils ont assumées comme bons citoyens de l'institution – gestion administrative, création de programmes d'enseignement, relations avec les entreprises et les anciens étudiants, etc. – est disqualifiée par l'exigence de publication. L'organisation se trouve ainsi placée en face d'un dilemme entre poursuite de l'excellence au risque de saper les bases du prestige, et culture du prestige au risque d'une lenteur excessive dans la conquête d'une excellence indispensable à la survie.

On trouve en outre dans cette catégorie des établissements missionnaires, peu visibles sur les radars du prestige ou de l'excellence, mais à qui la montée en puissance de la « croyance dans les nombres » (Porter, 1995) ouvre l'opportunité de chercher à se faire valoir dans la nouvelle compétition pour le classement. À l'image des précédents, ils sont tentés de développer des politiques volontaristes tirées par les indicateurs qui jugeront de leur valeur d'excellence.

Les patriciens

Les *patriciens* jouissent d'une bonne assise locale en termes de prestige. Ils sont réticents à jouer le jeu de l'excellence, qu'ils jugent absurde au regard de la singularité des institutions académiques. Au contraire des nouveaux riches que sont les aspirants qui veulent reconquérir leur valeur honorifique dans un espace élargi en convertissant leur prestige en excellence, les patriciens se comportent comme une vieille aristocratie dont le prestige ne fait qu'exprimer une qualité intrinsèque établie par l'histoire et précieusement conservée par la sagesse du corps des pairs. Les critères exogènes qui fondent les comparaisons entre établissements ne leur sont certes pas inconnus. Pourtant ils y restent noblement indifférents voire ouvertement hostiles. Ils les jugent infondés épistémologiquement en même temps qu'attentatoires à leur propre intégrité institutionnelle.

Aux manies comptables des non-initiés – journalistes, bureaucrates, institutions internationales, etc. – ils opposent le capital d'un prestige fondé sur la préservation collégiale de modes de production d'une qualité qu'ils jugent intangible. Les rites d'adoubement des nouveaux membres dans le club exclusif que forme la communauté académique où les patriciens reconnaissent les leurs et dont ils tiennent les allogènes à distance, assurent donc la coexistence pacifique et la coopération entre égaux plutôt que la concurrence caractéristique du monde des aspirants. Fondés sur des affinités électives entre élites, dédaigneux du tout-venant académique, sûrs de la valeur intrinsèque de leurs produits – publications, formations, diplômés – et de la vulgarité de la concurrence, les patriciens s'interrogent peu sur la pertinence de leurs contenus pour leurs publics. Au contraire des aspirants qui cherchent à coller à la demande d'excellence, ils sont résolument attachés à une économie d'offre dont ils se jugent seuls qualifiés à maîtriser la qualité.

Les patriciens cherchent à résister aux perturbations induites par des politiques qui pourraient affecter leur prééminence traditionnelle dans leur sphère d'action, et en particulier à l'introduction d'outils analytiques de performance qui pourraient introduire un doute sur leur valeur honorifique et affaiblir la capacité de soutien de leurs réseaux sociaux.

Les missionnaires

Les *missionnaires* disqualifient l'idée même de prestige et dénoncent la dangerosité du principe d'excellence. Ils ont une vision égalitaire d'une mission de service public assurée par des institutions formées de personnels soumis aux mêmes statuts et réglementations et offrant des prestations de même nature dans un esprit désintéressé. Le prestige leur est donc une valeur étrangère. Quant à la quête d'excellence sur critères impersonnels et acontextuels, elle ne peut selon eux qu'exacerber une concurrence coûteuse pour des bénéficiaires sociaux contestables, qui ne peuvent que creuser les inégalités et nuire à la mission intégrative de l'éducation. Les missionnaires imputent la variabilité sociale de la valeur honorifique des établissements au jeu de variables exogènes comme le lieu d'implantation ou la sélectivité sociale des recrutements, qui affectent tout à la fois l'image de l'établissement et les performances de ses étudiants, et qui ne dit donc rien des qualités intrinsèques de leur activité.

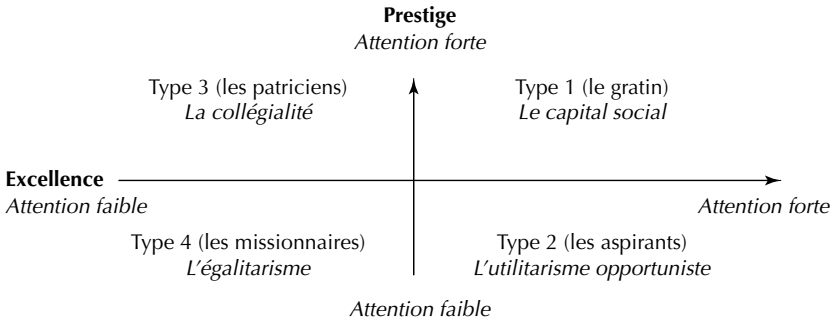
Les missionnaires se font les chantres des vertus de la continuité du service public d'enseignement supérieur. Ils refusent de considérer les tâches de formation visant l'insertion professionnelle de jeunes sur leur marché du travail local comme un métier spécifique ou comme le « sale boulot » d'un monde universitaire dont la noblesse se consacrerait à la recherche. Il importe donc de n'en dégrader aucune par des outils de mesure qui creuseraient les écarts entre ressources affectées à des missions de service public non hiérarchisables. Ils se montrent en conséquence réticents à des dispositifs d'évaluation et de financement appuyés sur l'évaluation hétéronome de leur production, qui ne rendraient pas justice à la variabilité des missions des universitaires selon les publics qu'ils recrutent et auxquels ils veillent à ajuster leur offre.

LES APPARIEMENTS ENTRE DIMENSIONS HONORIFIQUES ET INSTRUMENTALES DE LA QUALITÉ

Au concret, les positionnements des établissements en termes de qualité s'associent-ils à des modes particuliers de fonctionnement? Les arrangements internes qui les gouvernent diffèrent-ils selon le type d'attention qu'ils portent et que leur environnement physique et institutionnel leur permet de porter aux dimensions de l'excellence et du prestige? L'hypothèse est plausible et testable. Tels outils de gouvernance interne peuvent soutenir ou faire obstacle à tel positionnement stratégique, permettre ou empêcher l'établissement d'allouer son attention à telle ou telle dimension honorifique de la qualité. Les organisations sont tyranniques, car ce sont des ensembles sociaux et humains lestés de fortes inerties, chargés de jeux politiques et émotionnels entre acteurs, inscrits dans des postures cognitives, culturelles

et normatives qui leur donnent une sensibilité et une réactivité inégale aux modifications des environnements dans lesquels elles opèrent.

Figure 2: Typologie de l'instrumentation organisationnelle des établissements en fonction de l'attention qu'ils portent aux deux dimensions honorifiques de la qualité



Le capital social

Les établissements du gratin associent une attention stratégique forte à la poursuite de l'excellence et du prestige. Leur mode de fonctionnement les rapproche de la bureaucratie organique identifiée par Burns et Stalker (1961). Néanmoins ils s'en distinguent par un trait: en termes d'instrumentation organisationnelle, leur capital social leur permet de jouer à un très haut niveau d'une tension productive entre les sphères apparemment opposées du professionnel et de l'administratif d'une part et de l'individuel et du collectif d'autre part.

Leur rationalisation administrative est certes poussée. Elle fait appel à un outillage gestionnaire moderne et étendu qui mobilise des métiers, des *process* et des procédures allant de la communication externe aux achats, en passant par l'administration des programmes de formation et de recherche, le contrôle financier ou encore la recherche de financements auprès de publics divers, anciens élèves, organismes publics nationaux et internationaux, entreprises.

Pour autant la rationalisation administrative n'impose pas sa chape de plomb à une logique d'action de type professionnel, bien au contraire. Elle entretient le court terme et les affaires courantes sans pour autant marginaliser la veille et l'anticipation stratégiques de moyen terme. La gouvernance légitime et garantit la légitimité prioritaire de la sphère académique, de ses représentants, de ses références en termes de qualité et d'idéaux. Du coup les normes professionnelles sont explicitement intégrées dans l'organisation bien qu'elles y soient exogènes, et s'expriment sur des scènes de

débat à tous les niveaux et sur la plupart des thématiques qui comptent. Ces contre-feux évitent des travers communs à la bureaucratie comme la centralisation effrénée ou les effets de silos et permettent de maîtriser le risque que l'horizon de gestion de l'organisation échappe au contrôle de ses clients internes. Le recours à des procédures et l'exigence de conformité aux règles n'empêchent pas l'esprit d'entreprise, cela à tous les étages. Des normes partagées et souvent implicites régulent les marges de jeux et le contenu des rôles entre le centre et la base, entre la sphère administrative et la sphère académique. Il existe une régulation sociale légitime par des règles qui fondent une connaissance partagée et un espace d'interprétation commun des situations. On sait jusqu'où aller sans aller trop loin, on partage un même langage pour agir, on raisonne de la même manière face aux problèmes, on s'accorde mutuellement confiance.

Une forte pression à jouer collectif s'exerce en outre sur les composantes, départements, centres de recherche, membres du corps académique. En effet, l'appartenance de ses composantes à une institution du gratin leur offre la ressource considérable d'une marque notoire. Mais elle induit en retour une obligation de résultat, qui, au-delà des obligations formelles, repose sur la perception partagée du devoir moral de contribuer au bien collectif, de l'enrichir et de le cogérer avec les autres membres de la communauté locale d'appartenance qu'est l'établissement. La mutualisation est entretenue par des règles en usage qui font sens pour le corps académique, qu'elle concerne l'allocation de subventions de recherche, la création de chaires ou la promotion de sa réputation.

L'instrumentation conforte les valeurs collectives par ses multiples déclinaisons au niveau micro-organisationnel. Elles multiplient en interne les occasions d'interaction sociales sans considération de statut formel – pratiques de thé collectif, tours de piste des candidats à un poste, séminaires ouverts, etc. Elles encouragent le développement de communautés denses d'intérêts, notamment entre disciplines différentes. Des recouvrements partiels et des redondances entre composantes tempèrent les effets d'évitement mutuel qui naîtraient d'une emprise trop forte de silos disciplinaires ou de l'appropriation exclusive de thématiques. Les initiatives de création de nouveaux départements ou centres de recherche sont considérées comme normales et saines, et non pas comme de mauvaises manières cachant de sombres intérêts de pouvoir et de mandarinat. Le souci de la carrière future des étudiants fait partie de la contribution normale de chaque enseignant. L'irrévérence est acceptée et même récompensée, pourvu qu'elle soit jugée constructive dans les débats internes. Plus généralement, la pression à jouer collectivement est au cœur de la socialisation. Le capital social est un bien commun dont la préservation et l'expansion dictent des devoirs qui poussent les composantes individuelles à ne pas se comporter comme des passagers clandestins.

L'utilitarisme opportuniste

Les aspirants portent une attention prioritaire à l'excellence en recourant à un type d'instrumentation organisationnelle dont le vecteur, l'utilitarisme, vise à aligner leurs composantes sur cette nouvelle conception de la qualité. Elle guide les comportements de ses membres par des stimuli financiers directs ou indirects, qui encouragent les contributions individualisées aux objectifs d'excellence définis par des instances d'évaluation et de classification tierces, la performance collective se résumant à la somme des productions individuelles labellisées. L'engagement de *knowledge workers* est réduit aux acquêts. Ils s'achètent sur le marché et y retournent lorsque l'échange marchand est jugé insatisfaisant par l'établissement – performance insuffisante – ou par l'universitaire – capacité de négociateur de meilleures conditions salariales. Car aucune loyauté n'est requise hors des clauses du contrat de travail. L'activité de recherche vise à améliorer un score individuel et collectif qui définit la valeur de marché de l'établissement et de ceux qui y contribuent. La finalité de connaissance importe moins que la cote du support de publication. L'instrumentation comptabilise ex post, elle ne stimule pas l'inventivité ex ante. L'exploitation de la science normale prévaut sur l'activité d'exploration. L'interdisciplinarité est indiscipline, source de risque pour la productivité qui articule la relation entre managers et académiques. Les autres fonctions constitutives des missions traditionnelles de l'enseignement supérieur, comme l'enseignement et la gestion d'équipes ou de programmes, ne sont ni valorisées en termes de récompenses ni recherchées par les académiques.

L'organisation peut-elle engendrer des solutions aptes à contrebalancer la déstructuration interne des finalités d'action, des types de production et des formes de coopération interne, et avec quel effet sur le prestige originel? En d'autres termes, ce modèle d'excellence est-il soutenable? Les aspirants peuvent-ils régénérer les contenus en détruisant les bases instrumentales qui ont contribué à leur prestige en fidélisant leur communauté académique et en consolidant ses investissements dans l'invention de contenus et la participation aux processus endogènes de régulation? L'utilitarisme opportuniste repose sur le pari qu'une institution académique peut être gérée durablement comme un kit, en réduisant l'activité académique à un service marchand articulant intérêt individuel et bien collectif sur les comportements égoïstes de membres polarisés par la réussite matérielle à court terme. Les composantes intermédiaires entre la direction et les individus ne sont ni des instances autonomes de décision et de débat, ni des références identitaires ni des lieux de régulation sociale. Elles se contentent de déconcentrer une gestion administrative que le centre contrôle et subordonne à l'atteinte d'un rang plus élevé dans les classements d'excellence année après année. Les formes de sociabilité sont réduites aux acquêts, mieux elles importent peu. L'évitement mutuel entre disciplines ou départements, l'absence du bureau,

voire l'inscription de l'activité dans des institutions tierces sont acceptables pour autant que les produits d'excellence des résultats soit comptabilisables en interne.

L'utilitarisme opportuniste subordonne la sphère professionnelle à la sphère organisationnelle. Sa forme l'apparente au modèle d'organisation mécaniciste ou bureaucratique : forte spécialisation des tâches, standardisation prononcée des procédés, recours à des systèmes de planification, structure administrative élaborée (Burns & Stalker, 1961). Pour autant, bureaucratization ne signifie pas paralysie. Les aspirants bricolent des solutions au coup par coup à mesure que des déstructurations internes provoquées par une économie de cueillette affectent le tissu social et l'offre de l'établissement. Par exemple, dans les formations continues de haut niveau, en substituant aux ressources humaines permanentes appauvries par les incitations à la publication des enseignants à temps partiel eux-mêmes professionnels, qui peuvent y trouver des satisfactions d'image. La bureaucratization supporte une forte centralisation du pouvoir, un autoritarisme qui peut si nécessaire fouler aux pieds l'ethos académique. La procéduralisation construit les contraintes qui permettent d'imposer la marche forcée vers le nouvel horizon de l'excellence. La stratégie appartient à la seule autorité hiérarchique et est mise en œuvre de façon discrétionnaire si celle-ci le juge nécessaire.

La collégialité

Les patriciens portent une attention prioritaire et soutenue à leur prestige. Leur instrumentation organisationnelle emprunte beaucoup au modèle de la bureaucratie professionnelle (Mintzberg, 1982).

L'établissement est constitué comme une somme de parties autosuffisantes, confiées à la représentation bienveillante d'un *primus inter pares*, et distinctes de la ligne hiérarchique administrative. Les présidences d'établissements et les directions de composantes sont trop démunies de marges de jeu et de légitimité hiérarchique pour imposer sinon penser une ligne stratégique d'ensemble. Elles exercent leurs fonctions selon « un principe du moindre pouvoir » (Karpik, 1995) qui adoucit fortement la contrainte extrinsèque qu'exerce toute organisation sur ses membres, en s'appuyant sur la confiance implicite dans les relations interpersonnelles entre pairs. L'administration, organisée selon une ligne hiérarchique distincte, constitue une fonction subalterne parmi d'autres confiée à de fidèles et braves serviteurs. L'organisation est ainsi conçue comme un contenant mis au service du prestige de ses membres plutôt que comme un principe actif impulsant une dynamique collective.

Faute de finalités collectives partagées, prendre des initiatives dans l'organisation est suspect a priori. Débattre, mettre les cartes sur table,

expliciter des différends sur les horizons de l'institution ou sur l'organisation des moyens, négocier sont autant de mauvaises manières. La stratégie en acte résulte donc des mises en forme organisationnelles accumulées au fil de l'histoire – division du travail entre et au sein des composantes, procédures, passe-droits. Elle favorise une politique distributive qui tend à conserver les positions acquises. Les ressources sont allouées en fonction de la hiérarchie acquise du prestige et des statuts. Elles sont éventuellement complétées par les apports propres des universitaires au segment d'organisation où ils officient. La tyrannie des moyens et des traditions est assouplie par des arrangements locaux propres aux composantes. Piloter l'organisation, c'est donc d'abord gérer des incidents.

Le prestige des universitaires dérive d'abord du statut de leur discipline. La communauté de référence est essentiellement disciplinaire, supralocale ou cosmopolite, dépassant largement les frontières formelles de l'établissement ou d'une de ses composantes. C'est l'étendue et la qualité des réseaux professionnels hors de l'institution locale d'affectation plus que d'appartenance qui confèrent visibilité et puissance à l'institution. Cosmopolite, la collégialité est aussi corporatiste. Le prestige d'une communauté, donc celui de ses membres, découle en large partie du degré d'exclusivité de son contrôle sur la sélection, la formation, le placement et la carrière de ses membres tout au long de leur vie, donc de sa capacité à imposer des critères de distinction sociale et professionnelle dans son domaine nationalement ou internationalement.

En d'autres termes, l'institution locale externalise sa gestion des ressources humaines, sa politique scientifique, les définitions de la pertinence et de l'excellence vers des communautés professionnelles hors ses murs. Elle leur sert de structure d'accueil, d'agences locales de communautés pour l'essentiel gérées par « silo disciplinaire ». Chaque profession ou discipline est ainsi gouvernée par des critères foncièrement irréductibles les uns aux autres, qu'il s'agisse du type de recherche, de pédagogie ou de la régulation sociale entre membres. L'institution locale n'a d'autre choix que de faire confiance à la capacité des professionnels présents en son sein de fournir des ressources en termes d'image et de réputation sur des scènes extérieures plus vastes auxquelles elle n'a pas accès.

Cela ne signifie pas que la notoriété de ces communautés professionnelles soit indépendante du prestige des institutions existantes. Aussi veulent-elles prendre pied dans les établissements les plus prestigieux si elles en sont encore absentes, et, avec les ressources, en particulier politiques, dont elles disposent, résistent-elles aux évolutions qui pourraient affecter l'image des institutions locales les plus prestigieuses où elles sont établies.

La collégialité comme règle pour l'action facilite l'alliance de fait entre le local et le cosmopolite, entre l'établissement et la profession. Elle rend possible le gouvernement d'un groupe de pairs en important de l'extérieur

des normes de pouvoir légitime, par exemple l'ancienneté et le grade. Elle permet la coexistence de communautés professionnelles différentes en n'imposant pas de critères uniformes et rigides pour arbitrer entre elles.

L'égalitarisme

Les missionnaires accordent peu d'attention au prestige ou à l'excellence comme critère de pertinence de leur activité. D'autres priorités les mobilisent de fait, et elles ne sont pas nécessairement partagées de la même façon par tous, loin s'en faut.

Leur instrumentation partage à première vue plusieurs traits avec celle qu'adoptent les patriciens. L'organisation est subordonnée à la logique professionnelle. L'outillage gestionnaire reste peu investi et n'est guère valorisé. L'autorité hiérarchique est faible, car suspecte par principe. L'organisation fonctionne comme une juxtaposition de silos spécialisés chacun autour d'une fonction ou d'un domaine particulier qui différencient des composantes qui coopèrent peu entre elles de manière directe et spontanée. La coordination entre composantes mobilise de forts coûts de transaction, chacune tendant à ériger sa compétence ou son domaine d'activité en mission essentielle pour le collectif, en mobilisant des argumentations essentiellement normatives. Le centre peine à trancher clairement entre les missions, sauf à moduler à la marge. Tout changement stratégique est perçu comme un risque pour soi et pour l'institution. Comme chez les patriciens, il est très difficile voire impossible de redéployer des ressources à partir d'arbitrages stratégiques qui sont vécus comme des atteintes majeures aux valeurs pour l'action dont les professionnels sont les champions. Les membres tendent à exploiter les biens communs de façon opportuniste, sans trop de ménagement, et sont sujets à dégradation. L'objectif tacite acceptable par tous est de persévérer dans son être. Les composantes évoluent entre confiance et défiance dans un contexte de jeux politiques investis affectivement, qui produisent des formes de décision peu transparentes. Le collectif est centrifuge.

Néanmoins un fait majeur différencie l'instrumentation des missionnaires de la bureaucratie professionnelle qui est la caractéristique des ordres locaux des patriciens. Alors que ces derniers dérivent leur prestige de sources exogènes, tenant à l'appartenance de leurs membres à des professions fortement constituées à un niveau supralocal, les priorités des missionnaires sont formulées par des professionnels dont les références sont endogènes à l'institution locale. Les professionnels localistes tirent leur compétence d'apprentissage sur le tas et d'expériences mal transposables ailleurs, formateurs de savoir-faire ou de solutions ad hoc dans des réseaux fonctionnels locaux. Leur mobilité entre domaines ou tâches est limitée, ce même à l'intérieur d'un même établissement. Leur pouvoir tient à leur maîtrise des fonctions qui comptent localement pour telle ou telle mission.

La faiblesse du gouvernement collectif interdit qu'un arbitrage fort émerge spontanément et soit consensuellement toléré. Du coup, l'instrumentation accentue la déconnection de l'établissement, appauvrit son écoute des évolutions de la société, de la concurrence et de la demande de formation et de recherche, même si certaines de ses composantes sont enchâssées dans des réseaux dont la pertinence et le renouvellement sont parfois problématiques. À chacun son domaine d'action en interne, à chacun son réseau local plus ou moins actif, à chacun sa solution pour le collectif, à chacun son agenda. L'égalitarisme est un mode d'instrumentation qui alloue à chacun une même priorité de traitement, même si des différences subtiles ou informelles peuvent s'installer, cette allocation se faisant sans référence explicite à des critères de choix, comme si une priorité en valait bien une autre, le tout en respectant le maintien de niches routinières spécialisées auxquelles est consentie de fait une autonomie fonctionnelle poussée.

CONCLUSION

Les descriptions typiques que nous proposons doivent naturellement s'entendre comme « accentuations unilatérales de la réalité », c'est-à-dire qu'elles négligent les tensions entre types qui se développent au sein des établissements concrets. Elles négligent en particulier ce que ces tensions et leurs modes de résolution peuvent devoir, à l'intérieur d'un établissement, aux rapports de pouvoir entre disciplines, comme aux rapports entre les objectifs poursuivis par la présidence et les valeurs et normes prévalentes parmi les personnels académiques⁵. On peut cependant penser que chaque type tend à affecter les valeurs propres de ses membres, par socialisation interne, mais aussi, à travers les recrutements et mobilités diverses, par des appariements affinitaires entre types d'établissements et types d'universitaires. Ce phénomène ne peut que s'accroître avec le développement de stratégies de ressources humaines actives au sein d'établissements dont les moyens d'attirer à eux des universitaires conformes à leurs valeurs et objectifs se sont accrus.

Reste à savoir comment chacun de ces types se comportera en dynamique. Les patriciens et les missionnaires seront-ils condamnés à la disparition dans les limbes de la médiocrité? Les aspirants parviendront-ils à surmonter les obstacles sur le chemin vers le gratin? Ces derniers parviendront-ils à conserver la double préoccupation qui fonde leurs *capabilités*? On peut penser que, du poids et de la forme des incitations à l'*excellence* en particulier, dépendra la capacité des divers types que nous avons décrits à résister à leur alignement au long d'une échelle unidimensionnelle.

5. C'est en particulier l'objet du projet ANR PrestEnce que d'explorer ces dimensions.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Bleiklie, I. (2008, septembre). *Excellence and the diversity of higher education systems*. Texte présenté à la 21^e conférence annuelle de CHER, *Excellence and diversity in higher education. Meanings, goals and instruments*, Pavie.
- Brunsson, N. & Jacobsson, B. (2000). *A world of standards*. Oxford : Oxford University Press.
- Burns, T. & Stalker, G. M. (1961). *The Management of Innovation*. Oxford : Oxford University Press.
- Cret, B. (2007). *Émergence et concurrence des labels d'accréditation*. Thèse de doctorat, Sciences Po Paris.
- De Terssac, G. (1992). *Autonomie dans le travail*. Paris : PUF.
- DiMaggio, P. J. & Powell, W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147-160.
- Djelic, M.-L. (2006). Marketization: from intellectual agenda to global policy-making. In M.-L. Djelic & K. Sahlin-Anderssen (Ed.), *Transnational governance : institutional dynamics of regulation* (pp. 53-73). Cambridge : Cambridge University Press.
- Eymard-Duvernay, F. & Marchal, E. (1997). *Façons de recruter. Le jugement des compétences sur le marché du travail*. Paris : Métailié.
- Karpik, L. (1989). L'économie de la qualité. *Revue française de sociologie*, 30(2), 187-210.
- Karpik, L. (1995). *Les avocats. Entre l'État, le public et le marché. XIII^e-XX^e siècle*. Paris : Gallimard.
- Karpik, L. (1996). Dispositifs de confiance et engagements crédibles. *Sociologie du travail*, 38(4), 527-550.
- Le Galès, P. & Lascoumes, P. (Éd.). (2004). *Gouverner par les instruments*. Paris : Presses de Sciences Po.
- March, J. G. (1962). The Business Firm as a Political Coalition. *Journal of Politics*, 24, 662-678.
- Merton, R. K. (1973). *The Sociology of Science. Theoretical and Empirical Investigations*. Chicago : University of Chicago Press. (Première publication 'Recognition and excellence'. Instructive ambiguities. In A. Yarmolinsky (Ed.), *Recognition of Excellence* (pp. 419-438). New York : The Free Press, 1960).
- Michaud, C. & Thoenig, J.-C. (2009). *L'organisation et ses langages. Interpréter pour agir*. Québec : Presses de l'Université Laval.
- Mintzberg, H. (1982). *Structure et dynamique des organisations*. Paris : Éditions d'Organisation.
- Musselin, C. (2003). Editorial. *European Journal of Education*, 38, 333-339.
- Musselin, C. (2005). *Le marché des universitaires. France, Allemagne, États-Unis*. Paris : Presses de Sciences Po.

- Musselin, C. & Paradeise, C. (2002). Le concept de qualité: où en sommes-nous? *Sociologie du travail*, 44(2), 255-260.
- Paradeise, C., Reale, E., Goastellec, G. & Bleiklie, I. (2009). Universities Steering between Stories and History. In C. Paradeise, E. Ferlie, I. Bleiklie & E. Reale (Ed.), *University governance: Western European comparative perspectives* (pp. 227-246). Dordrecht: Springer Publishing Company.
- Porter, T. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Power, M. (1997). *The Audit Society*. Oxford: Oxford University Press.
- Serow, R. 2000. Research and teaching at a research university. *Higher Education*, 40, 449-463.
- Whitley, R. & Gläser, J. (Ed.). (2007). *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*. Dordrecht: Springer.

La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone: nouveau paradigme politique et médiation des experts

Christian Maroy* & Catherine Mangez**

***Universités de Montréal et de Louvain,
CRIFPE et GIRSEF¹**

****Université de Louvain, GIRSEF**

Les politiques éducatives sont souvent présentées comme des réponses à des problèmes auxquels sont confrontés les sociétés ou les systèmes scolaires. Ainsi par exemple, les élèves sortant sans qualification posent problème car leur insertion professionnelle et sociale semble compromise. Une telle appréhension des politiques éducatives met alors l'accent sur les différences d'intérêt et d'idéologie parmi les acteurs (politiques ou scolaires) participant à la formation des politiques, sur les négociations et rapports de force qui les incitent à adopter, de façon plus ou moins cohérente, telle ou telle solution. Les politiques sont des réponses aux problèmes mais peuvent varier en fonction des valeurs de la coalition politique dominante, des contraintes budgétaires, politiques ou institutionnelles dont elle doit tenir compte.

Loin de réduire le politique à la formulation de « réponses » aux problèmes, l'approche cognitive des politiques publiques (Muller, 1995, 2000) s'interroge sur la façon dont le politique (et plus largement la société) en

1. Respectivement « Centre de recherche interuniversitaire sur la formation et la profession enseignante » et « Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation ».

arrive à définir les questions et problèmes auxquels il doit faire face. Elle fait du politique un champ de débats et de controverses où s'élabore de façon conflictuelle une vision de la société ou d'un secteur de la société, de ses finalités ou de ses problèmes. Dans cette approche, les changements des politiques éducatives ne résultent pas seulement de changements factuels de l'école ou de son environnement social. Ils procèdent aussi des transformations des catégories cognitives ou normatives à partir desquelles la réalité sociale ou scolaire est appréhendée. Ainsi Muller ou Hall ont souligné le rôle des paradigmes politiques dans des domaines aussi différents que la politique économique (Hall & Taylor, 1993) ou aéronautique (Muller, 2000). Comme Muller l'avance, un référentiel² d'action publique est un processus de « construction d'un rapport au monde », un « mécanisme de fabrication d'images, d'idées, de valeurs qui vont constituer une vision du monde » (Muller, 1995, p. 157). Bref, c'est une matrice cognitive et normative qui tend à définir non seulement les buts de la politique et le genre d'instruments mobilisables pour les atteindre, mais également la nature même des problèmes que les décideurs politiques sont censés aborder. Par analogie avec les paradigmes scientifiques de Kuhn (1983), les paradigmes politiques tendent à être d'autant plus influents qu'ils deviennent hégémoniques et que les catégories cognitives et normatives qu'ils mobilisent prennent un caractère d'évidence.

Cet article propose une étude de la construction de la politique de pilotage et d'évaluation dans le secteur de l'enseignement en Belgique francophone, appréhendée à la lumière d'une telle approche cognitive. À partir du début des années 1990, on assiste à un développement important de dispositifs d'évaluation externe et de pilotage qui ne peut simplement être interprété comme une réponse fonctionnelle à des changements structurels internes et externes ou en relation avec l'évolution des équilibres de forces politiques en présence. Pour comprendre la discontinuité que cette politique introduit dans les politiques scolaires belges, il faut prendre la mesure du changement de paradigme de politique scolaire qui la sous-tend : l'école est considérée de plus en plus comme un système de production de compétences individuelles, dont il faut garantir l'efficacité et l'équité par un « pilotage » adéquat et la promotion de nouveaux modes de régulation basés sur le modèle de l'État évaluateur. Il faut, d'autre part, souligner le rôle d'acteurs institutionnels clés dans la coconstruction, l'importation et la traduction de ce paradigme dans les réalités institutionnelles et politiques de la Belgique³.

2. Nous utiliserons ici les termes de paradigme et de référentiel, de façon interchangeable, même si certaines nuances devraient être introduites.

3. Ce texte est dérivé d'un des rapports de la recherche européenne Know and Pol, conduite entre 2008 et 2009 par le Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation (GIRSEF) (voir Mangez, C., Maroy, Cattonar, Delvaux & Mangez, E., 2009). Sa méthodologie est basée sur des entretiens auprès de témoins privilégiés, mais

Nous présenterons d'abord le contexte du système scolaire belge avant de mettre en exergue l'orientation des politiques éducatives au cours des deux dernières décennies, en particulier en matière d'évaluation et de pilotage. Nous montrerons ensuite que cette formulation d'une réponse politique aux problèmes du système éducatif belge s'inscrit dans un paradigme du « système de production scolaire » qui n'était guère en usage dans les années 1980. Dans un quatrième temps, nous insisterons sur les conditions qui ont pu favoriser l'émergence et le succès de ce nouveau paradigme au sein du monde politique belge, en montrant tout à la fois le rôle de conditions structurelles « favorables » et celui d'entrepreneurs de changement.

UN SYSTÈME SCOLAIRE FRAGMENTÉ ET DÉCENTRALISÉ

Le système scolaire belge s'inscrit dans une démocratie qualifiée de consociative par les politologues, dans laquelle plusieurs institutions⁴ permettent à des communautés sociologiques différentes (sur le plan linguistique mais aussi sur un plan philosophique/religieux) de vivre ensemble tout en préservant, au moins en partie, leur autonomie d'existence collective (Dumont, 1999). Le système scolaire belge s'est ainsi construit sur un principe central de *liberté d'enseignement*, affirmé dès la création de l'État belge en 1830. Il s'agit en réalité d'une double liberté : celle d'organiser un enseignement et celle de choisir son école. Ce principe a rendu possible une initiative scolaire multiple et décentralisée. Divers pouvoirs organisateurs (PO) ou réseaux ont ainsi proposé un enseignement en bénéficiant, encore aujourd'hui, d'une forte autonomie⁵. Tout au long de l'existence de la Belgique, les politiques scolaires ont été dès lors le lieu de luttes et de compromis politiques qui se structurent essentiellement autour de la coexistence des différents réseaux d'enseignement et des mondes sociologiques et sociopolitiques⁶ qui les soutiennent (Deprez, 1970). Par comparaison avec d'autres systèmes et jusqu'à une période récente, les politiques scolaires portent au fond très peu sur le curriculum, la pédagogie ou les finalités de

aussi sur l'analyse d'un vaste corpus documentaire d'articles de presse et de documents scientifiques ou politiques.

4. Par exemple, un scrutin proportionnel (plutôt que majoritaire), l'octroi et le maintien d'une certaine autonomie aux acteurs collectifs organisés (dans le domaine de la santé, de l'éducation, du travail) et le financement par l'État des activités qu'ils organisent.

5. Ces PO sont structurés en trois réseaux : un réseau d'enseignement « libre » principalement catholique, un réseau officiel développé par des autorités publiques locales (les Communes, les Villes et les Provinces) et un réseau officiel organisé par l'État, c'est-à-dire la Communauté française, qui est donc simultanément un réseau et le pouvoir subsidiant. Les deux premiers réseaux sont subsidiés par la Communauté française mais ne sont pas organisés par elle.

6. On parle de « piliers » pour désigner ces mondes sociologiques sous-jacents aux institutions politiques.

l'éducation, car toutes ces dimensions sont pour l'essentiel des prérogatives des réseaux d'enseignement (Maroy & Dupriez, 2000). En effet, dans cette démocratie consociative et dans le système scolaire en particulier, des règles du jeu importantes structurent les modes de relations entre acteurs organisés. D'une part, la politique scolaire doit respecter l'autonomie des différents PO/réseaux (en matière de finalités éducatives, de programmes, etc.). D'autre part, un impératif de discrétion intersegmentaire s'impose à ceux-ci. Il faut (sauf crise) respecter l'autonomie de l'autre en ne cherchant pas à connaître ou à évaluer publiquement son mode de fonctionnement ou son efficacité. Autrement dit, la discrétion à l'égard des autres segments du système devra être préservée par les politiques scolaires, pour maintenir ce que Rémy appelait les « compromis de coexistence » entre « piliers » et réseaux constitutifs de la réalité scolaire et plus largement sociopolitique de la Belgique (Mangez, C. *et al.*, 2009 ; Mangez, E. 2009).

Mais ces réalités scolaires et politiques vont être érodées à la fois par des évolutions endogènes à la société belge (sécularisation et individualisation sociales) et des changements externes, que sont la globalisation économique et l'eupéanisation des politiques éducatives. Des « nouvelles » politiques scolaires en Communauté française de Belgique (CFB) se développent au cours des années 1990 qui, d'une part vont afficher des objectifs pédagogiques au niveau du système dans son ensemble (rognant ainsi l'autonomie des différents réseaux d'enseignement), d'autre part vont grignoter progressivement l'impératif de « discrétion » qui importait pour leur coexistence. Nous allons présenter l'essentiel de ces politiques et les enjeux qu'elles représentent, avant de souligner comment elles s'inscrivent dans un changement de paradigme politique.

DE NOUVELLES POLITIQUES SCOLAIRES DE PILOTAGE ET D'ÉVALUATION

Ce système scolaire très fragmenté et décentralisé fait l'objet de profondes réformes depuis le milieu des années 1990, notamment suite à une crise budgétaire et sociale qui accompagne la « communautarisation »⁷ de l'enseignement, sur laquelle nous reviendrons plus loin. Nous en présenterons ici les grandes orientations avant de faire une chronique plus détaillée de la construction des politiques d'évaluation et de pilotage.

7. En 1989, c'est-à-dire la décentralisation de l'essentiel des compétences d'enseignement de l'État fédéral belge aux entités fédérées (les Communautés linguistiques, notamment francophone).

Les orientations générales des politiques scolaires

En 1997, un décret central « sur les missions et structures de l'enseignement obligatoire » (appelé décret « missions »)⁸ va marquer la politique scolaire ultérieure. D'une part, c'est la première fois qu'un texte légal donne à tous les réseaux d'enseignement des missions communes, sur la base d'un débat social plus large que le débat parlementaire. D'autre part, ce décret organise « les structures propres à les atteindre » afin d'opérationnaliser ces missions et il contribue par là à redéfinir considérablement les modes de régulation du système scolaire. Ce décret va servir de référence à de nombreux décrets ultérieurs. Trois grandes orientations des politiques scolaires sont impulsées par ce décret (et ceux qui l'ont suivi) (Draelants, Dupriez & Maroy, 2003 ; Maroy & Dupriez, 2000).

DÉVELOPPEMENT DE RÈGLES CENTRALES ET TRANSVERSALES À TOUS LES RÉSEAUX D'ENSEIGNEMENT

Une première tendance est celle du renforcement de normes légales communes cadrant davantage le curriculum et les pratiques pédagogiques (mais aussi la formation initiale et continuée des enseignants, leur statut, etc.), ce qui a pour conséquence de rapprocher la situation des réseaux subventionnés de celle que connaît le réseau d'État. Ces normes communes se manifestent surtout dans la définition de référentiels pédagogiques et l'adoption d'un curriculum basé sur l'approche par compétences articulée autour d'une pédagogie constructiviste, différenciée, par cycles. Les politiques scolaires cherchent à rendre l'offre éducative plus homogène et à « harmoniser » les pratiques éducatives, jugées trop disparates et dénoncées comme une des sources des faibles performances du système notamment en termes d'inégalités (Crahay, 1997).

VALORISATION ET ENCADREMENT DE L'AUTONOMIE DES ÉTABLISSEMENTS

Néanmoins, cette tendance à la standardisation accrue est contrebalancée par une tendance a priori paradoxale : celle de « valoriser » l'initiative et le projet local, à travers la promotion d'une logique de participation au niveau des établissements. En CFB, comme ailleurs en Europe, l'établissement est donc valorisé et perçu comme un niveau d'action décisif, d'où l'attention qui sera donnée ensuite au rôle et à la formation des directeurs d'établissement (décret de 2007 sur le statut des directeurs)⁹, appelés à assumer une mission managériale et pédagogique.

8. On trouvera ce décret à l'adresse : http://www.galilex.cfwb.be/fr/leg_res_01.php?ncda=21557&referant=l01

9. On trouvera ce décret à l'adresse : http://www.galilex.cfwb.be/fr/leg_res_01.php?ncda=31886&referant=l01

DÉVELOPPEMENT DE L'ÉVALUATION ET DU PILOTAGE

L'orientation des politiques scolaires à la fin des années 1990 peut paraître paradoxale. D'une part, les grandes orientations du système scolaire sont de plus en plus établies au niveau central par la fixation d'objectifs à atteindre (notamment les compétences à maîtriser) et de stratégies à suivre pour y parvenir (notamment pédagogiques: structure en cycles, référentiels de compétences, évaluations externes communes, promotion de la pédagogie différenciée). D'autre part, l'autorité publique centrale valorise l'initiative et l'autonomie des établissements. Ce paradoxe apparent entre autonomie et contrôle disparaît cependant avec le développement ultérieur de divers dispositifs qui permettent de vérifier dans quelle mesure les objectifs sont atteints, ce qui suppose la mise en place d'un « pilotage » et de diverses formes d'évaluation externes, participant à ce que plusieurs auteurs ont appelé le modèle de la régulation par les résultats ou de l'État évaluateur¹⁰. Ces dispositifs vont surtout se développer dans les années 2000, mais dès le début des années 1990, l'idée germe et on assiste à un développement lent de dispositifs de pilotage qui s'institutionnalisent fortement dans la décennie suivante. Nous allons en retracer la chronologie, préalable indispensable pour fonder empiriquement notre interprétation ultérieure des conditions d'émergence de cette politique.

La construction lente du pilotage du système scolaire

Au départ, l'idée de piloter le système par ses résultats est suggérée par une étude de l'Organisation de coopération et de développement économiques (OCDE, 1993) sur l'état du système scolaire en Belgique, reprenant d'ailleurs des idées déjà lancées par certains chercheurs comme De Landsheere (1992). Cette étude, entamée en 1991 et publiée en 1993, constate que des lacunes et des freins empêchent un « pilotage plus rationnel » du système d'enseignement qui permettrait d'en améliorer la qualité et les résultats: absence d'objectifs généraux s'imposant à tous les établissements; absence de « culture d'évaluation »; services de statistiques trop peu développés; moyens financiers mal répartis; réseaux d'enseignement peu enclins à coopérer. Selon l'OCDE, ce n'est qu'en créant de nouveaux modes de

10. Dans ce modèle, l'État définit les objectifs du système en termes de résultats ou de compétences à atteindre aux différents niveaux d'enseignement, accorde de l'autonomie aux établissements pour définir les moyens de les atteindre, et enfin met en place des dispositifs d'évaluation des résultats des élèves par des tests externes, ou des audits d'établissements portant sur les pratiques pédagogiques ou questionnaires au sein des établissements. Ces nouvelles formes de régulation permettent ainsi de mettre en place une « responsabilisation » de la part des équipes locales au regard des objectifs fixés par les autorités, que les auteurs anglo-saxons décrivent comme des nouvelles formes d'*accountability* (Broadfoot, 2000; Maroy, 2006).

régulation, et notamment des instruments d'évaluation externe, que les résultats du système s'amélioreront.

C'est notamment sur cette base qu'est créée en 1993 une « structure légère de pilotage » qui met en place diverses enquêtes sur les acquis des élèves. En 1997, le décret « missions » reprendra le terme « pilotage » pour désigner les groupes de travail inter-réseaux chargés de définir les référentiels de compétences à atteindre¹¹. On notera cependant que l'évaluation externe est « oubliée » dans ce décret: tout se passe comme si on envisageait un pilotage à vue, sans instruments d'évaluation externe, personne ne semblant vouloir prendre la responsabilité politique d'instaurer des mécanismes de contrôle très étrangers au fonctionnement habituel des institutions scolaires belges (en raison notamment du principe de discrétion, évoqué plus haut).

La publication fin 2001 de l'enquête PISA 2000 et son retentissement médiatique ont cependant contribué à remettre le problème de l'efficacité de l'école au centre du débat (Cattonar, Mangez, E., Delvaux, Mangez, C. & Maroy, 2009). Parallèlement, des experts ont souligné que le diagnostic le plus alarmant sur l'enseignement en Communauté française était la très grande dispersion des résultats selon les élèves, c'est-à-dire l'inégalité du système.

Une proposition de décret a alors été discutée puis votée pour créer un « dispositif de pilotage » du système (décret du 27 mars 2002)¹². Une commission de pilotage « inter-réseaux » a ainsi été instituée, qui tend à faire des services de l'inspection et de l'administration, mais aussi des représentants des réseaux, des organisations syndicales, des parents, les acteurs clés de la mise en œuvre et du suivi du pilotage: mise en œuvre d'indicateurs du système, d'évaluations externes, de suivi statistique des élèves, remise

11. Ce sont les socles de compétences, les compétences terminales et les profils de formation et de qualification qui constituent autant d'objectifs d'apprentissage pour les différents niveaux de la scolarité. Ces référentiels sont produits à partir de 1998 par des groupes de travail inter-réseaux, mis en place suite au décret « missions ». Ce sont les premiers standards d'apprentissage, les premiers objectifs pédagogiques que les acteurs du système éducatif fixent communément et par décret. Le travail d'élaboration de ces référentiels est complexe et laborieux. Au niveau des socles de compétences (balises à 12 ans), le travail a duré de janvier à novembre 1998. Ce travail a pris appui sur un document provisoire daté de 1994 qui proposait des socles de compétences provisoires en français et en mathématiques, et sur le décret « école de la réussite » de 1995 (décret relatif à la promotion d'une école de la réussite dans l'enseignement fondamental du 14/03/1995) qui définissait des socles de compétences. Un décret de 1999 (décret portant confirmation des socles de compétences du 26/04/1999) a approuvé ces socles. Le travail de définition des compétences terminales (balises à la fin de l'enseignement obligatoire, 18 ans) s'est organisé entre 1999 et 2005 et a été confirmé par plusieurs décrets. La définition des profils de formation et de qualification concernant l'enseignement technique et professionnel est quant à elle toujours en cours.

12. On trouvera ce décret à l'adresse: http://www.gallilex.cfwb.be/fr/leg_res_01.php?ncda=26663&referant=l02

d'avis au gouvernement sur l'état du système et sur les orientations de la formation continue, etc.

Si à partir de 1994, des épreuves externes à valeur diagnostique ont été organisées périodiquement à l'initiative de la structure légère de pilotage, elles étaient facultatives. Depuis le décret du 2 juin 2006¹³, tous les élèves de 2^e et 5^e années primaires et de 2^e et 4^e années secondaires¹⁴ sont soumis obligatoirement à une épreuve externe. Ces épreuves sont supposées contribuer à la « régulation du système » et apporter un feed-back aux enseignants (sur le niveau de leurs élèves) assorti de pistes didactiques à mobiliser en classe. Ce même décret institutionnalise une épreuve certificative, le Certificat d'étude de base (CEB), en fin de 6^e primaire. Cette épreuve innove dans la mesure où elle est obligatoire et commune à tous les réseaux, et introduit un lien direct entre la réussite et l'obtention du CEB. Par la suite, un troisième décret (2009)¹⁵ a encore renforcé le dispositif d'évaluation externe des acquis des élèves de l'enseignement obligatoire.

Enfin, le service d'inspection de l'administration, important instrument de pilotage et d'évaluation, a été profondément réformé en 2007. Depuis le début des années 1990, les acteurs impliqués dans la réflexion autour du pilotage du système éducatif font en effet référence à une nécessaire réforme de l'inspection. Le projet était déjà dans les cartons du gouvernement de 1992, mais il mettra plus de 15 ans à se concrétiser (Maroy, 2009).

Pour bien comprendre la portée de ce nouveau texte législatif¹⁶, il faut dire un mot du système préexistant. Jusqu'en 2007, l'inspection est un corps de fonctionnaires dépendant de l'administration, qui intervient principalement pour établir une évaluation individuelle des compétences pédagogiques des enseignants, à la demande du chef d'établissement ou du PO. Par ailleurs, l'inspection est chargée d'un travail de vérification administrative. Le nouveau décret introduit trois nouveautés principales :

13. Décret relatif à l'évaluation externe des acquis des élèves de l'enseignement obligatoire et au certificat d'études de base au terme de l'enseignement primaire.

14. Ce décret instaure une épreuve externe commune certificative à la fin de la 2^e secondaire dont la participation est libre. Par ailleurs, il propose un « test d'enseignement secondaire supérieur » (TESS) qui vise à « accompagner les établissements scolaires dans le processus d'évaluation certificative au terme de l'enseignement secondaire ». Une épreuve externe en 4^e secondaire est prévue par décret mais n'a en réalité pas encore eu lieu faute d'avoir réussi à construire cette épreuve. Vu la disparité des filières et niveaux d'enseignement, il est difficile de produire une épreuve commune à tous les élèves à ce niveau-là.

15. Décret du 30 avril 2009.

16. Il s'agit du décret de la Communauté française de Belgique du 8 mars 2007 relatif au Service général de l'inspection, au Service de conseil et de soutien pédagogiques de l'enseignement organisé par la Communauté française (*Moniteur* n° 167 du 5 juin 2007, p. 30319).

- Le recentrage de l'inspection sur des questions pédagogiques. Le contrôle administratif des établissements est pris en charge par un autre service de l'administration. La nature des missions des inspecteurs est désormais essentiellement « l'évaluation et le contrôle du niveau des études » (d'un cycle ou de plusieurs cycles d'un établissement scolaire) en référence aux compétences définies comme objectifs pédagogiques¹⁷.
- Le caractère « inter-réseaux » de l'inspection : alors que les inspecteurs du secondaire provenaient exclusivement du corps enseignant du réseau de l'État, l'inspection scolaire recrutera désormais sur la base des mêmes critères dans tous les réseaux.
- Des « services ou des cellules de conseil et de soutien pédagogique » sont créés dans chaque réseau, reconnaissant ainsi des services qui existaient plus ou moins fortement dans les différents réseaux (Delvaux, Giraldo & Maroy, 2005). Ces services auront notamment pour fonction de « conseiller et d'accompagner les enseignants, les équipes pédagogiques et les écoles pour lesquels le service général d'inspection a constaté des faiblesses et des manquements »¹⁸, en tenant compte d'une « note d'information » rédigée par les services d'inspection ou des résultats obtenus aux évaluations externes. Ces services devront donc collaborer avec les services d'inspection, mais seront hiérarchiquement subordonnés aux réseaux¹⁹.

Si on met l'ensemble de ces réformes en perspective, force est de constater qu'on assiste à la mise en place progressive d'un État évaluateur : l'État, avec les responsables des réseaux ou les organisations syndicales, installe un nouveau mode de régulation du système dont un élément important sera constitué par les évaluations externes. Ces politiques scolaires ne peuvent cependant être simplement comprises comme des réponses à des évolutions de l'environnement (certes bien présentes). En effet, elles procèdent d'une nouvelle lecture de cet environnement et de la réalité de l'école, connectée à la montée en puissance d'un nouveau paradigme de politique scolaire, considérant l'école comme « un système de production » plutôt que comme une institution.

17. Ils sont en outre chargés de vérifier le respect des programmes ou la cohérence des options pédagogiques prises par l'école en regard des options pédagogiques prescrites par les décrets. Ils sont aussi chargés de missions relatives au pilotage central du système : appui aux évaluations externes, à la Commission de pilotage, etc.

18. Article 17 du décret du 8 mars 2007 précité, p. 30287.

19. La réforme de 2007 définit donc clairement deux missions : la mission de l'inspecteur et celle du conseiller pédagogique, qui sont assurées par des personnes différentes. Si la réforme a mis tellement de temps pour aboutir, c'est notamment en raison de l'opposition des inspecteurs qui voyaient la redéfinition de leurs missions d'un mauvais œil. En effet, les inspecteurs ont depuis toujours eu une mission de contrôle, mais pour certains niveaux d'enseignement, celle-ci s'accompagnait d'une mission de soutien, de conseil. Cela est maintenant « interdit » par le nouveau décret, en tout cas dans l'enseignement subventionné.

Après avoir décrit plus en détail ce changement de paradigme, nous montrerons que la montée du nouveau paradigme a été favorisée par deux types de conditions : 1) des changements structurels de l'environnement politique et socio-économique qui ont rendu l'ancien paradigme de plus en plus fragile et décalé ; 2) la présence d'un certain nombre d'entrepreneurs de changement, dont les propriétés et l'alliance ont favorisé le basculement de la vision de l'école et des politiques scolaires en Belgique et qui ont contribué à importer et aménager le nouveau paradigme, dont l'origine et l'impact dépassent clairement le cadre belge.

UN CHANGEMENT DE PARADIGME POLITIQUE : DE L'INSTITUTION AU SYSTÈME DE PRODUCTION SCOLAIRE

Notre thèse est que les nouvelles politiques dérivent d'un changement de paradigme de politique scolaire : l'école est de moins en moins définie comme instance de socialisation et comme « institution » et tend à être considérée comme « système de production scolaire », producteur de compétences et savoirs utiles pour les individus qui composent la société. Autrement dit, les politiques de même que les acteurs de base du système scolaire tendraient à penser le système scolaire non plus dans les termes du droit et de la sociologie (institution), mais dans les termes de la théorie systémique des organisations (l'école comme système de transformation d'*inputs* en *outputs*) et de l'économie (fonction de production du capital humain).

En Belgique francophone en particulier, le système scolaire est progressivement considéré par les politiques moins comme une institution plurielle, organisée en divers réseaux, au service de fonctions de socialisation importantes pour des collectivités clairement identifiées (société catholique, société laïque) que comme une organisation, « un système » qu'il convient de « piloter », de « gouverner » de la façon la plus efficace possible.

Dans le paradigme de l'institution, l'école est supposée assurer la socialisation des êtres sociaux, constituer simultanément – et parfois de façon contradictoire – des citoyens, des travailleurs et des personnes, selon les principales fonctions relevées par Durkheim (1922). Dans la démocratie consociative belge, contrairement à ce que l'on constate dans l'État républicain français (Dubet, 2002), plusieurs programmes institutionnels, correspondant à différents réseaux, se sont toujours concurrencés, tout en établissant de façon dynamique différents types de compromis de coexistence²⁰.

20. On peut l'illustrer de façon exemplaire à partir d'un article de synthèse de Van Haecht (1985) dans lequel elle présente « les noyaux durs des projets éducatifs » qui, tout au long de l'histoire scolaire opposent en Belgique catholiques et non-catholiques à « propos de l'organisation de l'enseignement », en particulier lors des deux épisodes de « guerres

Dans ce contexte et jusque dans les années 1980, les politiques scolaires belges ont du mal à définir des finalités partagées et une « politique » unifiée : au contraire, les oppositions entre les finalités et pédagogies de l'école publique et de l'école chrétienne sont tenaces, les principales divergences entre ces deux « mondes sociologiques » portant sur la répartition des moyens et la place respective des deux types d'écoles et de projets scolaires. Le projet de dépasser les oppositions par une « école pluraliste » reste dans les faits lettre morte, même s'il est prévu dans le pacte scolaire de 1959²¹. Tout au long des guerres scolaires, la politique scolaire concerne donc relativement peu le « projet institutionnel », les finalités ou la pédagogie qui seraient à promouvoir au niveau de l'ensemble du système scolaire.

Néanmoins, les politiques scolaires en Belgique au cours des années 1990 ont impliqué la définition d'objectifs généraux à l'échelle du système (les « missions » de l'école et les référentiels de compétences). Cela a entraîné la mise en avant d'une norme pédagogique de référence, la standardisation croissante de règles du jeu à l'échelle du système et enfin, la mise en place de modes de régulation qui agissent sur les résultats des écoles et de leurs élèves.

Ces politiques scolaires sont « nouvelles » pour une double raison : d'une part, elles se construisent à l'échelon du système et, d'autre part, elles visent à modifier la « boîte noire » de toutes les écoles (quel que soit le réseau) pour agir sur les conditions de l'efficacité éducative. Autrement dit, elles se réfèrent moins aux finalités axiologiques associées à un programme institutionnel et se formulent davantage comme des politiques visant à améliorer l'efficacité ou la justice du système dans son ensemble.

Cette politique éducative à l'échelle de la Communauté française est inspirée d'un paradigme du système de production scolaire que nous allons contraster par rapport au paradigme de l'institution scolaire, selon trois dimensions clés de tout référentiel d'action publique (Muller, 1995).

scolaires » (1878-1884 et 1954-1958). Comme elle le défend dans son texte, cette opposition entre tenants de l'école « officielle » et tenants de l'école religieuse catholique met en jeu deux « modèles forts différents de l'élève à former : le "citoyen autonome" est le type d'homme valorisé par le projet laïc, le "bon chrétien" est le type d'homme favorisé par le projet catholique », même si des courants pédagogiques transversaux se développent (Van Haecht, 1985, p. 10). Ces philosophies éducatives tendent à opposer « une conception éducative laïque où prédomine une pédagogie de l'éclaircissement – avec une dominante méthodologique existentielle – et une conception éducative catholique où prédomine une pédagogie de la révélation – avec une dominante méthodologique essentielle » (p. 10).

21. Loi modifiant certaines dispositions de la législation de l'enseignement du 29/05/1959. On trouvera le texte complet de cette loi à l'adresse: http://www.gallilex.cfwb.be/fr/leg_res_01.php?ncda=5108&referant=l02

Les valeurs de référence

Dans le paradigme du système de production scolaire, la recherche d'une meilleure performance, l'amélioration de la qualité et de l'efficacité de l'organisation scolaire tendent à devenir une préoccupation permanente. Cependant cette préoccupation « instrumentale » n'est pas dénuée de toute dimension axiologique, comme celle de valoriser la recherche d'efficacité pour elle-même. À l'opposé, le paradigme de l'institution scolaire impliquait de préciser le type de valeurs qui étaient au cœur de l'entreprise éducative, valeurs renvoyant soit à la collectivité, à l'institution sociétale dans son ensemble (comme dans le projet de l'école officielle), soit à une « communauté particulière » (catholique); valeurs diverses (et parfois communes) comme par exemple le pluralisme, la démocratie, la libre conscience, l'épanouissement de la personne, l'égalité, la solidarité, etc. Dans le paradigme du système de production scolaire, ces différentes valeurs ne sont pas frontalement évacuées, puisque l'efficacité peut être recherchée par rapport à la poursuite d'objectifs théoriquement variés, qui doivent cependant être visibles, mesurables, si bien qu'en pratique, toutes les finalités de référence ne seront pas poursuivies au même degré.

Les normes définissant des écarts entre le réel perçu et le réel souhaité

Dans le nouveau paradigme, le souhaitable est défini en termes de performances scolaires des élèves (relatives à l'acquisition de savoirs, compétences, savoir-faire, savoir-être). Plus précisément, ce sont des résultats mesurables qui constituent l'opérationnalisation de ce souhaitable. Des références chiffrées (*data*) deviennent des « standards » par rapport auxquels on confronte les résultats effectifs. Les écarts à ces standards dessinent alors les objectifs à poursuivre par la politique et la définition d'indicateurs et de standards devient clé dans la conduite des politiques (Ozga, 2009).

Par opposition, le souhaitable était antérieurement défini en termes de valeurs, de normes ou de propriétés référées à la socialisation des individus dans des entités collectives d'appartenance: ainsi, entre les réseaux d'enseignement belges, certaines figures de référence incarnant les valeurs à socialiser pouvaient être partagées (idéal de l'humaniste) mais déclinées à la lumière de pédagogies renvoyant soit à une référence chrétienne révélée, soit à une référence laïque valorisant « l'éclairement » et la raison (Van Haecht, 1985).

Les théories de l'action

Les théories de l'action concernent les relations causales types mobilisées par les décideurs politiques, pour orienter le domaine d'action scolaire vers le souhaitable. Dans le paradigme montant, le problème clé à résoudre est, on l'a vu, celui de la « qualité des apprentissages », envisagée sous l'angle prioritaire de l'efficacité ou de l'équité. Les nouveaux modèles de régulation « postbureaucratiques » (État évaluateur et quasi-marché) des systèmes d'enseignement (Maroy, 2006, 2008) peuvent alors être considérés comme des théories de l'action utilisées (consciemment ou non) par les politiques pour définir des voies d'amélioration de la qualité des apprentissages au sein du système de production scolaire :

- La qualité et les résultats seront améliorés si on promeut une « évaluation » et un processus d'apprentissage dans le système (c'est la théorie implicite du modèle de l'État évaluateur).
- La qualité sera promue si on promeut la « compétition » dans le système (modèle du quasi-marché)²².

Autrement dit, ce sont des théories de la « gestion » ou de la « gouvernance » mobilisées pour orienter les politiques scolaires. Par ailleurs, les nouveaux courants de recherche en sciences de l'éducation, autour des établissements ou des enseignants « efficaces » (Dumay & Dupriez, 2009) sont aussi mobilisés pour définir ce qu'il convient de promouvoir comme mode de gestion des établissements ou comme pratiques pédagogiques utiles pour assurer telle ou telle acquisition scolaire.

À l'inverse, dans le paradigme de l'institution, ce qu'il s'agit de promouvoir, ce sont les qualités sociales et intellectuelles des types d'individus à « socialiser » par les institutions d'enseignement. La question du *comment* ces valeurs s'incarnent dans une organisation, un curriculum, une pédagogie, une formation des enseignants est alors cruciale. Autrement dit la théorie de l'action suppose de mobiliser des théories philosophiques, sociologiques, didactiques ou pédagogiques relatives au mode de formation et de socialisation des agents de l'institution scolaire (ethos attendus des professeurs et de la direction), au type de choix de contenu et de curriculum à privilégier, au mode de relation pédagogique envisagé ou aux implications normatives du mode d'évaluation des élèves. Autrement dit, ces processus constituent des moyens « d'incarner », de concrétiser les options normatives privilégiées par les institutions d'enseignement.

Ce changement de paradigme politique n'est pas spécifique au contexte politique et scolaire belge ; il opère très largement dans l'ensemble des

22. Ces modèles peuvent bien entendu être combinés dans les pratiques et faire l'objet de retraductions diverses, prenant en considération les contraintes pragmatiques et symboliques diverses d'acteurs hétérogènes tant du point de vue de leurs situations que de leurs logiques.

démocraties et des systèmes scolaires occidentaux et au-delà, mais avec des temporalités, des modalités et des acteurs du changement qui sont chaque fois singuliers. Notre propos est maintenant de préciser quelles conditions spécifiques ont joué dans la fragilisation de l'ancien paradigme et l'adoption du nouveau dans le contexte belge. L'analyse de la genèse du nouveau paradigme et de son hégémonie progressive à un niveau transnational ne peut être développée dans le cadre de cet article. Limitons-nous à avancer que la conceptualisation de l'école comme système de production scolaire, transformant des *inputs* en *outputs* au travers d'une technologie et d'une organisation éducative est davantage portée par les économistes du capital humain ou la théorie des organisations, vision qui sera largement diffusée et promue dès les années 1970 par l'OCDE notamment (Pons, 2008; Rizvi & Lingard, 2006)²³.

CONDITIONS STRUCTURELLES ET ACTEURS DU CHANGEMENT

Comment expliquer que le paradigme du « système de production scolaire » ait acquis progressivement droit de cité en Belgique francophone dans un monde scolaire traversé par des oppositions institutionnelles? Notre thèse est que la transformation du cadrage cognitif et normatif partagé par les acteurs du secteur éducatif est favorisée par la conjonction, à un niveau national, de conditions socio-économiques contribuant à « mettre en crise » le fonctionnement de l'école et la pertinence de politiques scolaires inspirées par le paradigme de l'institution. Par ailleurs, le travail d'un certain nombre d'entrepreneurs institutionnels a contribué à forger, diffuser et légitimer de nouvelles orientations aux politiques scolaires, inspirées du nouveau paradigme. Cela dans un contexte international favorable au changement.

23. On pourrait argumenter que les travaux des sociologues de la « reproduction sociale » (Bourdieu & Passeron, 1970) thématisent également l'école comme un « système » de reproduction des inégalités sociales et insistent sur le fait que « l'institution scolaire et ses fonctions officielles » (thématisée par Durkheim en insistant sur sa vertu émancipatrice) cachent une fonction de légitimation et de reproduction des inégalités. Il nous semble quant à nous que la conception du « système d'enseignement », chez Bourdieu en tout cas, n'implique en rien une conception fonctionnelle et « technocratique » du système en fonction des besoins du système économique et des demandes de la « société ». L'école est une institution dont le sens et les fonctions sont liés chez lui à une société de classes, une institution créatrice dès lors de distinctions symboliques, de socialisations différentielles et d'inégalités sociales, irréductible à une fonction de production de qualifications en fonction des demandes sociales. L'école ne peut donc en aucune manière être considérée pour Bourdieu comme « un système de production scolaire » au sens où les économistes l'entendent (voir en particulier, Bourdieu & Passeron, 1970, pp. 209-220).

Conditions structurelles

Plusieurs conditions structurelles ont contribué à fragiliser les orientations antérieures des politiques scolaires, limitées alors à la gestion des équilibres institutionnels et financiers et renvoyant les questions des finalités éducatives et méthodes pédagogiques au niveau des réseaux. Tout d'abord, des sacrifices budgétaires et sociaux vont s'imposer au cours des années 1990. La conjonction de politiques éducatives coûteuses et de la communautarisation du système d'enseignement contraint le gouvernement à imposer des économies. Celles-ci déclenchent un large mouvement de protestation sociale qui, exceptionnellement, déborde des clivages en termes de piliers. Les crises dans le secteur de l'enseignement ont été nombreuses depuis la naissance de la Belgique. Cependant, celle qui émerge en 1990 est d'une « ampleur sans précédent » (Grootaers, 2005, p. 32) et n'oppose pas une tendance philosophique à une autre, ni une tendance idéologique à une autre. Elle rassemble dans la rue des enseignants, qui sont présents en nombre, des étudiants, des parents qui tous s'allient pour s'opposer au pouvoir politique et revendiquer une réelle politique éducative qui dépasse les questions économiques et budgétaires et qui s'attache aux missions fondamentales de l'école. Des coordinations entre enseignants se développent à un niveau inter-réseaux. Le mouvement porte un contre-discours de type « qualitatif » qui s'oppose au discours politique quantitatif et budgétaire.

Ce contexte national tendu du début des années 1990 est révélateur d'une crise de sens de l'école et fragilise la légitimité des politiques scolaires conduites jusqu'alors. La demande pour des politiques axées sur les finalités de l'école et du système scolaire dans son ensemble est maintenant réclamée par la base. C'est donc une double crise, budgétaire d'une part, sociale et symbolique de l'autre, qui crée une condition favorable à l'émergence d'un nouveau cadrage symbolique des politiques scolaires.

Le deuxième élément structurel est le contexte international où l'on voit, dans d'autres pays, une transformation déjà bien avancée des systèmes éducatifs. De manière assez générale, en Europe, un modèle type proche du modèle de l'État évaluateur se développe. « The current model is for the State to set educational goals and priorities and evaluate outcomes, leaving it to individual education districts and schools to comply » (Arnové, 2008, p. 113). Et cela, alors même que la diffusion de l'information sur les différents systèmes éducatifs s'intensifie, notamment par le biais des grandes organisations comme l'OCDE. Ces évolutions sont concomitantes avec l'évaluation négative des performances de notre système éducatif par différentes enquêtes internationales et nationales (IDE²⁴, IEA²⁵, OCDE), à partir du milieu des

24. IDE, investissements et planification en matière d'enseignement.

25. IEA, *International Association for the Evaluation of Educational Achievement*, connue en français sous le nom AIE, Association internationale pour l'évaluation des rendements scolaires.

années 1980, et la prise de conscience des profondes inégalités entre élèves mais aussi du manque de connaissance et d'information dont disposaient l'administration et les pouvoirs politiques par rapport à leur système éducatif.

Acteurs entrepreneurs

Mais c'est aussi l'action d'entrepreneurs institutionnels (Campbell, 2004) qui a progressivement forgé/importé/promu de nouvelles représentations de la réalité scolaire et contribué à transformer le cadrage cognitif dans lequel se pensaient les politiques scolaires. En développant des outils de l'action publique, en important de nouvelles connaissances dans le champ scolaire, en donnant un certain sens aux événements et en formant des coalitions, ces acteurs deviennent des figures porteuses du nouveau paradigme en Belgique et de véritables entrepreneurs. Ils pointent les anomalies du système et contribuent à mettre en crise l'ancien référentiel.

La structure légère de pilotage mise en place informellement en 1994 par le secrétaire général de l'administration de l'époque et développée en marge des instances officielles de négociation dans le secteur, mérite d'être mentionnée. Par leurs actions, les membres de cette structure ont (re)défini les problèmes, leurs causes et les solutions à apporter ; ils importent de nouvelles idées dans le champ scolaire et les diffusent. Face à la crise de gestion du système, face à la crise sociale évoquée, ils dénoncent les problèmes en termes d'équité et d'efficacité et proposent d'évaluer les dispositifs et arrangements institutionnels existants à cette aune. Le fait que cette coalition²⁶ d'entrepreneurs soit composée d'acteurs situés sur différentes scènes (administration, inspection, monde académique, politique), et simultanément issus des deux piliers, catholique et laïc, de la société belge, contribue à l'impact important de cette structure légère. Cette instance va lancer plusieurs actions, notamment des évaluations externes (les premières communes à tous les réseaux) et une réflexion autour des objectifs de l'enseignement et de son pilotage, fondements théoriques et pratiques des décrets qui suivront sur les « missions », le « pilotage » et l'évaluation externe.

Un second groupe d'acteurs se situe dans la sphère académique. Certains ont participé à la structure légère de pilotage. Pour décrire le réel, ces chercheurs ont apporté de nouveaux termes (pilotage, régulation, évaluation externe, système, efficacité, rendement, performance, etc.) qui vont s'implanter progressivement (De Landsheere, 1992, 1994 ; Grisay, 1982). Des variantes entre leurs discours sont bien sûr observables, mais ils ont tous

26. Le terme « coalition » est à entendre ici comme un rassemblement, une « agrégation » d'acteurs autour d'un idéal commun et de catégories normatives et cognitives proches, qui s'inscrivent dans le paradigme émergent. Il ne s'agit pas d'un regroupement stratégique en vue de défendre des intérêts individuels ou même les intérêts d'un groupe d'acteurs.

participé à l'importation d'idées nouvelles, de problématisations innovantes en Communauté française par rapport au paysage éducatif belge traditionnel, faisant cohabiter plusieurs réseaux aux programmes institutionnels spécifiques. Une partie de ces chercheurs sont des économistes qui ont porté un discours sur l'efficacité de l'enseignement, en prouvant, chiffres à l'appui, que notre enseignement n'était pas sous-financé par rapport à ses voisins mais que les recettes devaient être utilisées plus efficacement. Ces économistes parlent de « rendement qualitatif du système éducatif », « performances des élèves », etc. (Delvaux, Deutsch, De Villé, Martou, Vandenberghe, 1995 ; Moutard, 1990). Une autre partie de ces chercheurs plutôt pédagogues et sociologues va aussi développer un discours sur l'inefficacité du système au vu du nombre important de redoublements (Crahay, 1992, 1996). Ce constat, bien relayé dans la presse et dans les rapports internationaux sur notre système éducatif, est un autre élément cognitif qui a contribué à la promotion du nouveau paradigme en Belgique francophone. La double position de la plupart de ces chercheurs sur la scène nationale et internationale est certainement un élément qui leur confère une place de choix dans l'action publique et qui contribuera à l'influence de leurs discours.

Un troisième groupe (plus restreint) d'entrepreneurs comprend les acteurs politiques « purs », qui vont relayer ce discours sur la scène politique et permettre la concrétisation d'idées dans des projets politiques et des textes légaux. Ils apparaissent plus tardivement sur la scène (milieu des années 1990).

Tous ces entrepreneurs vont non seulement contribuer par leurs écrits à construire un diagnostic des problèmes à résoudre (manque de qualité, inefficacité et inéquité du système, absence de connaissance), mais ils vont aussi forger et diffuser les termes et les outils à promouvoir pour y faire face. De par leurs positions intermédiaires entre plusieurs mondes (international/national ; pilier laïc/pilier catholique ; réseau officiel/réseau libre), plusieurs formations politiques et plusieurs disciplines (économie, sociologie, psychologie et sciences de l'éducation), ils vont contribuer à rendre acceptables des pistes de solution de type *accountability* auprès des acteurs institutionnels clés au sein du champ scolaire.

Ressources cognitives

Ces entrepreneurs institutionnels n'agissent pas seulement en fonction de leurs expériences propres, ou en mobilisant leur réseau de relations. Ils s'appuient de surcroît sur divers types de connaissances clairement identifiés comme adjutants à l'entreprise de promotion d'une politique de pilotage et d'évaluation du système scolaire en Belgique francophone. Ces ressources

cognitives sont formalisées dans des documents²⁷ qui circulent dans la sphère de décision publique et semblent émaner d'organismes « neutres » et « extérieurs ». En réalité, plusieurs acteurs/entrepreneurs institutionnels agissant par exemple dans le cadre de la structure légère de pilotage ont participé de près ou de loin à la production de ces rapports.

La radioscopie de Delvaux et Vandenberghe (1992) est le fruit d'une recherche interuniversitaire chargée de réaliser un état des lieux de notre enseignement. Il s'agit en réalité d'une des réponses à la crise et aux grèves de 1990 qui résultent d'un accord entre ministres et syndicats décidant de soumettre l'école à une analyse de ses besoins et de ses moyens. Cependant, en raison des contraintes budgétaires de l'époque, l'intérêt des politiques se porte principalement sur le volet analysant l'utilisation des ressources budgétaires (Delvaux & Vandenberghe, 1992). Si le terme de pilotage n'est jamais mentionné dans ce document, celui-ci n'en promeut pas moins une vision de l'enseignement en tant que « système » et met au centre du débat la question de l'efficacité en montrant combien l'allocation des ressources n'est pas adaptée aux défis du décrochage scolaire. Cette radioscopie préfigure aussi le thème de la régulation par les résultats, les auteurs expliquant dans leur conclusion qu'il semble « urgent de passer d'une gestion par les normes à une gestion par les objectifs » en matière d'éducation.

Les deux rapports publiés par l'OCDE (1991, 1993) sont un autre repère cognitif significatif. Ils soulignent « l'absence de pilotage » dans notre enseignement, une pratique excessive des redoublements, une formation de base commune limitée, des insuffisances dans la recherche en éducation et la « faible culture d'évaluation » en Belgique. On y parle de problèmes de qualité de nos enseignements, liés notamment à l'inexistence de dispositifs de pilotage du système²⁸. Bien que ce rapport ne soit pas médiatisé auprès de l'opinion publique, il est par contre largement diffusé auprès des acteurs de l'action publique et c'est un électrochoc pour une série d'entre eux.

Ces publications phare vont cristalliser certains argumentaires largement repris par la suite et joueront un rôle de publicisation et de diffusion des problèmes à traiter et des solutions à esquisser sur le mode du pilotage et de

27. Nous exposons ici brièvement le contenu de deux rapports que l'on sait avoir joué un rôle majeur en matière de changement cognitif et normatif dans le secteur éducatif: la radioscopie de Delvaux et Vandenberghe (1992) et le rapport OCDE sur le système éducatif belge (1993). Il n'est évidemment pas exclu que d'autres documents aient influencé les débats scolaires de l'époque.

28. « La Belgique connaît une faible culture de l'évaluation dans et sur l'enseignement; et aussi des réticences pour donner un minimum de dimension scientifique à sa gestion et à ses modalités de suivi. En effet, l'état d'esprit d'indépendance qui a prévalu jusqu'alors dans les réseaux d'enseignement n'a pas permis d'atteindre – d'accepter? – le minimum de transparence nécessaire à une saisie rapide et correcte des caractéristiques de son fonctionnement » (OCDE, 1993, p. 113).

l'évaluation. À un titre ou à un autre, ces documents, qui seront beaucoup lus et commentés, participent aussi à la structuration du nouveau paradigme.

Par ailleurs, certains entrepreneurs institutionnels les plus anciennement investis dans la promotion du pilotage ont participé à la promotion d'épreuves externes dans les différents réseaux d'enseignement (épreuves cantonales et interdiocésaines), avant ou parallèlement à la mise en place de la structure légère de pilotage. Ces outils, l'expertise des professionnels qui les mettent en place, les argumentaires sur leur fonction et bien-fondé, ont également constitué des points d'appui cognitifs dans l'instauration et la construction d'une légitimité pragmatique de la politique de pilotage lorsqu'elle s'incarne dans des épreuves d'évaluation externes non certificatives puis certificatives.

CONCLUSION

Nous avons proposé une étude de la construction de la politique de pilotage et d'évaluation dans le secteur de l'enseignement en Belgique francophone, appréhendée à la lumière d'une approche cognitive des politiques publiques. À partir du début des années 1990, on assiste à un développement de dispositifs d'évaluation externe et de pilotage. Or cette politique de mise en œuvre d'un État évaluateur ne peut pas seulement être interprétée comme une réponse fonctionnelle à des (changements de) conditions structurelles internes et externes (par exemple le poids budgétaire très important de l'éducation après la communautarisation de l'enseignement). Par-delà ces conditions structurelles favorisantes (crise financière et sociale à l'interne, montée d'une globalisation des politiques éducatives et mise en comparaison des systèmes à l'externe), il faut souligner que les nouvelles politiques scolaires censées répondre à ces problèmes sont largement dérivées de la montée d'un nouveau paradigme de politique scolaire, qui conçoit l'école non plus comme une institution mais avant tout comme un système de production scolaire. Ce paradigme contribue à thématiser l'école comme une organisation productive, visant des résultats mesurables et visibles, opérationnalisant, à des degrés divers, les principales finalités socialement assignées au système scolaire. Les problèmes à résoudre par la politique scolaire dérivent dès lors d'une vision souvent « quantifiée » du souhaitable, même si les cibles à atteindre sont subsumées sous le terme ambigu de qualité scolaire. Les théories de l'action inspirant les voies et moyens à suivre par les politiques pour atteindre ces objectifs dérivent de diverses théories (relatives à la gouvernance et gestion scolaire, à l'efficacité pédagogique de l'établissement ou de l'enseignant) dont le point commun est penser les moyens pour améliorer les performances du système, en termes d'efficacité, d'efficience, mais aussi d'équité.

Nous avons par ailleurs insisté sur le rôle important d'entrepreneurs institutionnels qui ont contribué à coproduire, à importer, adapter et traduire ce nouveau paradigme politique en l'ajustant au contexte belge; ils ont donc contribué à diffuser une « réponse politique » aux diverses crises qui ont marqué l'histoire scolaire belge au début des années 1990. Schématiquement, il s'agissait désormais de considérer l'école comme un système de production, dont il faut garantir l'efficacité et surtout l'équité par un « pilotage » adéquat.

Cette vision de l'école et de la politique éducative ne s'est pas construite de façon « diffuse » ou par des acteurs politiques habituels (agents professionnels des partis ou des cabinets ministériels par exemple). Elle a été notamment favorisée par un groupe « informel » rassemblant des entrepreneurs académiques de plusieurs universités, des responsables de haut niveau de l'administration et de l'inspection, combinant par leurs histoires professionnelles du capital social dans plusieurs réseaux scolaires. Ces acteurs vont contribuer par leurs écrits à construire un diagnostic des problèmes à résoudre (insuffisante efficacité, inéquité, manque de connaissance du système), mais ils vont aussi forger et diffuser les termes et les outils à promouvoir pour y faire face (pilotage du système, évaluation externe, etc.). De par leurs positions intermédiaires entre plusieurs mondes (international/national; réseau officiel/réseau libre; analyse économique/analyse psychométricienne du « rendement scolaire ») et plusieurs disciplines (économie et sciences de l'éducation), ils vont contribuer à rendre acceptables ces pistes de solution auprès de plusieurs acteurs institutionnels clés au sein du champ scolaire (notamment l'administration, les partis, les réseaux).

En définitive, l'émergence de politiques d'évaluation externe en CFB introduit une discontinuité importante dans les politiques scolaires belges, discontinuité qui ne peut simplement s'expliquer par des transformations factuelles de l'environnement politique et social, ou en relation avec l'évolution des équilibres de forces politiques en présence. Pour la comprendre, il faut d'une part prendre la mesure du changement de paradigme de politique scolaire qui sous-tend cette nouvelle politique et, d'autre part, souligner le rôle d'acteurs institutionnels clés dans la coconstruction et la traduction de ce paradigme en Belgique.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Arnove, R. (2008). Education Reform Around the World: To What Ends? In J. Resnik (Ed.), *The Production of Educational Knowledge in the Global Era* (pp.109-120). Rotterdam: Sense Publishers.
- Bourdieu, P. & Passeron, J. C. (1970). *La reproduction. Éléments pour une théorie du système d'enseignement*. Paris: Minuit.

- Broadfoot, P. (2000). Un nouveau mode de régulation dans un système décentralisé: l'État évaluateur. *Revue française de pédagogie*, 130, 43-55.
- Campbell, J. L. (2004). *Institutional change and globalization*. Princeton: Princeton University Press.
- Cattonar, B., Mangez, E., Delvaux, B., Mangez, C. & Maroy, C. (2009). *Usage et circulation des enquêtes PISA en Communauté française de Belgique* (KnowandPol research report, Belgian Education Team, R&D EU Sixth Framework Program). Consulté le 31 janvier 2011 dans <http://www.knowandpol.eu>
- Crahay, M. (1992). *L'échec scolaire en Communauté française de Belgique: analyse et perspectives d'amélioration: notes de synthèse préparées à la demande de la Commission pour l'Éducation et la Formation de la Communauté française de Belgique* (Rapport de recherche). Université de Liège ULg, Département d'éducation et formation.
- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire?* Bruxelles: De Boeck.
- Crahay, M. (1997). *Une école de qualité pour tous*. Bruxelles: Labor.
- De Landsheere, G. (1992). *Le défi de l'éducation*. Communication présentée au colloque *La Wallonie au futur*, Institut Jules Destrée, Région Wallonne.
- De Landsheere, G. (1994). *Le pilotage des systèmes éducatifs*. Bruxelles: De Boeck.
- Delvaux, B., Deutsch, L., De Villé, P., Martou, F. & Vandenberghe, V. (1995). *La politique de l'enseignement: bilan et perspectives* (Études et Opinions). Louvain-la-Neuve: IRES.
- Delvaux, B., Giraldo, S. & Maroy, C. (2005). *Réseaux et territoires. Les régulations intermédiaires dans le système scolaire en Communauté française*. Louvain: Presses Universitaires de Louvain.
- Delvaux, B. & Vandenberghe, V. (1992). *Radioscopie de l'enseignement en Communauté française de Belgique: les dépenses de personnel dans l'enseignement fondamental et secondaire de la Communauté française* (Rapport de recherche). IRES-UCL Université catholique de Louvain, Département des sciences économiques.
- Deprez, G. (1970). La guerre scolaire et sa pacification. *Recherches sociologiques*, 2, 61-91.
- Draelants, H., Dupriez, V. & Maroy, C. (2003). *Le système scolaire en Communauté française* (Collection Dossiers du CRISP). Bruxelles: Éditions du Centre de Recherche et d'Information Socio-Politique.
- Dubet, F. (2002). *Le déclin de l'institution*. Paris: Seuil.
- Dumay, X. & Dupriez, V. (2009). *L'efficacité dans l'enseignement. Promesses et zones d'ombre*. Bruxelles: De Boeck.
- Dumont, H. (1999). La pilarisation dans la société multiculturelle belge. *La Revue Nouvelle*, 3, 46-75.
- Durkheim, E. (1922). *Éducation et Sociologie*. Paris: PUF.

- Grisay, A. (1982). *Rendement en français, notes et échecs à l'école primaire : les mirages de l'évaluation scolaire* (Rapport de recherche). Université de Liège ULg, Laboratoire de pédagogie expérimentale.
- Grootaers, D. (2005). Les mutations de l'égalité des chances à l'école, *Courrier hebdomadaire du CRISP*, 1893, 4-43.
- Hall, P. A. & Taylor, R. C. R. (1993). Policy, Paradigms, Social Learning and the State: The Case of Economic Policy-Making in Britain. *Comparative Politics*, 25(3), 275-296.
- Kuhn, T. S. (1983). *La structure des révolutions scientifiques*. Paris: Flammarion.
- Mangez, C., Maroy, C., Cattonar, B., Delvaux, B. & Mangez, E. (2009). *The construction of steering and evaluation policy in French-speaking Belgium: a cognitive approach* (Rapport de recherche à la Commission européenne, KnowandPol, WP 10, Belgian Education Team, R&D EU Sixth Framework Program). Consulté le 31 janvier 2011 dans <http://www.knowandpol.eu>
- Mangez, E. (2009). De la nécessité de discrétion à l'État évaluateur. *La Revue Nouvelle*, 7-8, 32-37.
- Maroy, C. (2006). *École, régulation et marché. Une comparaison de six espaces scolaires en Europe* (Collection Éducation et Société). Paris: PUF.
- Maroy, C. (2008). Vers une régulation post-bureaucratique des systèmes d'enseignement? *Sociologie et Sociétés*, 40(1), 31-54.
- Maroy, C. (2009). Réforme de l'inspection et montée de la régulation par les résultats en Belgique. Incidences des institutions, des intérêts et des idées. In G. Pelletier (Éd.), *La gouvernance en éducation. Régulation et encadrement dans les politiques éducatives* (pp. 53-72). Bruxelles: De Boeck.
- Maroy, C. & Dupriez, V. (2000). La régulation dans les systèmes scolaires. Proposition théorique et analyse du cadre structurel en Belgique francophone. *Revue française de pédagogie*, 130, 73-87.
- Moutard, F. (1990). L'enseigne...ment. *La Revue Nouvelle*, 9, 8-17.
- Muller, P. (1995). Les politiques publiques comme construction d'un rapport au monde. In A. Faure, G. Pollet & P. Warin (Éd.), *La construction du sens dans les politiques publiques. Débats autour de la notion de référentiel* (pp. 153-179). Paris: L'Harmattan.
- Muller, P. (2000). L'analyse cognitive des politiques publiques: vers une sociologie politique de l'action publique. *Revue française de science politique*, 50(2), 189-208.
- Organisation de coopération et de développement économiques. (1991). *Examens des politiques nationales d'éducation : Les systèmes éducatifs en Belgique: similitudes et divergences*. Bruxelles-Eupen: Ministerie van de Vlaamse Gemeenschap (Departement Onderwijs), Ministère de l'Éducation, de la Recherche et de la Formation, Verwaltung der Deutschsprachigen Gemeinschaft, Abteinkung Unterricht.

- Organisation de coopération et de développement économiques. (1993). *Examen des politiques nationales d'éducation : Belgique*. Paris: Éditions de l'OCDE.
- Ozga, J. (2009). Governing education through data in England: from regulation to self-evaluation. *Journal of Education Policy*, 24(2), 149-162.
- Pons, X. (2008). *L'évaluation des politiques éducatives et ses professionnels*. Thèse de doctorat, Sciences Po Paris.
- Rizvi, F. & Lingard, B. (2006). Globalization and the Changing Nature of the OECD's Educational Work. In H. Lauder, P. Brown, J. Dillabough & A. H. Halsey (Éd.), *Éducation, Globalization & Social Change* (pp. 247-260). Oxford: Oxford University Press.
- Van Haecht, A. (1985). *Perspectives historiques. Valeurs laïques. Valeurs religieuses dans l'enseignement: une mission*. Communication au colloque École publique, École privée, sl.

L'évaluation des performances scolaires des élèves : un instrument d'évaluation des politiques éducatives ?

Nathalie Mons* & Marcel Crahay**

***Université Paris-Est Marne-la-Vallée, LATT¹**

****Universités de Genève et de Liège**

INTRODUCTION

Depuis trois décennies environ, l'ensemble des secteurs de l'action publique dans les pays de l'OCDE sont submergés par le déferlement d'une rhétorique considérant l'évaluation comme l'outil cardinal de la bonne gouvernance. Cette idéologie ou « culture d'évaluation » se déploie autour de l'idée noyau d'une nécessaire intégration des évaluations scientifiques dans la formulation des politiques publiques. La manifestation la plus forte de ce courant de pensée s'incarne dans la notion programmatique d'*evidence-based policy* (Demailly, 2005 ; Moran & Malott, 2004).

En éducation, des dispositifs d'évaluation quantitative des acquis des élèves, désormais observables dans la quasi-totalité des pays (Mons, 2009), s'affirment comme le socle d'un mouvement plus général d'*accountability*. À la suite de Broadfoot (2000) et de Harris et Herrington (2006) parmi d'autres, on peut définir cette notion, souvent traduite en français par le néologisme responsabilisation², comme le couplage de mesures de la

1. Laboratoire techniques, territoires et sociétés.

2. Pour une analyse du passage du concept de responsabilité, classique en droit et en philosophie, au néologisme de responsabilisation, développé à partir des années 1970 dans le contexte d'un affaiblissement de l'État-providence et d'un développement des approches néo-libérales du lien politique et social, voir l'article introductif de Charbonneau et Estèbe

production d'un secteur donné (concernant l'école, il s'agit en particulier des acquis des élèves), l'évaluation des écarts entre objectifs et résultats de l'action publique ainsi que la publicisation de ces évaluations en direction du grand public et/ou de la hiérarchie administrative et politique (notions de *reporting* ou reddition des comptes). Selon les configurations nationales, ces mesures de performances sont enserrées dans des dispositifs de sanctions/récompenses plus ou moins contraignants qui dessinent du côté européen une *accountability* douce par opposition à des régimes plus durs dans les pays anglo-saxons.

Bien qu'en lien avec la tradition d'évaluation des connaissances pratiquée de longue date au sein de l'école, les évaluations des performances scolaires des élèves visent désormais – c'est du moins ce qui est proclamé au niveau des discours (Lessard, 2006; Lessard & Saussez, 2009) – à appréhender, sur une base décontextualisée et comparative, les acquis des élèves, le plus souvent dans un objectif de mesure des performances des unités opérationnelles composant les systèmes scolaires: les écoles, les classes et/ou les enseignants ou, dans certains pays, les collectivités locales lorsque des compétences en matière de scolarité leur ont été conférées. Ces nouvelles politiques éducatives renouvellent les modes de régulation des systèmes scolaires, qui évoluent sensiblement depuis une stricte conformité au respect de normes vers une orientation davantage marquée par une évaluation des performances scolaires (Crahay, Audigier & Dolz, 2006; Maroy, 2008; Mons, 2007). Ce faisant, elles interrogent les responsabilités des acteurs de l'école. Les enseignants, et plus largement les établissements scolaires et les collectivités locales, sont sommés de rendre des comptes sur les performances des élèves dont ils ont la charge.

Si la responsabilisation des acteurs locaux devient ainsi un des mécanismes clés du jeu scolaire, pour autant les responsables nationaux³ et les choix politiques qu'ils développent sont ou, en tout cas, devraient être au cœur de cette nouvelle gouvernance par l'évaluation. L'article présenté ici interroge les relations entre l'évaluation des politiques éducatives nationales et ces nouveaux mécanismes d'*accountability*, dont les épreuves standardisées des acquis des élèves sont le socle. En suivant Duran (1999), nous définirons ici l'évaluation des politiques publiques comme

l'appréciation raisonnée des actions définies et mises en œuvre par des autorités publiques pour leur contribution au traitement de problèmes publics qu'elles ont identifiées comme nécessitant leur identification, appréciation intervenant sur la base d'une connaissance approfondie des conséquences que ces actions induisent comme des processus qui les produisent. (p. 184)

(2001) du numéro spécial de *Lien social et Politiques* sur « l'appel à la responsabilité à l'ordre du jour ».

3. Ou des autorités locales supérieures en charge de la formulation des politiques éducatives dans les pays fédéraux.

L'objectif de cet article est donc d'analyser dans quelle mesure, au-delà de la rhétorique politique, les résultats des évaluations standardisées sont bien incorporés dans l'évaluation des politiques publiques d'éducation, à la fois en mobilisant des apports théoriques qui font le lien entre ces pratiques de *testing* et l'évaluation des réformes en éducation et en étudiant les politiques éducatives réellement mises en œuvre dans ce domaine dans les pays de l'OCDE. Pour ce faire, dans un premier temps, l'article se propose d'examiner les modèles théoriques qui sont régulièrement invoqués pour justifier les réformes mettant l'accent sur l'évaluation dans le secteur public en général et dans le secteur éducatif en particulier. Il faut – pensons-nous – distinguer, d'une part, les courants du *New Public Management*, de l'*Evaluation policy* ou de l'analyse des politiques publiques qui s'inscrivent dans une réflexion trans-sectorielle sur la réforme de l'État et les politiques publiques et, d'autre part, les courants théoriques qui se sont développés en éducation (en particulier, la théorie du capital humain en économie de l'éducation et le mouvement de la *School effectiveness*). Selon notre analyse, si les théories relatives à la réforme de l'État confèrent bien une nouvelle place à l'évaluation des politiques publiques, et donc à celle des acteurs politiques œuvrant à un niveau macro, les courants théoriques qui se sont développés en éducation privilégient davantage l'évaluation des unités et des acteurs opérationnels que sont les établissements scolaires ou les collectivités locales.

Dans un second temps, l'article présentera, à partir de quelques exemples nationaux diversifiés, les usages des programmes d'évaluations standardisées réellement mis en œuvre et leurs liens avec l'évaluation des réformes en éducation. Il ressort de ces exemples concrets que les relations entre l'analyse des performances des élèves et les choix nationaux sont peu développées. À l'exception de quelques réformes, comme celles portant sur les curricula, les mesures quantitatives des acquis des élèves contribuent rarement à l'évaluation des politiques nationales. En conclusion, l'article exposera quelques pistes d'interprétation pour expliquer ce décalage entre une rhétorique politique centrée sur la responsabilisation des acteurs et les usages réels des dispositifs d'épreuves standardisées fort en retrait en ce qui concerne l'évaluation de l'action des acteurs politiques.

CADRES THÉORIQUES DE L'ÉVALUATION DES POLITIQUES PUBLIQUES

Les 40 dernières années ont été marquées par le développement, à travers des prismes disciplinaires variés (l'économie, la science politique, la sociologie) de théories généralistes plus ou moins normatives faisant une place renouvelée à l'évaluation des politiques publiques. Avec des orientations différentes, les mouvements de l'*Evaluation policy*, de l'analyse des

politiques publiques et le courant pragmatique du *New Public Management* ont insisté sur la nécessaire mesure des performances de production des services publics et l'intégration de ces résultats dans la formulation des politiques publiques. Si certains mouvements théoriques en éducation sont en lien avec ces cadres théoriques trans-sectoriels, ils se sont par contre davantage centrés sur l'évaluation et la responsabilisation des acteurs opérationnels des systèmes scolaires au détriment d'une évaluation des politiques éducatives. Pour cette analyse de la littérature théorique, nous nous fondons ici sur une revue de la littérature extensive qui a intégré des corpus à la fois européens et nord-américains⁴.

La justification de l'évaluation dans les politiques publiques

S'appuyant sur le modèle des sciences expérimentales dans la veine des *policy sciences*, le mouvement américain de la *Policy Evaluation* (Spencehauer, 2003, cité par Mons & Pons, 2009a), courant de pensée pragmatique qui trouve son origine aux États-Unis dans la concurrence entre le Congrès et le pouvoir exécutif dès la fin du XIX^e siècle, se propose de découper en trois séquences les dispositifs d'action politique. Les décideurs doivent tout d'abord clairement expliciter le projet de leurs politiques (objectifs et moyens d'action), afin de pouvoir dans un deuxième temps choisir les méthodes de mesure et les indicateurs pertinents et enfin déterminer, par la comparaison, les effets d'une politique. Ce triptyque « projet-mesure-analyse des effets » est censé fonder toute évaluation de politiques.

Figure 1 : La vision séquentielle des politiques publiques proposée par le mouvement de l'*Evaluation policy*



Comme l'illustre la figure 1, l'évaluation s'inscrit alors dans une vision séquentielle et rationnelle de l'action publique : l'élaboration et la mise en œuvre des politiques publiques obéissent à un enchaînement d'étapes logiques ; quant aux objectifs des politiques publiques, ils peuvent être atteints si les bons moyens sont mis en face d'objectifs clairement définis. Ce courant de pensée reflète une vision *top-down* de l'action publique

4. Quatre bases de données bibliographiques ont été mobilisées pour cette recherche documentaire : FRANCIS (Base de données bibliographiques multidisciplinaire en sciences humaines et sociales avec une forte représentation de la littérature française et européenne), ERIC (*Educational Resources Information Center*, base de données en éducation gérée par le ministère de l'Éducation américain), et les bases de données (ouvrages et périodiques) du laboratoire IREDU de l'Université de Bourgogne.

dans laquelle les politiques publiques seraient mises en œuvre telles que décidées dans leur cadre légal – réglementaire ou législatif – par les acteurs locaux. L'évaluation des politiques publiques est ici un outil majeur de l'étape finale (« résultats ») et doit donc permettre la confrontation entre les objectifs initiaux et les performances atteintes au final.

En opposition avec cette vision rationaliste de l'action publique, certains sociologues qui se réclament du courant de l'analyse des politiques publiques donnent aussi un rôle crucial à l'évaluation, en particulier pour la compréhension des effets des politiques publiques (Duran, 1999). Cette nouvelle intelligibilité des politiques est censée, par le biais de la saisie des effets produits qu'elle permet, influencer la formulation des politiques publiques, ce lien entre connaissance et action publique étant plus postulé que démontré. Face à une action publique qui est plus que jamais reconnue comme incertaine et complexe car trans-sectorielle et fragmentée entre des acteurs multiples, dans une posture qui n'est pas non plus exempte d'une certaine normativité, les théoriciens de ce courant recommandent le développement de dispositifs d'évaluation des politiques publiques comme un mécanisme de connaissance des effets toujours potentiellement imprévus des politiques publiques.

L'évaluation ne trouve pas sa source dans les faillites de ce que l'on a pu appeler les faillites de l'État providence. Elle atteste aussi des transformations profondes de l'action publique [...]. La maîtrise des politiques publiques est devenue d'autant plus délicate que celles-ci gèrent des problèmes politiques complexes, diffus, malaisément cernables dans des logiques techniques simples et des expertises spécialisées. La formulation et la mise en œuvre de telles politiques font appel à une multiplicité d'acteurs portés par des intérêts eux-mêmes multiples et elles développent des conséquences plurielles, le plus souvent inattendues. Cette nouvelle intelligence de l'action publique conditionne la recherche de modes de connaissance des politiques publiques plus appropriés que les approches administratives classiques, comme le renouvellement des techniques de pilotage et de contrôle de l'action. (Duran, 1999, p. 182)

Dans ce cadre sociologique, l'évaluation ne s'inscrit pas dans une vision pragmatique des politiques publiques, elle vise à construire connaissance et expérience en vue de l'action future: « Si la rationalité des techniques classiques d'aide à la décision était de nature prospective, fondée sur la clarté des objectifs, la rationalité de l'évaluation devient rétrospective, basée sur l'apprentissage des conséquences de l'action passée » (Duran, 1999, p. 187). Pour Majone (1980), les politiques publiques étant « des théories du changement social », il s'agit dès lors à travers l'évaluation d'en valider ou non les hypothèses. Du point de vue de Duran et Monnier (1992),

on s'éloigne sans nul doute aujourd'hui d'une pensée qui trouvait dans la planification et la programmation l'expression la plus achevée d'une intervention

publique rationnelle pour une approche davantage centrée sur l'ajustement des processus d'action qui réhabilite de ce fait la part d'aléas mais aussi de liberté propre à toute histoire humaine. (p. 239)

Le mouvement du *New Public Management*, qui se trouve au cœur de la réforme de l'État dans différents pays de l'OCDE, a lui aussi insisté sur la mesure des performances des secteurs publics. Cependant, dans le cadre de cette nouvelle gouvernance, il s'agit moins de s'attacher à une vision globale de l'action publique qu'à un suivi des résultats des entités opérationnelles considérées comme cibles de transformation (Hood, 1991)⁵. Né aux États-Unis, ce courant d'analyse du fonctionnement des administrations s'impose à partir des années 1970 dans un contexte de remise en cause de l'action de l'État. S'installe progressivement, dans les administrations de l'économie et des finances, une orthodoxie budgétaire influencée par les thèses monétaristes (Jobert, 1994; Siné, 2006, cité par Pons, 2008). Dans ce contexte, l'objectif principal est d'assainir les dépenses publiques, donc de rationaliser l'action de l'administration. Face à ce contexte de crise, le *New Public Management* va développer un ensemble de lignes directrices permettant de guider la rénovation de l'administration dans le sens d'un meilleur coût-efficacité.

Quatre principaux piliers du *New Public Management* vont donner à l'évaluation standardisée un rôle très spécifique dans les nouveaux agencements institutionnels :

- les productions des services publics sont mesurables ;
- elles doivent donc être mesurées grâce à des outils spécifiques dont la validité doit être scientifiquement assurée ;
- les acteurs publics opérationnels qui bénéficient de davantage d'autonomie sont redevables de leurs actions devant les gestionnaires du système (c'est le modèle de l'évaluation gestionnaire) et/ou devant les citoyens (modèle de l'évaluation démocratique) ;
- l'organisation publique doit être régulée par les résultats (*outputs*) et non plus exclusivement par des contrôles procéduraux orientés sur les ressources (*inputs*).

5. Le *New Public Management* constitue en quelque sorte une théorie de management à appliquer localement et est donc moins concerné que les autres courants examinés ci-dessus par l'évaluation des choix politiques.

Cadres théoriques en éducation: l'évaluation des acteurs opérationnels est privilégiée

Pour les économistes, dans la lignée de la théorie du capital humain (Becker, 1964), l'éducation est non plus considérée comme un coût, mais comme un investissement dont va résulter un capital. L'agrégation des compétences individuelles va refléter partiellement la force économique des pays. Dans cette perspective, les interrogations sur la réalité des apprentissages, auxquelles les évaluations standardisées cherchent à répondre, sont donc fondamentales. « Dans la logique d'une économie centrée sur "l'innovation" ou "la connaissance", un discours prêchant la nécessité d'investir dans les ressources humaines se fait de plus en plus entendre, qui se traduit dans les demandes de voir le "niveau général d'éducation" s'élever » (Maroy, 2005, p. 18). Cependant, selon les économistes, cette amélioration de la qualité de l'éducation ne doit pas se faire à n'importe quel prix, et ce d'autant plus que la massification de l'éducation par un accès quasi universel au secondaire supérieur dans les pays de l'OCDE pèse sur le budget des États.

Face à ces contraintes budgétaires, les économistes mettent en avant leur capacité à apporter des solutions pour « produire » plus à moindre coût – c'est-à-dire éduquer mieux avec des budgets stables voire moindres –, en définissant les paramètres institutionnels sur lesquels les politiques doivent jouer pour optimiser le système. Pour ce faire, l'analyse économique utilise le paradigme de la fonction de production, importée du monde de l'entreprise, l'entreprise combinant des facteurs de production (les *inputs*) pour élaborer un produit, l'*output*. Appliqué à l'éducation, ce modèle induit la transposition suivante: l'école assemble différents *inputs* (matériels pédagogiques, professeurs aux caractéristiques particulières [nombre d'années de formation, ancienneté, etc.]...) pour produire de l'éducation, qui peut être évaluée d'un point de vue quantitatif (nombre d'années de scolarisation, par exemple) ou qualitatif (mesure des acquis des élèves). Aussi, l'économiste a-t-il besoin de données relatives aux *inputs* et aux résultats des systèmes éducatifs pour faire « tourner » ses modèles statistiques et mettre en évidence les configurations institutionnelles efficaces. Les données relatives aux résultats qualitatifs des écoles (acquis des élèves) sont collectées par des tests standardisés. Les économistes prônent donc un nouveau regard sur le fonctionnement des systèmes éducatifs appuyés par une régulation par les résultats (Woessmann, 2007). Si les politiques peuvent bien être l'objet de l'analyse de l'évaluation (l'autonomie des établissements, l'évaluation des élèves, le niveau des salaires...), le plus souvent l'explication des mécanismes en jeu favorise une vision micro centrée sur les acteurs locaux (par exemple la théorie de l'agence ou du choix publics rationnels dans les travaux de Woessmann, 2007).

Les choix de stratégies et d'organisation au niveau de l'établissement sont également considérés comme cruciaux quant aux résultats des élèves.

C'est le fameux slogan « *School matters* » (« L'école compte ») du courant de la *School effectiveness* (Hargreaves, 1996; Normand, 2006). La qualité du leadership, de la discipline à l'école et en classe, des relations entre enseignants et élèves, etc. sont des facteurs fondamentaux de la réussite scolaire. Empiriquement, les études de ce courant montrent que des écoles accueillant des publics identiques en termes de catégories socioprofessionnelles des parents peuvent présenter des résultats scolaires très différents. Ce mouvement a fortement influencé la construction des nouveaux dispositifs d'évaluation standardisée qui se sont centrés sur la responsabilité des écoles, responsabilité interne envers les autorités décisionnelles ou externe vis-à-vis de la société civile, au détriment de l'évaluation des politiques éducatives et des acteurs qui en ont la charge, comme nous allons le voir dans cette seconde partie.

POLITIQUE D'ÉVALUATION STANDARDISÉE DANS L'OCDE : UNE ÉVALUATION DES ACTEURS POLITIQUE EN RETRAIT

Les concepts de responsabilisation des acteurs ainsi que la mise en œuvre de dispositifs d'*accountability* ont marqué les décennies 1990 et 2000. Parmi ces outils qui revêtent des formes diverses (dispositifs d'évaluation quantitatifs et/ou qualitatifs des établissements, suivi des carrières enseignantes), un instrument s'est plus particulièrement développé, sans doute parce qu'en lien direct au moins formellement avec une des fonctions traditionnelles de l'école : l'évaluation standardisée des élèves. Il s'agit d'administrer, à des cohortes d'élèves fréquentant les classes d'un certain niveau du cursus, une (ou des) épreuve(s) portant sur les connaissances et compétences scolaires, épreuves dont l'élaboration, l'administration et la correction sont homogénéisées (De Landsheere, 1992, 1994).

Cet outil n'est pas nouveau dans les systèmes éducatifs des pays développés. Dès la fin du XIX^e siècle, au Canada, des tests standardisés ont été mis en œuvre à l'entrée ou la fin du secondaire (Traub & Maraun, 1994). La Corée du Sud a aussi développé une forte tradition de tests depuis les années 1950 (INCA, 2009). En Europe, des épreuves nationales uniformisées, souvent à visée certificative, se sont imposées dans les années 1960-1970 au Royaume-Uni, en Irlande, en Suède ou encore en France (Eurydice, 2009). Mais ce sont les années 1990 et la première décennie après 2000 qui marquent à la fois une expansion de cet instrument et un renouvellement de ses usages, par une intégration dans de nouveaux modes de gestion des systèmes éducatifs.

Force est de reconnaître que cet outil d'évaluation est à la fois le plus généralisé dans la famille des instruments d'évaluation en éducation

et constitue – du moins, dans les discours – le socle des dispositifs d'*accountability* mis en place dans un grand nombre de pays. Son analyse offre ainsi un bon observatoire des dispositifs d'évaluation qui se sont développés dans le secteur scolaire. Cet instrument a fait l'objet d'une étude dans le cadre du Réseau européen Eurydice (2009)⁶ sur laquelle repose notre analyse. Nous nous centrons donc ici sur l'étude de cet instrument, de façon à analyser à travers cet outil précis les relations entre la mesure quantitative des performances scolaires et l'évaluation des politiques éducatives, telle que définie en introduction de ce texte.

Des appareils statistiques nationaux pauvres qui permettent difficilement l'évaluation des politiques publiques

L'étude des différentes contributions des pays européens dans le cadre de l'enquête d'Eurydice (2009) ainsi qu'une recherche documentaire menée par nous sur les pays hors de l'Union européenne (UE) permettent plusieurs constats. On observe, premièrement, une quasi-généralisation des dispositifs d'évaluation standardisée des élèves dans les pays de l'OCDE. Deuxièmement, la quasi-totalité des pays qui développent ces dispositifs l'assortissent d'un jeu d'indicateurs de suivi des performances des élèves au niveau national, fédéral ou régional (si des autorités infranationales sont en charge de l'éducation). Troisième constat : même les outils qui ne visent pas en priorité l'évaluation du système éducatif font l'objet d'un travail d'agrégation des résultats et, le plus souvent, d'une présentation dans un rapport national même si ces outils n'ont pas été conçus dans cette perspective et,

6. Le Réseau européen Eurydice, qui réunit des unités opérationnelles dans chacun des pays de l'Union européenne, vise à travers des études thématiques à analyser les cadres légaux et les mises en œuvre des politiques éducatives des pays membres de ce réseau. À partir d'une grille de contenu standardisée élaborée par l'Union européenne, les experts des différents pays renvoient des contributions nationales qui consistent en un corpus conséquent pour chaque étude. Ces contributions sont intéressantes à un triple égard. Descriptives et historiques, elles présentent dans le détail les cadres légaux et la genèse des politiques éducatives conduites dans chacun des pays. S'ouvrant à des références scientifiques, exigées dans les grilles de contenus, elles permettent d'avoir accès à différentes recherches développées en langue nationale et souvent méconnues. Enfin, constituant une littérature grise et non des rapports scientifiques, elles sont aussi un matériau de choix pour l'analyse du discours autour de la genèse et des motivations des politiques éducatives. Cet article est fondé sur l'analyse des contributions des pays européens à l'étude *Les évaluations standardisées des élèves en Europe : objectifs, organisation et utilisation des résultats* (Eurydice, 2009) qui visait à mieux connaître les outils développés sous ce label générique, au primaire et dans le premier cycle du secondaire, dans les différents États membres, la genèse de ces politiques et leurs usages notamment en termes d'intégration de ces instruments de connaissance dans la formulation des politiques éducatives. L'analyse de ce corpus européen a été complétée par une recherche documentaire d'articles scientifiques à la fois sur les pays européens et sur d'autres pays de l'OCDE, notamment les États-Unis.

donc, ne présentent pas a priori de validité pour cet usage. Cela peut être le cas des tests utilisés dans la gestion des parcours scolaires des élèves (pour permettre l'accès à certains niveaux scolaires ou l'orientation vers des filières) ou ceux qui servent aux diagnostics des difficultés d'apprentissage des élèves, comme au Pays-Bas, en Islande, dans certains *Länder* allemands ou en France. Il y a, dans ces cas, confusion quant au niveau des fonctions d'évaluation : des épreuves conçues dans une perspective pronostique ou diagnostique sont utilisées à des fins sommatives. Or, les épreuves pronostiques reposent sur le principe de sélectionner des items discriminatifs et prédictifs de la réussite future dans le niveau scolaire ou la filière considérés, alors qu'une épreuve sommative (ou épreuve-bilan) doit porter sur un échantillon d'items aussi représentatif que possible des connaissances et/ou compétences qui constituent les objectifs du cycle d'enseignement évalué. De même, une épreuve diagnostique suppose que l'on privilégie les items qui permettent d'explorer spécifiquement les composantes d'une séquence d'apprentissage qui constituent des nœuds de difficulté (voir sur ce point les travaux d'Allal, 1979, 1982 ; Cardinet, 1982, 1985, 1986, 1987, 1988 ; Crahay, 2007). Bref, comme l'a défendu il y a longtemps Cardinet (1985), il convient de développer des instruments d'évaluation propres à la fonction qui lui est assignée (pronostique, diagnostique, formative et sommative ou bilan). De même, pour Mottier Lopez et Crahay (2009), entre *la régulation des apprentissages et le pilotage des systèmes*, il y a tension en ce sens que les opérations de mesure adaptées à la première fonction ne sont pas celles qui sont adaptées à la seconde fonction et vice versa. Or, il ressort d'une première analyse des dispositifs d'évaluation standardisée dans l'OCDE qu'il est fait un usage à la fois fréquent et abusif des résultats de certaines épreuves pronostiques ou diagnostiques dans une perspective de description des acquis des élèves au niveau national.

De surcroît, une analyse fine des indicateurs développés au niveau national et de leurs usages (Mons, 2009) met en évidence l'indigence des dispositifs statistiques développés ainsi que la faible utilisation de leurs résultats dans le cadre d'évaluations des politiques éducatives. Ceci a pour conséquence que les répercussions de ces données ou informations sur le fonctionnement des systèmes sont le plus souvent réduites.

Ces outils statistiques apparaissent, tout d'abord, trop incomplets eu égard aux objets appréhendés pour permettre la mise en place de dispositifs réellement opérationnels d'évaluation des politiques éducatives et pour interpeller les décideurs sur la pertinence et l'efficacité de leurs politiques. Les évaluations standardisées portent, premièrement, dans la quasi-totalité des pays de l'OCDE qui en développent, sur un nombre limité de disciplines : généralement, la maîtrise de la langue nationale, les langues étrangères, les mathématiques et, moins souvent, certaines disciplines scientifiques. Rares sont les pays qui présentent un large spectre de disciplines ou d'objets évalués. Notamment, l'éducation civique, l'histoire et la géographie ainsi

que le développement socio-affectif sont souvent les laissés-pour-compte des évaluations nationales. La France fait exception car elle a multiplié les évaluations dites bilan sur un nombre large de matières ou d'attitudes (les compétences et attitudes civiques par exemple). Il en va de même des Pays-Bas, dont le spectre des tests est traditionnellement très large ainsi que du Danemark; ainsi, dans ce pays, en seconde année et à la fin de la scolarité obligatoire, les élèves passent près d'une vingtaine de tests sur des disciplines variées (danois, biologie, géographie, etc.). Dans la majorité des autres pays, les évaluations nationales portent sur trois disciplines (mathématiques, langue d'enseignement et une autre discipline). Ainsi, dans le cadre de l'étude d'Eurydice (2009) qui portait sur 35 États ou régions européennes, dans moins d'un cas sur cinq plus de trois disciplines faisaient l'objet d'un test. Bref, le plus souvent, le spectre des épreuves standardisées apparaît trop étroit pour pouvoir constituer une véritable évaluation des politiques éducatives des États.

Non seulement le spectre des objets évalués est trop strictement circonscrit aux matières dites principales, mais aussi l'analyse du contexte dans lequel se déroulent les apprentissages (caractéristiques des familles, des établissements ou collectivités locales) fait souvent l'objet de collecte d'information très parcellaire voire totalement inexistante. C'est le cas par exemple des caractéristiques socio-économiques des parents des élèves qui ne donnent pas lieu à un recueil de données dans tous les pays⁷. Comment, dans ces cas-là, discuter de l'efficacité des mesures prises pour lutter contre les inégalités sociales de réussite?

De manière générale, l'analyse des appareils statistiques développés dans les pays de l'UE (Eurydice, 2009) révèle une centration sur les résultats des établissements et non des politiques nationales. Cependant, il faut noter qu'en l'absence de données sur les caractéristiques des établissements, il n'est guère possible d'élaborer une analyse permettant d'approcher les raisons pour lesquelles certains établissements se montrent efficaces et d'autres pas. Alors que l'évaluation des politiques éducatives pourrait se pratiquer sur la base d'informations collectées sur des échantillons d'unités scolaires représentatives de la population mère, dans la très grande majorité des pays de l'OCDE, les données disponibles résultent de l'agrégation des résultats de tests administrés à l'ensemble des élèves. Ainsi dans le cadre des 35 États ou régions participant à l'étude d'Eurydice (2009), moins d'un tiers ont développé des enquêtes sur échantillon. Dans l'OCDE, on en trouve en France, en Finlande, dans la Communauté flamande belge ou en Espagne, ainsi qu'aux États-Unis. Le plus souvent, ces enquêtes sur échantillon cohabitent avec des épreuves administrées dans tous les établissements; moins de 10% des États ou régions ont développé des études exclusivement sur échantillon d'élèves. Ce constat interpelle car les opérations impliquant

7. À cet égard, la France, la Finlande, certains États américains et bientôt l'Italie, par exemple, font exception.

l'entièreté de la population sont onéreuses. À moindre frais, il serait possible d'obtenir des informations tout aussi satisfaisantes. En effet, dans le cas d'une étude portant sur un échantillon aléatoire simple suffisamment nombreux, on obtient des estimations de population dont l'erreur de mesure est faible. C'est ainsi que l'on procède dans le cadre des études PISA. Il faut donc se demander pourquoi tant d'États font le choix d'évaluer l'entièreté de la population d'un groupe d'âge alors qu'ils pourraient économiser de l'argent en opérant sur des échantillons représentatifs. Selon nous, ce choix n'est pas anodin ; il traduit un ciblage de l'évaluation des performances scolaires des élèves sur les établissements et les enseignants plutôt que sur l'évaluation des politiques nationales. Bien plus, on trouve peu d'opérations évaluatives qui ont pour finalité explicite de répondre à une question précise concernant une décision politique prise ou à prendre eu égard à l'organisation de l'école. Bien sûr, on trouve des exceptions. Citons, pour l'exemple, le cas de la France qui a planifié une étude pour mesurer l'effet de la réduction des effectifs des classes dans les écoles en zones d'éducation prioritaire (voir la contribution de Bressoux & Lima dans le présent ouvrage), mais ce type d'évaluation est l'exception plutôt que la règle. Au contraire, en général, les épreuves standardisées concernent l'ensemble des établissements (par opposition à un échantillon) et, de surcroît, elles se limitent à un nombre restreint de champs disciplinaires, explorant de façon indigente le contexte familial et les caractéristiques des établissements scolaires.

Quel usage est-il fait des évaluations standardisées ?

Dans presque tous les pays de l'OCDE, des données sont donc disponibles concernant les performances scolaires des élèves. Cependant, il résulte des caractéristiques des appareils statistiques développés au niveau où s'élaborent les politiques éducatives un usage très limité de ces informations. Plusieurs configurations d'usages de ces données dans le cadre d'évaluations de politiques peuvent, néanmoins, être repérées qui de fait ne concernent qu'une minorité de pays de l'OCDE.

Dans certains pays, caractérisés par la mise en place d'objectifs nationaux, l'utilisation des résultats des évaluations se limite à la présentation de statistiques descriptives des acquis des élèves au niveau national ; celles-ci ont un lien plus ou moins affirmé, selon les pays, avec les objectifs poursuivis par certaines politiques en matière d'éducation. On donnera deux exemples de cet usage des évaluations des performances scolaires, usage qui peut être considéré comme la forme la plus simple d'évaluation d'une politique éducative. Ainsi, en France, la Loi d'orientation de 2005⁸ a mis

8. On trouvera le texte complet de cette loi à l'adresse : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000259787&dateTexte=>

en place un socle commun de compétences, ensemble de compétences et connaissances devant être acquis à la fin de la scolarité obligatoire. Désormais, l'atteinte du socle est examinée à travers un test administré à un échantillon d'établissements, à la fin de la scolarité obligatoire. Cet indicateur est intégré dans les documents budgétaires qui, chaque année, permettent à la représentation nationale d'évaluer l'action du gouvernement en matière d'éducation. En Angleterre, on a pu observer une évolution dans le même sens. De fait, l'usage des épreuves standardisées a sensiblement évolué à l'arrivée du gouvernement Blair en 1997. Alors qu'à l'époque Thatcher, les tests visaient principalement l'évaluation des établissements et leur mise en concurrence à travers le palmarès des *League Tables*, ils servent désormais aussi à vérifier l'atteinte de cibles quantitatives déterminées par le ministère de l'Éducation. Des *keys stages* ayant été définis, les données des épreuves standardisées servent aujourd'hui à examiner dans quelle mesure les élèves atteignent désormais le niveau 4 en anglais et en mathématiques, l'objectif étant que 78 % des élèves y parviennent.

Il est un autre usage, celui consistant à recourir aux données des épreuves standardisées soit pour justifier une réforme curriculaire, soit pour l'évaluer. Dans la majorité des pays européens, les tests, étant centrés sur les disciplines dites de base, l'évaluation se concentre, quand elle a lieu, sur des réformes disciplinaires. Ainsi, au Portugal, les résultats des évaluations standardisées ont été utilisés pour lancer un plan d'action pour l'apprentissage des mathématiques en 2006-2007 et du portugais en 2007. De même, en Estonie, Lettonie ou en Roumanie, les programmes scolaires ont été rénovés suite à l'analyse des résultats aux épreuves standardisées. De même, en Allemagne, les évaluations des performances scolaires opérées dans les *Länder*, mais aussi et peut-être surtout le choc provoqué par les résultats des élèves allemands à l'enquête PISA 2000 ont accéléré la recherche d'une coordination entre les *Länder* sur les curricula. Notons que, dans ces différents exemples, les résultats aux épreuves standardisées sont invoqués pour justifier une réforme curriculaire, ce qui ne signifie pas pour autant que l'on utilise ces données pour évaluer l'effet de la réforme introduite. Ceci a été fait en Irlande. Dans ce pays, le test national NAMA⁹ fondé sur un échantillon représentatif vise à analyser les évolutions des acquis des élèves depuis l'introduction du programme d'enseignement primaire et à formuler des recommandations en termes d'enseignement et d'évaluation.

Les résultats des épreuves standardisées ont également été mobilisés – mais de manière ponctuelle – dans le cadre des analyses portant sur les disparités territoriales ou sociales au sein des systèmes éducatifs. Cet usage a ainsi pu donner lieu au développement ou au réaménagement de certaines réformes. Ainsi en France, les résultats des évaluations dites bilan ont été mobilisés dans la régulation de la politique de l'éducation prioritaire, à

9. *National Assessment of Mathematical Achievement.*

travers notamment la création du Réseau ambition réussite pour les établissements recevant les publics les plus en difficulté. En Irlande, les résultats des deux tests nationaux en mathématiques (NAMA) et en anglais (NAER¹⁰) ont également été mobilisés dans la poursuite de la politique de lutte contre les inégalités sociales. En Finlande et en Suède, les résultats aux tests nationaux sont utilisés pour analyser l'évolution des inégalités territoriales et sociales et, en fonction de cette analyse, attribuer les moyens aux écoles.

À l'exception de ces quelques exemples d'usages ponctuels des résultats des épreuves standardisées dans l'évaluation des politiques éducatives, rares sont les pays qui ont mis en place une procédure régulière d'observation et d'analyse des indicateurs quantitatifs produits au départ de ces épreuves. Dans la plupart des pays, on n'observe aucune velléité d'institutionnalisation, sous une forme ou une autre, d'un dispositif d'évaluation des politiques éducatives par leurs effets. Deux systèmes éducatifs font exception : la Communauté flamande de Belgique qui a défini un processus annuel d'analyse des indicateurs produits par ses services, de dialogue entre les parties prenantes et d'élaboration de recommandations sur la base de ces indicateurs, et la Finlande qui a mis en place un appareil statistique complet de mesure des acquis des élèves et d'observation des contextes familiaux et scolaires des apprentissages.

En définitive, on se trouve face à une situation paradoxale : alors que les textes fondateurs de la nouvelle gouvernance mettent en avant une rhétorique centrée sur la mesure de la performance, l'obligation de résultats et la responsabilisation des acteurs, l'analyse fine des dispositifs mis en place et des usages qui sont faits des données recueillies par les épreuves standardisées montre une très faible utilisation de ces données dans la formulation et/ou la régulation des politiques éducatives. Si l'on tient compte de certains dispositifs de pilotage dans lesquels les informations sont utilisées pour réguler l'école (cf. notamment Lafontaine, Soussi & Niedegger, 2009), il semblerait que l'on puisse conclure que ce sont davantage les acteurs opérationnels au niveau des établissements scolaires et des collectivités territoriales que ceux en charge de l'élaboration des politiques éducatives qui sont la cible du mouvement de reddition des comptes. Cette tendance est troublante car l'analyse réalisée par Eurydice (2009) atteste de la généralisation dans les pays de l'UE d'indicateurs statistiques agrégés, présentés notamment dans des rapports descriptifs nationaux. Bref, on observe une tendance très générale à disposer d'indicateurs se situant au niveau macro. Bien plus, cette tendance aboutit dans certains cas à l'utilisation de données issues d'épreuves standardisées initialement dédiées à d'autres usages qu'à une radioscopie du système et donc impropres à renseigner sur l'état des acquis des élèves à un niveau national (les épreuves de diagnostic individuel des difficultés des élèves par exemple). D'un autre côté,

10. *National Assessment of English Reading.*

L'analyse fine des opérations de *testing* des performances scolaires révèle que, bien souvent, les épreuves standardisées portent sur un nombre limité de champs disciplinaires et que le recueil d'informations sur les contextes et processus d'apprentissage est indigent. Enfin, ajouté à ces différents constats, la faiblesse de l'institutionnalisation de leurs usages à l'exception de quelques pays, oblige à conclure à un écart manifeste, en ce qui concerne les acteurs en charge de la formulation des politiques, entre la rhétorique de la responsabilisation et la réalité des dispositifs mis en œuvre. Bref, nos observations rejoignent et généralisent la conclusion de Broadfoot (2000) qui, comparant les systèmes d'*accountability* français et anglais et leurs relations avec l'élaboration des politiques scolaires, écrit : « Il est rare que des propositions spécifiques découlent des informations obtenues » (p. 44).

DISCUSSION

Comment expliquer cet écart entre, d'un côté, la banalisation de la rhétorique de la responsabilisation et de la culture de l'évaluation et, de l'autre, la faiblesse des dispositifs d'évaluation des politiques éducatives utilisant les résultats quantitatifs de l'évaluation standardisée des élèves ?

À cette question, il n'existe assurément pas de réponse unique, valable pour tous les systèmes d'enseignement. Chaque pays a son histoire propre qui influe sur ses modes de fonctionnement et sur la façon dont les politiques éducatives évoluent. Notre ambition ici est de proposer modestement quelques réflexions qui pourront peut-être servir à d'autres analystes à titre d'hypothèses à explorer. Pour éclairer le paradoxe dont nous avons fait le constat ci-dessus – la faible utilisation des résultats des épreuves standardisées des élèves et autres indicateurs dans l'évaluation et la formulation des politiques éducatives et ceci en dépit de la rhétorique de l'*accountability* –, il semble pertinent de replacer cet instrument d'action publique – l'évaluation des performances des élèves – dans le contexte général des transformations des États et des modes de gouvernabilité (Foucault, 1978, pp. 813-828) qui ont touché la très grande majorité des pays de l'OCDE depuis trois décennies. Par instrument d'action publique, nous entendons à la suite de Le Galès (2004),

un dispositif à la fois technique et social qui organise des rapports sociaux spécifiques entre la puissance publique et ses destinataires en fonction des représentations et des significations dont il est porteur. Chaque instrument d'action publique constitue une forme condensée et finalisée de savoir sur le pouvoir social et les façons de l'exercer. (p. 239)

Il en résulte que le choix de ces instruments n'est pas technique mais politique; contrairement à une vision techniciste de l'action publique, il renvoie à une dynamique structurante de reconfiguration de l'État.

Comme l'un de nous l'a souligné ailleurs (Mons, 2009), les dispositifs d'évaluation standardisée des performances scolaires sont à la croisée des nouvelles tendances qui caractérisent les politiques éducatives développées dans les pays de l'OCDE depuis les années 1980 (Mons, 2007). Pour en approcher la compréhension, l'évaluation standardisée doit, en effet, être mise en relation avec quatre évolutions récentes de nos systèmes éducatifs :

- la centration sur une mesure quantitative des apprentissages et la priorité donnée à des objectifs cognitifs au détriment d'objectifs de socialisation larges (Osborn, 2006), en lien avec le développement du concept de compétences dans la veine économiste de la théorie du capital humain;
- le développement d'un nouveau contrôle social des enseignants et des écoles par les responsables administratifs de l'éducation au sens large (districts, municipalités, administrations déconcentrées, régions suivant les pays) dans le cadre le plus souvent de réformes de décentralisation et d'autonomie scolaire (Maroy, 2008);
- l'évolution de la répartition des pouvoirs entre les acteurs centraux ou fédéraux et les responsables locaux qui voient ainsi leurs marges de manœuvre fortement encadrées (Broadfoot, 2000);
- le développement de la responsabilisation de l'École envers le grand public, en général, et les parents, en particulier, dans le cadre de nouvelles relations entre le politique, l'État, l'administration, d'un côté, et la société civile, de l'autre, ces nouvelles relations étant sous-tendues par l'avènement d'une « démocratie du public » dans laquelle la définition du bien commun n'est plus le seul monopole des dirigeants légitimes (Manin, 1996).

Ce contexte général dans lequel s'insèrent les évaluations standardisées explique à la fois la publicisation nationale dont elles sont l'objet et leur faible utilisation dans ce qui pourrait constituer des dispositifs d'évaluation des réformes en éducation, car implicitement elles poursuivent d'autres buts que celui-ci, buts qui s'avèrent concurrents voire contradictoires puisqu'un même outil, dans ce domaine, ne peut poursuivre efficacement plusieurs objectifs. Ces objectifs implicites sont au nombre de quatre.

Dans la quasi-totalité des pays de l'UE (Mons, 2009) et, plus généralement, de l'OCDE, le politique a recours aux statistiques issues des épreuves standardisées tout d'abord pour communiquer en direction du citoyen et de l'usager de l'école en les présentant comme un instrument de transparence et d'évaluation de l'action publique. Deux facteurs donnent une importance nouvelle à ces « instruments d'action publique fondés sur la connaissance » (Pons & Van Zanten, 2008) que sont les évaluations standardisées: d'une

part, le principe de plus en plus prégnant de reddition des comptes sur l'action publique (Duran, 1999) et, d'autre part, le rôle grandissant des espaces publics, dont l'espace médiatique, comme lieu d'échanges ou de controverses sur les politiques publiques constituant selon Jobert (1994) une des caractéristiques de la nouvelle action publique. Tenus de faire preuve de crédibilité, les politiques recourent dans leurs discours aux résultats nationaux des épreuves standardisées et des enquêtes internationales et les offrent aux citoyens, dans des « forums de communication politique » (Jobert, 1994), comme gage de la nouvelle transparence de l'action publique.

Ces outils, comme l'ensemble des instruments d'action publique fondés sur les connaissances – *benchmarking* nationaux, internationaux, etc. – sont aussi utilisés pour légitimer l'action publique et pour démontrer l'assujettissement du politique aux valeurs d'efficacité et d'efficience prégnantes aujourd'hui. À une légitimité du pouvoir qui reposait historiquement sur un système de valeurs partagées (grandes idéologies, doctrines philosophiques, religieuses...) ou sur les qualités individuelles des gouvernants (légitimité charismatique...), s'ajoute une légitimité rationnelle qui repose sur la production de résultats en termes d'action publique et la démonstration par le politique de sa capacité à solutionner des problèmes publics, comme le montrait déjà Weber (1971), à travers le concept d'éthique de responsabilité. Dans ce cadre, en éducation, les résultats des évaluations standardisées sont bien instrumentalisés pour communiquer en direction du citoyen et de l'usager de l'école et, parallèlement, pour attester de l'efficacité gouvernementale en exhibant des statistiques descriptives dans la très grande majorité des cas (cf. l'étude d'Eurydice, 2009) et, partant, assez rudimentaires.

S'il utilise bien ces données quantitatives pour manifester son souci de transparence et pour légitimer son action, le politique procède cependant avec « discernement » dans leur usage. La nécessaire transparence des résultats de l'action publique n'est revendiquée que dans les cas favorables. Dans une enquête sur les usages par le politique des études quantitatives internationales sur les acquis des élèves, en particulier PISA, Mons et Pons (2009b) ont montré dans le contexte français l'ambivalence du politique. Celui-ci n'hésite pas à limiter la communication sur les résultats négatifs de l'enquête qui pourraient être utilisés comme un bilan de son action publique quand la date de publication de celle-ci se rapproche de la fin de son mandat. Au contraire, il publicise largement les données présentées comme une évaluation d'un système éducatif en crise, quand elles peuvent s'accorder avec le tempo d'un programme de réformes. Les mêmes observations ont été réalisées dans les autres pays participant à cette enquête européenne *Knowledge and Policy* (Portugal, Hongrie, Écosse...) ¹¹. De même, dans une analyse bien documentée de la politique menée par l'administration Bush

11. On trouvera les résultats de recherche concernant ces pays à l'adresse: http://www.knowandpol.eu/fileadmin/KaP/content/Scientific_reports/Orientation3

à propos de l'enseignement de la lecture, Terwagne (2008) dénonce les abus et perversions de l'*evidence-based policy* rejoignant en cela Lessard et Saussez (2009). Dans le cas de ce qu'il appelle « la guerre de la lecture », Terwagne (2008) montre que les « preuves » convoquées par le politique sont soigneusement sélectionnées pour servir une option prise d'avance. De surcroît, entre les chercheurs, la compétition et même les conflits font rage : porteurs de thèses tout autant que de données ou de résultats, ils se disputent le privilège d'influer sur les décisions politiques. Chaque camp proclamant détenir la vérité, certains chercheurs n'hésitent pas à qualifier de non-scientifiques ceux qui, sur base d'autres preuves, ne pensent pas comme eux. Et dans cette lutte qui a pour enjeu le pouvoir sur le pouvoir, les équipes de recherche sont également en lutte pour l'obtention des budgets de recherche¹² et, partant, pour circonscrire ce qui relève ou non de la science et, donc, de la preuve.

L'évaluation standardisée sert aussi un troisième objectif le plus souvent non proclamé : la centralisation et la coordination des micro-interventions des acteurs locaux comme l'atteste l'ensemble des études de cas d'Eurydice (2009). Face aux mouvements de la décentralisation ou de l'autonomie scolaire ou dans le cadre des organisations historiquement fédérales, le développement de cet outil vise à encadrer les actions locales et à engager un mouvement de reddition des comptes du local – établissements scolaires et collectivités locales – vers le centre. Le fait que tous les États fédéraux, sans exception au sein de l'OCDE (le Canada, l'Allemagne, la Suisse...), ont développé cet outil est emblématique de cette reconfiguration des États qui réaménagent les relations entre centre et périphérie. L'analyse fine des dispositifs réellement mis en œuvre dans le cadre européen montre que ce n'est pas le principe de responsabilisation en direction de l'usager parents ou du citoyen qui prime mais bien le contrôle administratif interne. On remarquera, en effet, dans l'étude d'Eurydice (2009) un fort décalage entre le nombre d'États ou de régions – majoritaires – qui utilisent les épreuves standardisées comme outils d'évaluation des établissements par les responsables hiérarchiques et la faible publication de ces résultats en direction du grand public (moins d'un cas sur cinq). Au-delà de l'évaluation des établissements, le *testing* a également pour vocation d'orienter les pratiques pédagogiques des enseignants, par une imposition de standards qui se substituent aux programmes scolaires ou viennent les compléter (Mottier Lopez & Crahay, 2009).

Enfin, l'instrument sert aussi, implicitement, à déporter les responsabilités de l'action publique du niveau central vers le local, qui devient ainsi l'objet et l'acteur privilégié de cette évaluation. Si l'on rejoint Mayntz (1993) et

12. Sans oublier que, dans le cas de la lecture, le marché des manuels scolaires est important (en particulier aux États-Unis) et que les maisons d'édition interviennent également dans les questions ayant trait aux méthodes d'enseignement.

le diagnostic d'un problème de connaissance au sein des administrations ou Thélot (2002) et son « effet miroir », faute de pouvoir à la fois fonder et imposer une théorie de socialisation et d'apprentissage à l'école, faute de pouvoir penser et porter des solutions légitimes aux acteurs opérationnels, le politique en serait réduit à présenter les résultats des évaluations en espérant mettre en mouvement les acteurs opérationnels, dans la bonne direction.

Au total, plus qu'un outil de publicité de l'action publique, d'aide à la décision ou d'évaluation des réformes, l'évaluation standardisée s'impose de fait comme un instrument de reconfiguration, de transformation de l'action publique en éducation, visant à faire changer les pratiques pédagogiques et évoluer les rapports de pouvoir entre les acteurs traditionnels dans ce secteur, l'État, les professionnels de l'éducation que sont les enseignants et les personnels d'encadrement, les collectivités locales et les parents.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Allal, L. (1979). Stratégies d'évaluation formative: conceptions psychopédagogiques et modalités d'application. In L. Allal, J. Cardinet & P. Perrendoud (Éd.), *L'évaluation formative dans un enseignement différencié* (5^e éd., pp. 130-145). Berne: Peter Lang.
- Allal, L. (1982). *Cours: Évaluation pédagogique, régulation des processus de formation* (3^e éd., fascicule 1) [Polycopié]. Université de Genève: Centrale des polycopiés.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.
- Broadfoot, P. (2000). Un nouveau mode de régulation dans un système décentralisé: « l'État évaluateur ». *Revue française de pédagogie*, 130, 43-55.
- Cardinet, J. (1982). *La cohérence nécessaire dans le choix des procédures d'évaluation scolaire* (Rapport SIPRI, Cahier 4). Neuchâtel: IRDP.
- Cardinet, J. (1985). Des instruments d'évaluation pour chaque fonction. *Mesure et évaluation en éducation*, 8(1-2), 45-116.
- Cardinet, J. (1986). *Les modèles de l'évaluation scolaire*. Neuchâtel: IRDP.
- Cardinet, J. (1987). De nouvelles bases pour la certification. *Mesure et évaluation en éducation*, 10(3), 5-28.
- Cardinet, J. (1988). *Évaluation scolaire et pratique*. Bruxelles: De Boeck.
- Charbonneau, J. & Estèbe, P. (2001). Entre l'engagement et l'obligation: l'appel à la responsabilité à l'ordre du jour. *Lien social et Politiques*, 46, 5-15.
- Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire?* (3^e éd.). Bruxelles: De Boeck.

- Crahay, M., Audigier, F. & Dolz, J. (2006). En quoi les curriculums peuvent-ils être objet d'investigation scientifique? In F. Audigier, M. Crahay & J. Dolz (Éd.), *Curriculum, enseignement et pilotage* (Collection Raisons éducatives, pp. 7-37). Bruxelles: De Boeck.
- De Landsheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris: PUF.
- De Landsheere, G. (1994). *Le pilotage des systèmes d'éducation*. Bruxelles: De Boeck.
- Demailly, L. (2005). En Europe: l'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. *Éducation et Sociétés*, 15, 105-120.
- Duran, P. (1999). *Penser l'action publique*. Paris: L.G.D.J.
- Duran, P. & Monnier, E. (1992). Le développement de l'évaluation en France. Nécessité techniques et exigences politiques. *Revue Française de Science Politique*, 42(2), 235-262.
- Eurydice. (2009). *Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats*. Bruxelles: Commission Européenne.
- Foucault, M. (1978). *Dits et écrits, 1954-1988: Vol. 3. 1976-1979*. Paris: Gallimard.
- Hargreaves, D. H. (1996). *Teaching as a research-based profession: possibilities and prospects*. (Teaching training agency annual lecture). Londres: Teaching training agency.
- Harris, D. N. & Herrington, C. D. (2006). Accountability, standards and the growing achievement gap: lessons from the past half-century. *American Journal of Education*, 112, 209-238.
- Hood, C. (1991). A Public Management for All Seasons? *Public administration*, 69, 3-19.
- INCA. (2009). *Korea assessment arrangements*. Consulté le 10 mars 2010 dans le site web de Qualifications and Curriculum Development Agency: <http://www.inca.org.uk/1401.html>
- Jobert, B. (1994). *Le tournant néolibéral en Europe*. Paris: L'Harmattan.
- Lafontaine, D., Soussi, A. & Nidegger, C. (2009). Évaluations internationales et/ou épreuves nationales: tensions et changement de pratique. In L. Mottier Lopez & M. Crahay (Éd.), *Évaluation en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (Collection Pédagogie en développement, pp. 61-81). Bruxelles: De Boeck.
- Le Galès, P. (2004). Contrôle et surveillance: la restructuration de l'État en Grande-Bretagne. In P. Lascoumes & P. Le Galès (Éd.), *Gouverner par les instruments* (pp. 237-271). Paris: Presses de Sciences Po.
- Lessard, C. (2006). Le débat américain sur la certification des enseignants et le piège d'une politique éducative: evidence-based. *Revue française de pédagogie*, 154, 19-31.
- Lessard, C. & Saussez, F. (2009). Entre orthodoxie et pluralisme, les enjeux de l'éducation basée sur la preuve, Note de synthèse. *Revue française de pédagogie*, 168, 111-136.

- Majone, G. (1980). Policies as theories. *Omega*, 8(2), 151-162.
- Manin, B. (1996). *Principes du gouvernement représentatif*. Paris : Flammarion.
- Maroy, C. (2005). *Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe ?* (Les Cahiers de Recherche en Éducation et Formation n° 49). Louvain-la-Neuve : Université de Louvain.
- Maroy, C. (2008). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? *Sociologie et Société*, 40(1), 31-55.
- Mayntz, R. (1993). Governing Failures and the Problem of Governability. In J. Kooiman (Ed.), *Modern Governance. New Government-Society Interactions* (pp. 236-252). Londres : Sage.
- Mons, N. (2007). *Les nouvelles politiques éducatives*. Paris : PUF.
- Mons, N. (2009). *Les effets théoriques et réels de l'évaluation standardisée*. Bruxelles : Commission Européenne.
- Mons, N. & Pons, X. (2009a). Pourquoi le pilotage par les résultats? Une mise en perspective théorique et historique de ce nouveau mode de gouvernance. In N. Mons, J.-C. Emin & P. Santana (Éd.), *Piloter par les résultats* (pp. 71-82). Paris : CNDP.
- Mons, N. & Pons, X. (2009b). The reception of PISA in France. A cognitive approach of institutional debate (2001-2008). *Sísifo Educational Sciences Journal*, 10, 27-40.
- Moran, D. J. & Malott, R. W. (2004). *Evidence-based educational methods*. San Diego : Elsevier.
- Mottier Lopez, L. & Crahay, M. (Éd.). (2009). *Évaluation en tension. Entre la régulation des apprentissages et le pilotage des systèmes* (Collection Pédagogie en développement). Bruxelles : De Boeck.
- Normand, R. (2006). L'école efficace ou l'horizon du monde comme laboratoire. *Revue française de pédagogie*, 154, 33-43.
- Osborn, M. (2006). Changing the context of teachers' work and professional development: a European perspective. *International Journal of Educational Research*, 45, 242-253.
- Pons, X. (2008). *L'évaluation des politiques éducatives et ses professionnels. Les discours et les méthodes (1958-2008)*. Thèse de doctorat de science politique, IEP Paris.
- Pons, X. & Van Zanten, A. (2008). *KnowandPol littérature review : Knowledge circulation, regulation and gouvernance* (Rapport de recherche dans le cadre du projet européen KnowandPol). Consulté le 10 février 2011 dans http://www.knowandpol.eu/fileadmin/KaP/content/Scientific_reports/Literature_review/Pons_van_Zanten_EN.pdf
- Terwagne, S. (2008). La guerre de la lecture. 1996-2008. Les années Bush. *Caractères*, 31, 4-32.
- Thélot, C. (2002). Évaluer l'École. *Études*, 397, 323-334.

Traub, R. E. & Maraun, M. (1994). La mesure en éducation au Canada : le passé, le présent et l'avenir. *Mesure et évaluation en éducation*, 17(2), 21-48.

Weber, M. (1971). *Économie et société*. Paris : Plon.

Woessmann, L. (2007). International Evidence on School, Competition, Autonomy and Accountability : a Review. *Peabody Journal of Education*, 82(2-3), 473-497.

La place de l'évaluation dans les politiques éducatives : le cas de la taille des classes à l'école primaire en France

**Pascal Bressoux & Laurent Lima
Laboratoire des Sciences de l'Éducation
Universités Pierre-Mendès-France
et Joseph-Fourier, Grenoble**

En 2010, un ensemble de fiches du ministère français de l'Éducation nationale sont adressées aux inspecteurs d'académie en vue de la rentrée scolaire 2010-2011. L'une d'elles envisage une politique d'augmentation de la taille des classes primaires, et mobilise un ensemble de présupposés quant à la neutralité d'une telle mesure sur les résultats des élèves. En voici un extrait :

La diversité des pratiques dans les départements montre que toutes les marges en termes de taille des classes n'ont pas encore été utilisées. D'une manière générale, les méthodes les plus répandues ont pour objectif l'amélioration ou la préservation des taux d'encadrement dans un contexte d'augmentation relative des moyens dans les classes.

Hors cas ou situations spécifiques, les études et expériences les plus récentes indiquent que la diminution des effectifs dans les classes n'a pas d'effet avéré sur les résultats des élèves et que les très petites écoles ne s'avèrent plus toujours performantes.

La réduction ou le maintien de la taille des classes doit ainsi être réservée aux secteurs relevant de l'éducation prioritaire et être strictement accompagnée (pédagogie) et encadrée (évaluation).

L'augmentation de la taille des classes peut donc être globalement envisagée sans dégradation des résultats des élèves. (Ministère de l'Éducation nationale [MEN], 2010, fiche n° 1, p. 1)

Cet extrait provient de la section « exposé des motifs » de la fiche ministérielle intitulée *Augmentation de la taille des classes dans le premier degré* (version du 4/05/2010). Ces « motifs » justifient la mesure définie en tête de cette fiche, à savoir « contribuer efficacement à une augmentation du nombre moyen d'élèves par classe » (p. 1). Il s'agit de la première d'un ensemble de 13 fiches¹ qui portent sur l'enseignement primaire pour cinq d'entre elles, sur l'enseignement secondaire pour sept d'entre elles, la dernière étant de portée générale². Le but de cet ensemble de fiches consiste à « optimiser l'utilisation des moyens d'enseignement sans dégrader les résultats des élèves » (MEN, 2010, fiche n° 6, p. 1) c'est-à-dire, pour l'essentiel, à dégager les moyens que doivent mettre en œuvre les académies en vue de réduire le nombre de postes dans l'Éducation nationale (obtenir des « gains en emploi » énonce la fiche n° 6, p. 1).

En quoi cette fiche nous intéresse-t-elle dans le cadre de ce numéro de *Raisons éducatives* consacré à la place de l'évaluation dans les politiques éducatives ? Tout simplement en ce qu'elle justifie une politique éducative, au motif explicite (bien que non exclusif) de résultats d'évaluations obtenues par « les études et expériences les plus récentes ». Cette politique est fondée sur plusieurs constats. Tout d'abord, celui que la taille des classes en France n'a guère évolué dans la période récente, passant dans les écoles primaires (préélémentaire + élémentaire + ASH³) de 23,28 élèves par classe en moyenne en 1999 à 23,46 en 2009. Ensuite, un constat de disparité de taille des classes d'un département à l'autre en raison, pour l'essentiel, d'une plus ou moins grande ruralité : 15,13 élèves en moyenne par classe en Lozère, département fortement rural, mais 24,95 en Essonne, département fortement urbanisé. Les départements ruraux ont la caractéristique d'avoir beaucoup plus de petites écoles à classe unique ou à deux classes seulement. De fait, dans ces écoles, la taille des classes est souvent faible. Le regroupement des toutes petites écoles, lorsque leur proximité l'autorise, pourrait conduire à une meilleure optimisation en créant davantage de possibilités de créer des classes à effectifs non faibles. La grande part du budget de l'Éducation nationale étant constituée par les salaires des enseignants, on comprend l'intérêt du ministère pour ces questions : des classes à effectif plus élevé signifient l'emploi de moins d'enseignants, donc une réduction des coûts salariaux. Ainsi, dans la même fiche, le ministère estime que, théoriquement, c'est-à-dire en l'absence de contraintes liées à des impossibilités de regroupements d'écoles, etc., élever la taille moyenne des classes d'une unité « devrait se traduire, au niveau national, par une économie de près de 10 000 classes, soit un peu plus de 4 % du contingent total » (p. 3). On notera que c'est le terme « économie » qui est employé, et

1. Ces fiches sont accessibles à l'adresse : <http://www.cafepedagogique.net/lexpresso/Pages/2010/05/31052010Accueil.aspx>

2. Nous ne traiterons dans cet article que de l'enseignement primaire.

3. ASH concerne l'adaptation scolaire et la scolarisation des élèves handicapés.

non pas simplement celui de « réduction », ou encore de « baisse ». Même si le ministère prévoit que, dans les faits, « la hausse du seuil d'une unité ne pourra se traduire que par une remontée beaucoup plus limitée du nombre moyen d'élèves par classe (0,6 ou 0,7) » (p. 3), l'objectif visé est bien ici de faire des économies.

Évidemment, rien de mal à cela ; quoi de plus légitime que le souci d'un ministère de chercher à réduire ses dépenses ou, plus positivement encore, à les « optimiser » là où il estime pouvoir le faire ? Ce que l'on discutera plus avant dans cet article n'est pas le principe de réalisation d'économies, mais le fondement des motifs sur lesquels reposent ces économies. Il faut rappeler et souligner que l'évaluation n'est pas absente des préoccupations du ministère ; il se soucie donc aussi des effets possibles d'une telle politique sur les résultats des élèves et, pour cela, se fonde sur « les études et expériences les plus récentes ». Lesquelles précisément ? On ne peut le savoir avec certitude ; aucune bibliographie n'est donnée, aucune référence explicite n'est présente dans le texte. Mais on peut émettre une hypothèse qui nous semble raisonnable à ce sujet : une expérience de réduction de la taille des classes ayant été réalisée quelques années auparavant par le ministère, il apparaît plausible que les résultats de cette expérience d'envergure fondent en partie ses constats. Ces résultats ont été présentés dans la *Note Évaluation 05.03* (MEN, 2005). Par ailleurs, un avis du HCEE⁴ (présidé à l'époque par Claude Thélot, lui-même ancien directeur de la DEP⁵) ayant été donné au gouvernement suite au rapport effectué par Meuret (2001) sur la question des effets de la réduction de la taille des classes, on peut là encore raisonnablement supposer que la fiche du ministère est au moins en partie fondée sur cet avis.

Cet article a pour but de questionner le constat établi dans la fiche n° 1 du ministère concernant le manque d'effets avérés de la réduction de la taille des classes sur les acquis des élèves au niveau primaire. Nous montrerons que ce constat est largement erroné. Pour cela, nous allons revisiter le rapport Meuret (2001) ainsi que certains travaux antérieurs majeurs et, surtout, proposer de nouvelles exploitations de l'expérimentation de réduction de la taille des classes réalisée par le ministère de l'Éducation nationale (MEN).

4. Haut conseil de l'évaluation de l'école. Cet organisme devait remettre un avis annuel au gouvernement sur l'état de l'évaluation du système éducatif. Sur tout sujet qu'il souhaitait étudier, le HCEE commanditait un rapport à un expert : « L'avis [du HCEE] était politique : il devait non pas reprendre des éléments d'analyse, mais, au contraire, se concentrer sur les apports et les lacunes des évaluations existantes, et sur les recommandations politiques, pour l'essentiel à destination du ministre de l'Éducation nationale, soit pour développer ces évaluations, soit pour mieux les utiliser dans le pilotage du système éducatif » (Forestier & Thélot, 2007, p. 15).

5. Direction de l'évaluation et de la prospective, rebaptisée depuis Direction de l'évaluation, de la prospective et de la performance (DEPP). L'une comme l'autre est (était) une direction du ministère de l'Éducation nationale.

Notre laboratoire ayant été associé (avec d'autres) à la réalisation de cette expérimentation, il est possible d'explorer les données et d'en livrer des analyses détaillées.

LE RAPPORT MEURET, L'AVIS DU HCEE ET LA NOTE ÉVALUATION DU MEN

De nombreux travaux, d'origines diverses, ont traité la question des effets de la taille des classes sur les acquis des élèves. Si tous ne parviennent pas toujours aux mêmes résultats, un assez large consensus se dégage maintenant sur l'existence d'une relation entre la taille des classes et les acquisitions des élèves (e.g. Wang & Finn, 2000), pour peu qu'on distingue les niveaux scolaires pris en compte, les caractéristiques sociales des publics d'élèves, de même que la méthodologie de recherche employée. Sur ces points, le rapport réalisé par Meuret (2001) rédigé à l'attention du HCEE conclut que :

- Pour être efficace, la réduction doit être importante et amener les classes nettement en dessous de 20 élèves.
- Une réduction importante ne conduit qu'à un effet faible, de l'ordre de 0,2 écart type, soit, pour un élève médian, un gain de 8 rangs sur 100 ou de 2 rangs dans une classe de 25.
- Pour les élèves défavorisés, cet effet devient important. Il peut atteindre 0,4 écart type, soit un gain de 16 rangs sur 100.
- Cet effet perdure longtemps après que les élèves ont rejoint de grandes classes. (p. 29)

L'avis formulé par le HCEE (repris in Forestier & Thélot, 2007) suite au rapport Meuret est que la baisse des effectifs n'entraîne de gains notables que si i) elle est importante, ii) elle se produit aux premiers niveaux de la scolarité et iii) elle concerne les élèves de bas niveau socio-économique :

Ainsi il semble exister un effet positif – mais faible – sur les progrès des élèves, effet observé presque uniquement dans les petites classes de l'enseignement primaire, qui semble ne se produire que si l'on procède à une forte réduction de la taille des classes, et qui n'est vraiment visible que pour les enfants de familles défavorisées. (HCEE, 2001, cité par Forestier & Thélot, 2007, p. 212)

Dans le fil de l'avis émis par le HCEE, le ministère de l'Éducation nationale (MEN) a décidé de mettre en place une vaste expérimentation qui visait à tester les effets d'une forte réduction de la taille des classes dans des écoles accueillant des publics défavorisés, auprès de jeunes élèves de cours préparatoire (CP), première année d'école élémentaire en France. On peut signaler au passage qu'il s'agit là d'une remarquable prise en compte par un ministère d'un avis formulé sur le fonctionnement du système dont il a la charge.

Cette expérimentation a été conduite au cours de l'année scolaire 2002-2003 ; elle visait à réduire la taille des classes expérimentales à 10 élèves (entre 8 et 12 dans les faits). Elle a été mise en place dans 10 académies où la proportion d'écoles en zones d'éducation prioritaire (ZEP) était parmi les plus élevées et où pourrait se dérouler une expérimentation sur 100 classes avec autant de classes témoins servant de comparaison dans l'évaluation des effets produits par la réduction de la taille des classes. Le protocole de l'évaluation sera décrit en détail dans la partie « méthode » lorsque nous réexaminerons ces données.

Comme énoncé *supra*, les résultats et conclusions de l'évaluation produite par le MEN ont été présentés dans la *Note Évaluation 05.03* (MEN, 2005). Cette note insiste fortement sur le peu d'efficacité de la réduction de la taille des classes :

Dans les conditions qui ont été celles de l'expérimentation (sélection non aléatoire de classes situées dans des milieux particulièrement défavorisés, autonomie des enseignants en matière d'évolution de leurs pratiques pédagogiques), l'ensemble de ces constats doivent conduire à estimer qu'une réduction de la taille des classes de CP est, à elle seule, d'un intérêt pratiquement nul. (p. 5)

Le MEN fonde pour l'essentiel son constat sur trois faits. Tout d'abord, des gains jugés faibles en fin de CP : l'expérimentation a conduit à « consacrer des moyens (importants) à cette diminution [des effectifs] pour n'obtenir qu'un léger avantage en termes d'acquis en fin de CP » (MEN, 2005, p. 5). Il faut relever toutefois que les résultats concernant les acquis des élèves sont tirés de tableaux qui présentent les différences brutes des scores des élèves, sans aucun contrôle statistique de quelque variable que ce soit (cf. tableau 3, p. 2). Autrement dit, le MEN fonde ses conclusions à partir d'une hypothèse de comparabilité totale et directe entre les deux échantillons, supposés avoir par conséquent des caractéristiques initiales identiques. Ce que nous aurons l'occasion de remettre en cause. Le MEN fonde aussi ses conclusions sur le fait que les pratiques enseignantes observées semblent avoir été peu modifiées dans les classes réduites et sur le fait que, un an après la fin de l'expérimentation (fin du cours élémentaire première année [CE1]) les élèves des classes expérimentales ont des acquis qui sont comparables à ceux des élèves des classes témoins. Concernant ce dernier point, il faut insister sur le fait que les élèves des CP expérimentaux se sont retrouvés, comme les élèves des CP témoins, dans des classes de CE1 à effectifs « normaux ». Ce résultat est donc peu informant sur le bénéfice que pourraient obtenir les élèves à fréquenter *régulièrement*, au cours de leur scolarité, des classes à faibles effectifs. Toutes sortes de scénarios sont en effet possibles : bénéfiques qui « plafonnent » dans le temps, bénéfiques qui se cumulent ou encore bénéfiques qui s'estompent.

Nous venons de répertorier ici les éléments sur lesquels se fonde, selon toute vraisemblance, la fiche ministérielle adressée aux inspecteurs d'académie qui envisage une politique d'augmentation de la taille des classes. Ce constat est-il tenable ? C'est ce que nous étudions à présent en réexaminant, dans un premier temps, les travaux antérieurs sur les effets de la taille des classes puis, dans un second temps, les données de l'expérimentation CP à effectifs réduits.

UN RÉEXAMEN DES TRAVAUX ANTÉRIEURS

Le rapport Meuret pointait les nombreux cas de recherche où les résultats sont non significatifs, ne produisant donc aucune évidence sur l'effet positif (ou négatif) de la réduction de la taille des classes. Mais il pointait aussi avec raison le lien que ces résultats entretiennent avec la méthode utilisée. Si, en se fondant sur les études expérimentales, les résultats montrent un avantage clair aux classes à effectif réduit, les études corrélationnelles « donnent des résultats moins probants [...]. Certaines études concluent en faveur des grandes classes, d'autres en faveur des petites, la plupart à des effets non significatifs. Les modalités de la baisse, les mesures d'accompagnement, importent sans doute fortement » (Meuret, 2001, p. 29).

Or, les études corrélationnelles offrent des garanties beaucoup moins bonnes quant à la fiabilité de leurs résultats. Les conditions y sont en effet moins bien contrôlées que dans les études expérimentales randomisées. Surtout, on ne sait pas si les enseignants et les élèves qui fréquentent de petites ou de grandes classes sont effectivement comparables. Certes, il est possible de contrôler certaines variables, mais on ne peut être certain d'avoir contrôlé toutes les variables influentes. Les possibles biais de sélection font peser un risque redoutable sur les conclusions à tirer de ces études. C'est pourquoi un effort particulier des chercheurs a porté sur les moyens de contrôler au maximum les biais de sélection.

Ici se situe un nœud important du problème. Il faudrait pouvoir analyser des populations qui ne diffèrent que par le fait de fréquenter de grandes ou petites classes. La méthode qui s'impose de prime abord est la méthode expérimentale randomisée. Dans une même population, des individus sont extraits aléatoirement, qui vont fréquenter soit une grande classe, soit une petite classe. De cette manière, on s'assure que la constitution d'un groupe expérimental et d'un groupe témoin ont les mêmes caractéristiques de départ, aux fluctuations d'échantillonnage près. Dans la pratique, il n'est cependant pas aisé d'allouer aléatoirement élèves et enseignants dans des classes de petite ou grande taille. C'est pourquoi fort peu d'expérimentations ont été mises en œuvre en ce domaine. C'est pourtant ce qui a été réalisé par le projet STAR (*Student Teacher Achievement Ratio*) qui a débuté en

1985 dans le Tennessee (États-Unis). L'étude STAR reste à ce jour la seule expérimentation randomisée de grande envergure qui ait visé à tester les effets de la réduction de la taille des classes sur les acquisitions des élèves. De par son protocole et de par les nombreuses analyses et réanalyses de ses données, les résultats produits peuvent figurer parmi les plus solides et les plus fiables (Finn, 2002 ; Finn & Achilles, 1999 ; Finn, Fulton, Zaharias & Nye, 1999 ; Krueger, 1999). L'une des difficultés auxquelles s'affronte tout chercheur qui veut faire la synthèse des résultats d'un domaine d'étude particulier réside dans la confiance, et donc dans le poids à accorder dans les conclusions finales, à des études dont le niveau de fiabilité et de robustesse des résultats peut être fort varié. Dans le cas précis des effets de la taille des classes, négliger ces questions de fiabilité des résultats revient à placer sur le même plan le résultat d'études corrélacionnelles peu fiables, aux conditions mal contrôlées, et des études dont la qualité de contrôle rend les résultats beaucoup plus solides. Pourtant, force est de constater que, même en se fondant sur de nombreuses études de types et qualités très divers, les méta-analyses montrent malgré tout un effet positif de la réduction de la taille des classes (e.g. Goldstein & Blatchford, 1998 ; Slavin, 1990). On comprend cependant que, de par son importance dans le champ de recherche qui nous occupe, l'étude STAR mérite qu'on s'y attarde pour présenter en détail ses résultats.

En 1985, les enfants qui entraient dans les jardins d'enfants (*Kindergarten*) ont été alloués aléatoirement dans de petites classes (13-17 élèves), dans des classes « pleines » (22-26 élèves), ou dans des classes « pleines » avec un enseignant assistant à temps plein. Les enseignants ont eux-mêmes été affectés aléatoirement dans les classes. Les élèves ont été maintenus dans les mêmes conditions (petite classe, classe pleine, classe pleine avec assistant) pendant une période maximale de quatre années (jusqu'au *grade 3*) avec un nouvel enseignant alloué aléatoirement chaque année.

Plus de 6000 élèves de 329 classes dans 79 écoles ont participé à l'expérimentation la première année et près de 12 000 élèves ont participé à un moment donné aux quatre années d'étude. Plusieurs tests ont été administrés à chaque élève au printemps de chaque année : *Stanford Achievement Test* (SAT) (lecture, *word study skills*, mathématiques), et *Basic Skills First* (BSF)⁶ (lecture, mathématiques). Des données concernant les enseignants et les élèves (e.g., comportement en classe, redoublement...) ont été collectées régulièrement.

Les principaux résultats sont les suivants :

- Les classes à petit effectif ont obtenu des résultats significativement meilleurs dans chaque degré et dans chaque discipline. Pour le SAT,

6. Le BSF est un ensemble de tests *curriculum-referenced* qui ont été développés par l'État du Tennessee.

la différence entre les classes à petit effectif et les classes à effectif « normal » varie de 0,2 à 0,3 écart type dans chaque discipline. Pour le BSF, la différence entre les classes à petit effectif et les classes à effectif « normal » varie de 0,1 à 0,3 écart type dans chaque discipline (Finn, Gerber, Achilles & Boyd-Zaharias, 2001). Krueger (1999), réanalysant les données de STAR, a estimé qu'une réduction de 10% de la taille des classes conduisait à un gain dans les acquisitions de 10% d'écart type. Les effets étaient plus forts pour les élèves qui avaient commencé tôt le programme et qui avaient passé davantage d'années dans les classes à petit effectif.

- Les bénéfices sont plus importants pour les élèves des minorités ou les élèves fréquentant les classes des quartiers défavorisés (*inner-city schools*) que pour les élèves blancs scolarisés dans des écoles non urbaines. Les gains étaient jusqu'à deux ou trois fois plus importants dans le premier cas. Par exemple, la différence entre élèves blancs et noirs au test de maîtrise de la lecture au *grade* 1 était de 14,3% dans les classes à effectif « normal », mais elle était seulement de 4,1% dans les classes à petit effectif. Blancs et Noirs profitaient tous des petites classes mais les Noirs en profitaient davantage. Krueger et Whitmore (2001) ont conclu qu'au *grade* 3, l'écart entre Blancs et Noirs serait réduit de 38% si tous les élèves étaient dans des classes à petit effectif.
- Après le retour de tous les élèves dans des classes à effectif « normal », le bénéfice dû aux petites classes perdurait significativement pour tous les niveaux scolaires suivants et pour toutes les disciplines.
- Même si cela n'intéresse pas directement notre présente étude, signalons que la troisième condition de STAR, celle avec classes « pleines » et enseignant assistant supplémentaire, n'a pas donné de résultats probants. Les enseignants assistants n'ont bénéficié ni aux enseignants (ces derniers se disent même parfois un peu gênés: 71% déclarent préférer enseigner en petite classe plutôt que dans une classe à effectif « normal » avec un assistant), ni aux élèves (on n'enregistre aucun gain, quelle que soit la discipline, ni aucune amélioration du comportement; les élèves des classes « pleines » avec assistant ont des performances moins élevées que celles des élèves des classes réduites, elles ne sont pas meilleures que celles des élèves des classes « pleines » et sont même parfois moins bonnes).

Les études expérimentales telles que STAR impliquent des protocoles très lourds, de nombreuses réticences et obstacles à surmonter, et les pressions des parents pour que leurs enfants intègrent des classes à taille réduite n'ont pas manqué. Ces difficultés expliquent qu'une seule étude d'envergure de ce type ait été réalisée. Par ailleurs, même si les études expérimentales randomisées figurent incontestablement parmi les études qui ont produit les résultats les plus solides, elles ne sont pas sans limites. D'une part, on

ne peut évacuer la possibilité d'un effet Hawthorne (modification des comportements des participants du simple fait qu'ils sont intégrés à une étude) et, d'autre part, on ne peut transférer directement sans questionnement des résultats obtenus dans d'autres systèmes éducatifs au nôtre.

En l'absence de telles études expérimentales randomisées, il est toutefois possible d'employer une autre approche méthodologique par « variable instrumentale » par laquelle il s'agit d'utiliser une variable exogène au processus étudié (la méthode va conduire à ne retenir, parmi toutes les différences de tailles des classes, que celles qui sont dues au hasard). Ces méthodes ont l'avantage d'annuler quasiment la possibilité d'effets Hawthorne. Piketty (2004) a réalisé une étude en utilisant une telle approche. Il a pour cela exploité les données du panel primaire 1997 en France et, dans le fil des analyses réalisées par Angrist et Lavy (1999), il s'appuie sur le fait qu'en école primaire, lorsque l'effectif d'une classe dépasse 30 élèves, on ouvre souvent une deuxième classe, si bien que la taille des deux classes diminue brusquement. Ce processus crée des discontinuités exogènes qu'on peut exploiter pour analyser les effets de la taille des classes sur les acquisitions des élèves. Les estimations réalisées lui permettent de mettre ainsi au jour des gains importants liés à la réduction de la taille des classes, même pour une faible diminution des effectifs. Ces gains sont d'autant plus importants que le public est plus défavorisé. En raisonnant à coûts constants, Piketty indique que réduire la taille des classes ZEP à 18 élèves au lieu de 21,9 dans les faits – ce qui ne conduirait à porter la taille des classes hors ZEP qu'à 24,16 élèves, contre 23,3 dans les faits – réduirait de 40 % les écarts de performances entre classes ZEP et hors ZEP.

Par ailleurs, en restreignant leur analyse aux classes confiées à des instituteurs débutants, pour lesquelles il n'y a plus de relation entre la taille des classes et le niveau d'acquisitions de leurs élèves en début d'année⁷, et pour lesquelles il ne semble donc plus y avoir de biais de sélection, Bressoux, Kramarz et Prost (2009) montrent eux aussi un effet positif de la réduction de la taille des classes, celui-ci étant notablement important pour les élèves les plus faibles.

Ainsi, un large consensus apparaît dans la communauté scientifique sur le fait que la taille des classes exerce un effet sur les acquis des élèves au niveau primaire. La question de l'importance de cet effet est davantage discutée mais la plupart des études, et cela est confirmé par les méta-analyses (e.g. Goldstein & Blatchford, 1998; Slavin, 1990), concluent à un effet positif de la réduction de la taille des classes. Comment, dès lors, comprendre que le MEN énonce que c'est seulement dans des « cas ou situations

7. Les enseignants expérimentés tendent à enseigner dans des classes de meilleur niveau, souvent en ville, dans des classes aux effectifs plus importants, créant ainsi un biais de sélection.

spécifiques » qu'il y a un effet et que, en dehors de ces cas, « les études et expériences les plus récentes indiquent que la diminution des effectifs dans les classes n'a pas d'effet avéré sur les résultats des élèves » ? (MEN, 2010, fiche n° 1, p. 1). Ne s'agit-il pas là d'un argument qui repose sur les résultats de l'expérimentation conduite par le ministère et dont les premiers résultats ont été donnés par la DEPP (MEN, 2005)? Cette expérimentation portait sur des classes en zones d'éducation prioritaire et la DEPP concluait à des effets très modestes de la réduction de la taille des classes alors même que cette réduction était particulièrement forte puisqu'elle amenait les classes expérimentales à un effectif de 10 élèves.

Or, nous avons de bonnes raisons de penser que les conclusions émises par la *Note Évaluation* de la DEPP (MEN, 2005) sous-estiment les effets de la réduction de la taille des classes, en raison de la non-prise en compte de biais de sélection. C'est pourquoi nous proposons à présent une réanalyse des données de cette expérimentation.

UN RÉEXAMEN DES DONNÉES DE L'EXPÉRIMENTATION CP À EFFECTIFS RÉDUITS

Méthode

PARTICIPANTS

Un groupe de 100 classes expérimentales de CP a vu sa taille réduite à 10 élèves par classe. Un groupe de 100 classes témoins, à effectif « normal » a été choisi pour contrôler les effets de l'expérimentation. Dans les faits, les effectifs des 100 classes expérimentales variaient de 8 à 12 élèves par classe avec une moyenne égale à 10,45. Les effectifs des 100 classes témoins variaient quant à eux de 10 à 27 élèves par classe avec une moyenne égale à 21,29 (on relève une classe à 10 et une classe à 14 élèves dans cet échantillon). Toutes les écoles étaient situées dans des milieux défavorisés et appartenaient à une zone d'éducation prioritaire.

Au total, compte tenu des données manquantes et de l'exclusion de la classe témoin qui ne comprenait que 10 élèves⁸, 138 classes ont été retenues pour l'analyse, soit 1021 élèves. On dispose de 72 classes expérimentales (8 à 12 élèves par classe avec une moyenne égale à 10,43), soit 529 élèves, et de 66 classes témoins (14 à 27 élèves par classe avec une moyenne égale à 21,15), soit 492 élèves. L'attrition n'a pas conduit à modifier sensiblement

8. Les raisons qui faisaient que cette classe témoin comptait seulement 10 élèves étaient douteuses; était-ce le résultat de conditions particulières ou bien cette classe avait-elle faussement été attribuée au groupe témoin? En l'absence d'information, nous avons préféré écarter cette classe des analyses. Remarquons toutefois que les estimations produites avec ou sans cette classe sont similaires.

les moyennes d'effectifs par classe, ni les caractéristiques individuelles des élèves et des enseignants ce qui, de ce point de vue, ne crée pas de biais de sélection.

Il est fondamental à ce point de l'analyse de signaler que l'expérimentation n'a pas été randomisée: ni les élèves ni les enseignants n'ont été affectés aléatoirement aux différentes conditions, expérimentale ou témoin.

MATÉRIEL

Des renseignements sociodémographiques (genre, catégorie socio-professionnelle du père [PCS], langue maternelle) et scolaires (âge de début de scolarisation, maintien éventuel en grande section) sur les élèves ont été collectés. Un questionnaire aux enseignants les interrogeait sur leurs pratiques en classe et sur leurs caractéristiques individuelles (âge, genre, ancienneté, etc.).

Les acquis scolaires des élèves ont été évalués grâce à des épreuves standardisées. Ces évaluations portaient pour l'essentiel sur l'acquisition du langage et exploraient de manière spécifique les domaines suivants: compréhension orale, lexique, phonologie, reconnaissance de mots, sémantique, traitements complexes et écriture. Une partie des épreuves était cependant consacrée aux acquisitions en mathématiques. Nous traiterons dans cette étude du score d'acquisitions global des élèves, qui agrège les résultats obtenus par les élèves dans l'ensemble de ces domaines. Ce score global a été centré (moyenne fixée à 0) et réduit (écart type fixé à 1).

PROCÉDURE

Les élèves ont passé les épreuves standardisées évaluant leurs acquisitions en début (octobre 2002) et en fin d'année scolaire (juin 2003). Une évaluation intermédiaire (mars 2003) a aussi été réalisée mais elle ne sera pas prise en considération dans cet article. Dans les classes témoins, ces évaluations n'ont porté que sur 10 élèves choisis aléatoirement dans chaque classe.

Résultats

UNE ANALYSE DES POPULATIONS ÉLÈVES ET ENSEIGNANTS

Les caractéristiques des deux échantillons d'élèves sont très proches, quoique non strictement identiques, tant en ce qui concerne les caractéristiques sociodémographiques que les scores d'acquisitions. Concernant les premières, on enregistre 2 % d'enfants dont le père est cadre supérieur ou de profession intermédiaire dans le groupe expérimental, contre 3 % dans le groupe témoin; 43 % d'enfants dont le père est ouvrier dans le groupe expérimental, contre 36 % dans le groupe témoin. Cela traduit bien le caractère très défavorisé des deux échantillons.

Concernant le niveau initial d'acquisitions, on relève un score moyen de 0,045, avec un écart type égal à 0,98 pour les élèves du groupe expérimental, contre un score moyen de - 0,048 et un écart type de 1,02 dans le groupe témoin. Ces différences ne sont pas statistiquement significatives.

Il en va tout autrement pour les caractéristiques des enseignants en fonction de leur répartition dans les classes expérimentales ou témoins. L'analyse révèle en effet des différences notables pour ce qui concerne leur ancienneté, comme le montre le tableau 1. L'ancienneté d'enseignement en CP est moins élevée dans les classes à effectif réduit (la moyenne $M = 3,06$) que dans les classes témoins ($M = 7,65$). L'ancienneté dans la profession est moins élevée dans les classes à effectif réduit ($M = 8,28$) que dans les classes témoins ($M = 14,17$). Ces différences sont statistiquement significatives tant pour l'ancienneté d'enseignement en CP ($t = 4,22$; $p < .0001$) que pour l'ancienneté dans la profession ($t = 3,59$; $p < .001$).

Tableau 1 : Caractéristiques des enseignants selon CP réduit ou classes témoins

Variables	N	Moyenne	Écart type	Minimum	Maximum
CP réduit					
Ancienneté en CP	72	3,06	4,29	0	24
Ancienneté dans la profession	72	8,28	7,71	0	35
Classes témoins					
Ancienneté en CP	66	7,65	8,09	0	30
Ancienneté dans la profession	66	14,17	11,35	1	39

On peut considérer que ces différences sont dues à l'allocation non aléatoire des enseignants dans les classes expérimentales ou témoins. Une conséquence importante est qu'on ne peut pas tenir pour comparables les enseignants des deux groupes d'élèves. Or, il apparaît raisonnable de penser que l'ancienneté n'est pas indépendante de l'efficacité des enseignants dans la mesure où l'expérience doit favoriser une maîtrise des « gestes » du métier. Bref, il y a un biais de sélection évident et celui-ci risque d'affecter les estimations.

La DEPP ayant laissé les académies autonomes dans la répartition des écoles et enseignants, on ne peut pas connaître absolument le processus

par lequel l'allocation s'est effectuée. Peut-être a-t-on jugé localement que les enseignants les plus novices étaient ceux à qui devaient être confiées les classes expérimentales puisque, allégées en effectifs, il serait plus facile d'y enseigner. Mais il peut aussi s'agir de raisons autres (elles-mêmes peut-être pour partie liées à l'ancienneté). De plus, il faut remarquer que, bien que nous ayons pu capter certaines caractéristiques qui différencient les deux échantillons d'enseignants, nous ne pouvons pas assurer les avoir repérées toutes. Il sera ainsi impossible de contrôler ces éventuelles différences inobservées.

ESTIMATION DES EFFETS DE L'EXPÉRIMENTATION SUR LES ACQUIS DES ÉLÈVES

Dans la mesure où les élèves sont « emboîtés » dans des classes et que cette structure est susceptible de générer des corrélations intra-classes, nous utilisons des modèles multiniveaux (Bressoux, 2008).

Nous avons d'abord construit un modèle « vide » (i.e. sans variables explicatives) afin de décomposer la variance totale des scores d'acquisitions des élèves en une variance inter-classes et une variance intra-classes. Les estimations de ce modèle « vide » (modèle 1, tableau 2) révèlent une part de variance inter-classes de 19%. C'est dire que la plus grande part de la variation des acquisitions se situe entre les élèves à l'intérieur des classes mais qu'il existe aussi des différences, non négligeables et significatives, entre les classes. Il s'agit ensuite de modéliser ces variances afin de faire apparaître quels sont les facteurs qui influencent les acquis des élèves.

L'introduction, dans le modèle 2, du score initial des élèves contribue à réduire les parts de variance tant inter qu'intra-classes. Le score initial explique 58% de la variance intra-classes et 39% de la variance inter-classes. On note cependant que la part de variance inter-classes demeure significative, montrant ainsi qu'une grande part des différences finales entre classes ne sont pas dues à des différences initiales de performances. Des spécifications non linéaires n'ont pas conduit à un meilleur ajustement du modèle.

Le modèle 3 introduit un effet aléatoire du score initial. La variance des pentes est significative ce qui montre que la relation entre le score initial et le score final varie d'une classe à l'autre. Certaines classes sont donc plus égalisatrices que d'autres. Nous n'avons pas intégré la covariance entre les constantes et les pentes car un modèle intermédiaire a montré qu'elle n'était pas significative. Ainsi, dans cet échantillon, les classes les plus efficaces ne sont pas forcément les plus égalisatrices.

Afin de cerner au mieux les caractéristiques des élèves, le modèle 4 introduit la profession du père (ni le genre des élèves, ni d'autres caractéristiques individuelles renseignées dans l'étude ne jouent d'effet significatif, elles ne

sont donc pas introduites dans le modèle présenté). On y voit que les enfants de cadres supérieurs et professions intermédiaires ont, à score initial donné, de meilleurs scores de fin d'année que les enfants d'employés, ouvriers et « autres ». La différence est non significative avec les enfants d'agriculteurs et artisans (mais ceux-ci n'étaient pas nombreux dans l'échantillon, ce qui peut expliquer cette non-significativité). La profession du père ne réduit toutefois que très peu les parts de variance résiduelle intra et inter-classes.

Tableau 2 : Modèles expliquant les scores globaux d'acquisitions de fin d'année

Paramètres	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Effets fixes				
Constante	0,011 (0,047)	0,008 (0,035)	- 0,004 (0,035)	0,150 (0,061)*
Score initial global		0,735 (0,020)***	0,732 (0,026)***	0,727 (0,026)***
Profession du père (référence = cadre sup. et profession intermédiaire)				
Agriculteur/artisan				- 0,083 (0,096)
Employé				- 0,198 (0,078)*
Ouvrier				- 0,152 (0,062)*
Autre				- 0,213 (0,064)***
Effets aléatoires				
Variance inter-classes	0,191 (0,037)***	0,116 (0,020)***	0,118 (0,021)***	0,117 (0,021)***
Variance des pentes du score initial			0,035 (0,010)***	0,034 (0,010)***
Variance inter-élèves	0,806 (0,038)***	0,340 (0,016)***	0,307 (0,016)***	0,303 (0,015)***
-2 log L (déviante)	2814,20	1966,69	1940,36	1927,82

N = 1021

*p < .05; **p < .01; ***p < .001

Après avoir contrôlé l'ensemble des caractéristiques des élèves disponibles dans cette étude et qui exerçaient un effet significatif sur les acquis de fin d'année, il demeure une part de variance inter-classes inexpliquée qui n'est pas négligeable. Cela incite donc à rechercher quelles variables sont susceptibles d'expliquer ces différences de performances entre les classes et, tout particulièrement, si l'expérimentation de réduction de la taille des

classes est susceptible de contribuer à cette explication. C'est à cette tâche que s'emploient les modèles 5 à 8 (tableau 3).

Le modèle 5 introduit l'indicatrice de l'expérimentation. Cette variable (nommée « CP réduit » dans le tableau 3) exerce un effet significatif mais modeste sur les scores de fin d'année (le coefficient vaut 0,136). On pourrait être tenté à ce stade de conclure que la réduction de la taille des classes conduit certes à des effets positifs pour les élèves, mais que ces effets sont négligeables, dépassant à peine un dixième d'écart type. Cela reviendrait toutefois à supposer, au moins de manière implicite, que les échantillons d'élèves et d'enseignants sont comparables d'un groupe à l'autre. Concernant les élèves, nous avons vu d'une part que les caractéristiques observées des deux échantillons étaient très proches. D'autre part, ces caractéristiques ont été intégrées dans le modèle et corrigent donc les petites fluctuations existantes. Même si on ne peut assurer que les deux groupes d'élèves sont rigoureusement identiques, il reste que le contrôle de variables telles que le niveau initial et l'origine sociale permet sans doute « d'attraper » une bonne partie des différences qui ont pu être générées par le caractère non aléatoire de l'allocation des élèves dans les conditions expérimentale ou témoin. Il en va tout autrement des caractéristiques des enseignants : non seulement elles diffèrent notablement d'un échantillon à l'autre mais, de plus, elles n'ont pas été contrôlées dans les modèles. Le biais de sélection étant avéré, il faut bien sûr en tenir compte dans les estimations. C'est ce à quoi s'attachent les modèles suivants (modèles 6 à 8, tableau 3).

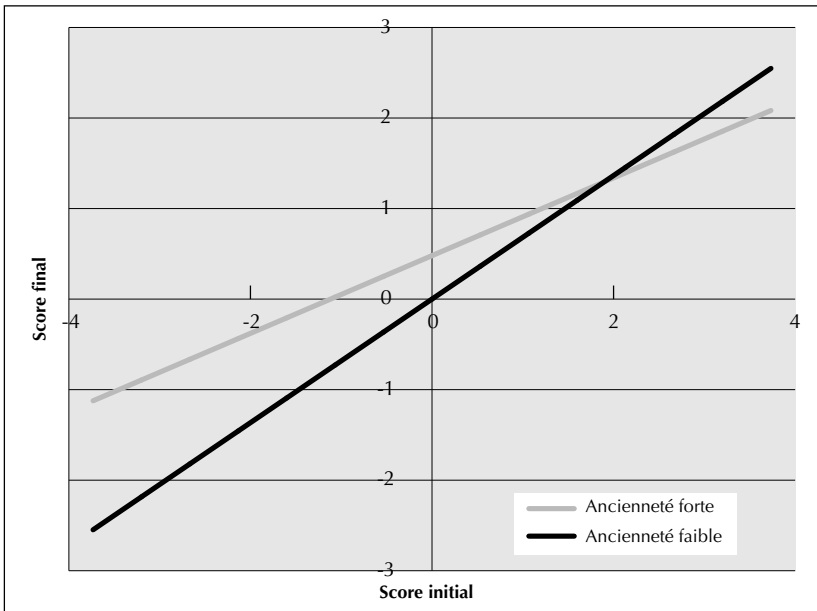
Après avoir testé alternativement les effets des différentes variables qui caractérisent l'ancienneté des enseignants (en CP, dans la fonction, dans l'école), nous avons retenu pour les modèles définitifs l'ancienneté en CP, qui se révèle la plus pertinente pour expliquer les acquis des élèves.

Le modèle 6 montre que l'introduction de la variable « ancienneté en CP » améliore l'ajustement du modèle aux données (décroissance de la déviance = 9,29; $p < .005$). La variance inter-classes résiduelle est quelque peu réduite. L'ancienneté en CP exerce un effet significatif et positif sur les scores de fin d'année; les enseignants qui bénéficient de plus d'ancienneté au CP sont en moyenne les plus efficaces. Si cet effet positif de l'ancienneté n'est pas surprenant à proprement parler, c'est surtout ce que l'introduction de cette variable produit sur le coefficient de l'indicatrice de l'expérimentation qui nous intéresse ici au premier chef. On voit en effet que le coefficient de l'indicatrice de l'expérimentation est beaucoup plus fort après avoir contrôlé l'ancienneté en CP des enseignants. L'effet de l'expérimentation passe ainsi de 0,136 (modèle 5) à 0,209 (modèle 6), soit une hausse de 54%. L'effet de l'expérimentation apparaît ainsi nettement plus fort. Cela provient du fait qu'on ne confond plus les effets de la réduction de la taille des classes avec l'ancienneté des enseignants. Or, attendu que i) cette variable joue un effet significatif sur les scores de fin d'année et ii) les classes réduites

avaient en moyenne des enseignants moins expérimentés, cela conduisait à sous-estimer les effets de la réduction de la taille des classes.

On n’a pas relevé d’interaction significative entre l’expérimentation et le niveau initial des élèves. Les élèves bénéficient dans la même mesure de la réduction de la taille des classes, quel que soit leur niveau initial. Cela peut sembler étonnant du fait que la plupart des études montrent un effet bénéfique plus important de la réduction de la taille des classes en faveur des élèves faibles ou défavorisés. Il faut toutefois se souvenir que les élèves de l’échantillon sont globalement très défavorisés et d’un niveau faible. Une variabilité plus grande dans l’échantillon, incluant par exemple des écoles de quartiers plus favorisés, aurait peut-être permis la détection d’effets différentiels de la réduction de la taille des classes en fonction du niveau socio-économique ou d’acquisitions des élèves.

Graphique 1 : Relation entre score initial et score final en fonction de l’ancienneté en CP des enseignants



En revanche, nous avons relevé une interaction significative entre le score initial et l’ancienneté des enseignants (modèle 7). Plus les élèves sont faibles, plus l’effet positif de l’ancienneté sur les acquisitions est important. En d’autres termes, les élèves faibles sont ceux qui bénéficient le plus de l’expérience des enseignants. En conséquence, les enseignants les plus

expérimentés ne sont pas seulement plus efficaces en moyenne, ils sont aussi plus égalisateurs (voir graphique 1). La prise en compte de cette interaction permet une meilleure spécification du modèle susceptible d'améliorer les estimations des autres variables introduites dans le modèle. C'est en effet ce qu'on constate puisque cela conduit à accroître encore l'effet de l'expérimentation qui passe de 0,209 (modèle 6) à 0,216 (modèle 7). Augmentation certes modeste, mais qui conforte l'importance de la réduction de la taille des classes.

L'indicatrice de l'expérimentation est une variable dichotomique qui introduit une rupture entre les « grandes » classes et les « petites » classes. L'écart moyen d'effectifs entre ces classes « grandes » (témoins) et « petites » (expérimentales) est de plus de 10 élèves par classe (10,72 exactement). Si l'on divise le coefficient issu du modèle 7 par l'écart moyen d'effectifs, on peut dire qu'un élève de moins par classe conduit à un gain d'acquisitions de 2,0 % d'écart type⁹.

On peut toutefois trouver une manière plus directe d'estimer l'effet marginal de la diminution d'un élève : il suffit pour cela de remplacer l'indicatrice de l'expérimentation par l'indicatrice de la taille des classes (i.e. prendre le nombre réel d'élèves par classe, indépendamment du groupe expérimental ou témoin). Par ailleurs, on peut toujours dans ce genre d'expérimentation craindre un éventuel effet Hawthorne : les enseignants des classes expérimentales n'auraient-ils pas modifié leur comportement du simple fait qu'ils savaient qu'ils bénéficiaient de ces conditions expérimentales ?¹⁰ En d'autres termes, leur meilleure efficacité serait-elle due à des causes exogènes à la taille des classes ? Nous nous sommes attachés à répondre à ces deux interrogations.

Le modèle 8 montre l'effet associé au nombre d'élèves (traité, donc, comme une variable quantitative qui varie de 8 à 27). Le coefficient associé à cette variable montre que le gain entraîné par la diminution d'un élève est de 0,022 soit 2,2 % d'écart type, une estimation tout à fait similaire à celle produite par l'indicatrice de l'expérimentation. On ne relève donc pas de différence notable entre les estimations produites par ces deux variables.

Il n'y a évidemment rien là de très étonnant dans la mesure où cela provient de la structure des données. Il est toutefois possible de tester plus directement l'éventualité d'un effet Hawthorne. Il suffit pour cela de tester l'effet de la taille des classes sur le seul groupe témoin. La variation de la

9. $(0,216/10,72)*100$.

10. Remarquons toutefois que les deux échantillons d'enseignants ont été visités à deux reprises en situation de classe et que, outre l'intérêt de prélever des données sur les pratiques effectives, cela plaçait les deux groupes d'enseignants en condition d'observation. On peut penser que cela a limité la possibilité d'émergence ou l'intensité d'un effet Hawthorne.

taille des classes (de 14 à 27) y est suffisamment importante pour qu'on puisse estimer son effet sur les apprentissages. De plus, ces enseignants ont été placés dans des conditions de classe « normales » pour lesquelles aucun effet Hawthorne (tout au moins aucun effet Hawthorne différentiel) ne peut être suspecté. Les estimations montrent un effet de la taille des classes de $-0,031$ qui, s'il n'est significatif qu'en tendance ($p < .09$) en raison d'un échantillon faible, se révèle plus élevé que celui estimé sur l'ensemble de l'échantillon. Sur les seules classes témoins, la réduction d'un élève par classe entraîne un gain d'acquisitions estimé à 3,1 % d'écart type. Il n'y a donc aucune raison de suspecter un quelconque effet Hawthorne dans les résultats produits par l'expérimentation CP à effectifs réduits.

À noter qu'on ne relève pas d'effet significatif du nombre d'élèves par classes dans le groupe expérimental, sans doute en raison d'une étendue trop faible (8-12 élèves). C'est aussi probablement ce qui explique l'effet légèrement plus fort que l'on enregistre sur le groupe témoin par rapport à celui enregistré sur l'ensemble de l'échantillon.

Tableau 3 : Modèles expliquant les scores globaux d'acquisitions de fin d'année

Paramètres	Modèle 5	Modèle 6	Modèle 7	Modèle 8
<i>Effets fixes</i>				
Constante	0,078 (0,071)	- 0,042 (0,080)	- 0,045 (0,080)	0,408 (0,109)***
Score initial global	0,726 (0,026)***	0,728 (0,026)***	0,773 (0,032)***	0,773 (0,032)***
Profession du père (référence = cadre sup. et profession intermédiaire)				
Agriculteur/artisan	- 0,082 (0,096)	- 0,085 (0,096)	- 0,084 (0,096)	- 0,091 (0,096)
Employé	- 0,190 (0,077)*	- 0,196 (0,078)*	- 0,189 (0,078)*	- 0,191 (0,078)*
Ouvrier	- 0,154 (0,062)*	- 0,154 (0,062)*	- 0,154 (0,062)*	- 0,158 (0,062)*
Autre	- 0,213 (0,063)***	- 0,217 (0,063)***	- 0,218 (0,063)***	- 0,223 (0,063)***
CP réduit	0,136 (0,069)*	0,209 (0,071)**	0,216 (0,071)**	
Nombre d'élèves par classe				- 0,022 (0,006)***
Ancienneté en CP		0,016 (0,005)**	0,016 (0,005)**	0,018 (0,005)**
Score initial global*Ancienneté en CP			- 0,009 (0,004)*	- 0,009 (0,004)*

Effets aléatoires				
Variance inter-classes	0,113 (0,020)***	0,102 (0,019)***	0,104 (0,019)***	0,100 (0,018)***
Variance des pentes du score initial	0,034 (0,010)***	0,033 (0,010)***	0,030 (0,009)***	0,030 (0,009)***
Variance inter-élèves	0,303 (0,015)***	0,304 (0,016)***	0,303 (0,015)***	0,303 (0,015)***
-2 log L (déviante)	1923,99	1914,70	1909,69	1906,28

N = 1021

*p < .05; **p < .01; ***p < .001

Discussion

La présente étude montre un avantage manifeste aux classes réduites par rapport aux classes à grand effectif. Nous ne pouvons donc que nous inscrire en faux face à la fiche adressée par le ministère aux inspecteurs d'académie qui énonce le non-effet (« pas d'effet avéré ») des tailles de classes. Certes, la fiche du ministère précise que ce non-effet se produit « hors cas ou situations spécifiques », mais peut-on assimiler les ZEP, qui représentent plus de 10 % des écoles en France à un « cas » ? De plus, partant du constat d'une absence d'effet avéré de la *diminution* des effectifs (constat largement critiquable comme nous l'avons vu), la fiche affirme qu'*augmenter* la taille des classes peut se faire sans conséquences sur les résultats des élèves sans plus faire référence aux spécificités indiquées *supra* : « L'augmentation de la taille des classes peut donc être *globalement* envisagée sans dégradation des résultats des élèves » (souligné par nous) (MEN, 2010, fiche n° 1, p. 1).

Du réexamen des données de l'expérimentation CP à effectifs réduits, il apparaît de manière très claire que l'expérimentation souffre de biais de sélection importants en ce qui concerne les échantillons d'enseignants. La non-prise en compte de ces biais de sélection conduit à sous-estimer fortement l'impact de la réduction de la taille des classes : entre le modèle 5 qui ne contrôle pas ces biais de sélection et le modèle 7, on enregistre un accroissement de l'effet de près de 60%. Or, la *Note Évaluation* de la DEPP (MEN, 2005) ne tenait pas compte de ce biais de sélection. C'est dire que l'effet qui a été publié dans cette *Note Évaluation* est très largement sous-estimé.

Prendre en compte l'ancienneté en CP des enseignants permet de contrôler une partie des biais de sélection qui résultent de l'allocation non aléatoire des enseignants aux conditions expérimentale et témoin. Il est

cependant fondamental d'insister sur le fait que ce processus d'allocation des enseignants aux conditions expérimentale et témoin nous est inconnu. Non seulement il ne résulte pas d'un tirage aléatoire, mais nous ne pouvons pas le connaître entièrement. Le fait d'avoir contrôlé l'ancienneté des enseignants est crucial pour cerner le mieux possible les effets de la réduction de la taille des classes, mais nous ne pouvons assurer que tous les biais de sélection ont été ainsi contrôlés. Nous pouvons tout au plus espérer que peu de biais perturbent nos estimations mais nous sommes dans l'incapacité de l'assurer.

En ce sens, nous devons sans doute admettre que, sauf à supposer que des éléments du processus d'allocation jouent en sens inverse (du point de vue de leurs effets sur les scores des élèves) à celui que nous avons détecté, l'estimation des effets de la réduction de la taille des classes que nous produisons ici est probablement une estimation basse de cet effet. Il nous est cependant impossible de connaître l'ampleur de cette éventuelle sous-estimation.

Nous pouvons toutefois rapprocher nos estimations de celles produites par d'autres études qui offraient de meilleures garanties de contrôle des biais de sélection. C'est le cas de l'étude de Piketty et Valdenaire (2006). Nos estimations, en termes de taille d'effet (i.e. évolution du score d'acquisitions par élève supplémentaire, en pourcentage d'écart type) sont très proches de celles enregistrées par Piketty et Valdenaire (2006) en ce qui concerne les acquisitions en français au CE1¹¹. Leur estimation sur les scores de français (à savoir – 2,2 % par élève supplémentaire) est même rigoureusement identique à celle que nous enregistrons dans le modèle 8. Cela accroît en apparence la « plausibilité » de nos résultats mais un examen approfondi de leurs résultats laisse grande ouverte la possibilité d'effets sous-estimés dans notre étude. La nôtre et la leur se distinguent en effet sur un certain nombre de points qui ne peuvent les rendre directement comparables.

Notre étude porte uniquement sur des classes de ZEP, ce qui n'est pas le cas de celle de Piketty et Valdenaire. Or, l'effet qu'enregistrent Piketty et Valdenaire en ZEP est beaucoup plus fort puisque le coefficient qu'ils retiennent est de 0,70 en ZEP contre 0,25 hors ZEP (Piketty & Valdenaire, 2006). Il ne s'agit toutefois pas d'un test direct sur les écoles ZEP *versus* hors ZEP, leurs tailles d'échantillons ne leur permettant pas d'estimations robustes¹², mais de coefficients obtenus en se concentrant sur les sous-échantillons constitués soit par l'ensemble des enfants dont la PCS parentale est défavorisée, soit par l'ensemble des enfants dont la PCS parentale est

11. En fait, il s'agit de l'effet estimé de la variation de la taille des classes de CE1 sur les résultats aux évaluations nationales de début de CE2.

12. Les estimations réalisées (uniquement en mathématiques) montrent un effet en ZEP plus de trois fois supérieur à l'effet hors ZEP (bien que le premier soit non significatif en raison d'effectifs trop faibles).

favorisée. L'apparente similitude de nos estimations globales masque donc en réalité de très grosses différences si l'on se restreint aux écoles ZEP.

Notre étude porte sur la première année élémentaire, la leur sur la deuxième¹³. Plusieurs travaux montrent un effet plus fort lorsque le niveau scolaire est plus précoce. Nos deux niveaux de classe sont toutefois très proches et, si différence il y a, elle devrait être très modeste et sans doute à l'avantage des CP.

Enfin, Piketty et Valdenaire notent un effet plus fort en mathématiques qu'en français. Dans le cas des mathématiques, ils relèvent un effet de $-3,0\%$ d'écart type par élève supplémentaire. Les épreuves d'acquisitions de notre étude portaient elles aussi pour l'essentiel sur le français, mais elles comportaient en plus un sub-test de mathématiques. Notre score global incorpore les résultats des élèves à ce sub-test de mathématiques qui devrait, si l'on suit les estimations de Piketty et Valdenaire, plutôt contribuer à augmenter l'effet. Afin de vérifier cela, nous avons estimé l'effet de la taille des classes sur le seul score de mathématiques. Le modèle (analogue dans sa spécification au modèle 8) produit en ce cas une estimation de $-1,8\%$ d'écart type par élève supplémentaire, inférieure à celle produite sur le score global. L'insertion du score de mathématiques dans ce dernier ne contribue donc pas à gonfler l'estimation de l'effet de la taille des classes sur les acquis en français.

Notre réexamen de l'expérimentation CP à effectifs réduits ne permet pas de conclure avec assurance sur la question de l'ampleur des effets de la réduction de la taille des classes. Une étude randomisée, qui aurait pour conséquence de contrôler les biais de sélection, serait ici fondamentale. Considérer les effets que nous avons dégagés comme importants ou non dépendra en grande partie du fait qu'on envisagera ou non des solutions alternatives. Ainsi, nos estimations montrent que, pour les élèves de niveau moyen, deux années d'expérience supplémentaires de leur enseignant en CP dépassent l'effet d'un élève en moins par classe. Pour les élèves faibles, une seule année supplémentaire suffit pour dépasser cet effet¹⁴. Des considérations en termes de rapport coût-efficacité entrent bien sûr en jeu : réduire de manière généralisée la taille des classes coûterait cher – de la même manière d'ailleurs qu'augmenter leur taille de manière généralisée permettrait d'économiser beaucoup (cf. fiche du ministère aux inspecteurs d'académie) – mais Piketty (2004) et Piketty et Valdenaire (2006) ont montré qu'on pouvait jouer sur la taille des classes en raisonnant à coûts constants : réduire la taille des classes en ZEP n'aurait qu'un faible impact sur la taille

13. Leurs données sur les CP sont insuffisantes pour produire des estimations fiables et distinctes des CE1.

14. D'après le modèle 8, une année supplémentaire d'ancienneté d'enseignement en CP se traduirait par un gain de $0,018 + 0,009 = 0,027$ point d'acquisitions pour un élève situé à un écart type en dessous de la moyenne aux épreuves initiales.

des classes hors ZEP et permettrait de réduire les écarts d'acquisitions entre ces deux types de classes de façon notable. En cela, Piketty s'oppose clairement à l'idée selon laquelle seule une réduction très importante des tailles de classes pourrait avoir un effet notable.

CONCLUSION

Tout aurait pu faire de cette question des effets de la taille des classes un modèle de lien entre évaluation et politique. L'appropriation par le HCEE de cette question tout d'abord, qui donne lieu à un rapport rédigé par un chercheur qui fait un état précis des connaissances sur le sujet. Un avis, ensuite, transmis au gouvernement par un organisme reconnu et indépendant, ce même HCEE, sur la base du rapport commandité. La saisie très rapide et sérieuse par le ministre de l'époque de cet avis et la décision de mettre en place une évaluation d'envergure (rare par la mobilisation qu'elle a dû générer et les moyens déployés) des effets d'une réduction forte de la taille des classes. L'association de la DEPP avec plusieurs laboratoires de recherche pour mener à bien cette expérimentation – hélas sans que ceux-ci puissent intervenir dans le protocole d'échantillonnage – et qui ont contribué à la construction des épreuves standardisées, des protocoles d'observations en classe, des questionnaires, etc.

Pourtant, on constate que la fiche envoyée par le ministère aux inspecteurs d'académie est en contradiction avec ce que l'on sait actuellement des effets de la taille des classes sur les résultats des élèves primaires. Plusieurs éléments peuvent être pointés à cet égard.

Tout d'abord, une expérimentation qui rate l'opportunité de dire des choses fortes et robustes sur les effets de la réduction de la taille des classes, faute de randomisation des échantillons. On pourrait souhaiter à cet égard que le ministère soit en mesure de s'emparer d'une question vive, récurrente, sur l'efficacité du système éducatif (taille des classes, classes à plusieurs niveaux, formation des enseignants, méthodes d'apprentissage de la lecture, etc.) et qu'il décide de conduire une évaluation qui permette de fournir des réponses robustes et rigoureuses aux questions posées. Cette évaluation serait confiée aux organismes les plus compétents (par appel d'offres ouvert, ou en faisant directement appel aux laboratoires et instituts spécialistes de la question) et devrait fournir les informations de la meilleure qualité possible, compte tenu des standards de la recherche. Il faudrait ainsi que le ministère soit capable de dégager des moyens importants pour conduire ces évaluations, en réduisant le nombre de « petites » évaluations conduites dans différents domaines par exemple, et qu'il soit en mesure de définir, avec l'aide de chercheurs, un ensemble de règles, sorte de cahier des charges à respecter pour assurer la plus grande validité possible à l'étude commanditée.

Le Département américain d'éducation a par exemple édicté un certain nombre de conditions méthodologiques qui lui permettent de classer les évaluations en fonction de leur degré de fiabilité, et donc de la confiance qu'on peut apporter à leurs résultats: la randomisation a été ainsi élevée au rang de *gold standard* pour la production d'*evidence*, notamment dans la perspective de mise en place de politiques qui soient éclairées par ces résultats (*evidence-based policies*) (U.S. Department of Education, 2003).

Le deuxième élément à relever provient de la *Note Évaluation* de la DEPP (MEN, 2005). Sont en cause ici, d'une part, les traitements réalisés sur les données, qui ne corrigent pas les biais de sélection et, d'autre part, la rédaction de la note elle-même, qui met peu l'accent sur des effets, pourtant positifs, en fin d'expérimentation. On peut ainsi se demander si un organisme d'État lié au ministère, en l'occurrence une de ses directions, est le plus approprié pour conduire ces évaluations d'envergure et si celles-ci ne devraient pas être confiées à des organismes indépendants (cf. le point *supra*). Le concours du ministère, et en particulier de la DEPP, est sans doute fondamental dans la mesure où il est quasiment le seul à pouvoir initier des études d'envergure (tirer des échantillons d'élèves ou de classes représentatifs sur le plan national, obtenir les autorisations et assurer la participation des inspections académiques et de circonscription, etc.) mais cela pourrait se limiter à une aide dans la réalisation de l'évaluation, l'essentiel de la conception et des traitements revenant aux organismes indépendants. Il s'agit de mettre en place une organisation institutionnelle qui permette de se doter des conditions optimales de production de résultats probants. À cet égard, les organismes indépendants ont une liberté de ton que ne peut se permettre un organe du ministère. Dans la situation que nous préconisons, ce dernier peut jouer son rôle plein de transmission d'informations, de synthèses au ministre, mais les deux versions, scientifique et administrative existent alors et il est ainsi possible de se référer aux deux.

Le troisième élément que nous voulons pointer est la fiche ministérielle elle-même, qui fonde ses directives sur une sélection très partielle, voire partielle, des informations disponibles sur le sujet. On ne peut que s'étonner qu'un ministère utilise les faibles résultats d'une recherche comme argument pour prendre des mesures qui vont dans le sens opposé aux conclusions de rapports antérieurs (Meuret, 2001 ; Piketty, 2004 ; Piketty & Valdenaire, 2006). Il ne faut sans doute pas attendre de rapport idyllique entre politique et évaluation. Comme le disait Thélot (1993), il serait bien naïf de penser que le politique suive les évaluations car les décisions politiques relèvent fondamentalement du politique, non de l'expertise, chacun suivant des objectifs parfois différents. Il existe souvent une tension qui peut être vive entre le politique et le scientifique. En ce sens, les *evidence-based policies* peuvent se révéler parfois bien illusoire. Pourtant, si l'on peut sans doute admettre que, pour des raisons qu'il juge supérieures, le politique ne suive

pas les recommandations des expertises et autres évaluations, il apparaît non justifiable que les *evidences* soient contournées. À la lumière de ce travail, il est aussi possible de questionner l'usage plus que discutable qui est fait d'une absence de résultats (par ailleurs largement critiquable comme nous l'avons montré) comme preuve de l'inverse de ce qui avait été établi auparavant.

La production d'*evidences* demeure donc pour nous fondamentale. Au moins peut-on attendre des politiques qu'ils cherchent à avoir les informations les plus complètes possibles et qu'ils assument clairement, lorsque c'est le cas, de ne pas suivre les évaluations sans en opérer une lecture partielle ou une retraduction tronquée. Et qu'ils soient ainsi en mesure d'anticiper les conséquences de leurs décisions. En l'occurrence, il est difficile de soutenir que la hausse de la taille des classes n'entraînera pas de baisse des performances des élèves. Cette baisse sera sans doute faible si, dans chaque classe, un seul élève était ajouté. Mais il est bien évident qu'une hausse moyenne d'un élève par classe (ou fut-elle seulement de 0,6-0,7 en raison des contraintes situationnelles) se traduira, pour les mêmes raisons situationnelles, par des écarts de hausse d'effectifs entre les classes. Ainsi, cette hausse sera parfois insensible mais elle sera aussi, en certains endroits, plus importante et donc plus délétère en moyenne pour les acquis des élèves. Dans le cadre d'une économie de la connaissance, où la compétition économique future semble devoir se jouer en grande partie, pour les pays occidentaux, sur leur capacité à former des individus à de hauts niveaux de qualification, il n'est pas sûr qu'une politique de hausse des effectifs scolaires procède d'un bon calcul coût-efficacité. Et il n'est pas interdit de penser non plus que les gains économiques engrangés maintenant ne se traduiront pas par des pertes ultérieures, dans ce même domaine économique qu'on pensait préserver.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Angrist, J. D. & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2), 533-574.
- Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles : De Boeck.
- Bressoux, P., Kramarz, F. & Prost, C. (2009). Teachers' training, class size and students' outcomes: learning from administrative forecasting mistakes. *Economic Journal*, 119, 540-561.
- Finn, J. D. (2002). Small classes in American schools: research, practice, and politics. *Phi Delta Kappan*, 83(7), 551-560.
- Finn, J. D. & Achilles, C. M. (1999). Tennessee's class size study: findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), 97-109.

- Finn, J. D., Fulton, D., Zaharias, J. & Nye, B. A. (1999). Carry-over effects of small classes. *Peabody Journal of Education*, 67, 75-84.
- Finn, J. D., Gerber, S. B., Achilles, C. M. & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2), 145-183.
- Forestier, C. & Thélot, C. (2007). *Que vaut l'enseignement en France?* Paris: Stock.
- Goldstein, H. & Blatchford, P. (1998). Class size and educational achievement: a review of methodology with particular reference to a study design. *British Educational Research Journal*, 24(3), 255-268.
- Haut Conseil de l'Évaluation de l'École. (2001). *L'effet de la réduction de la taille des classes sur les progrès des élèves. Avis du Haut Conseil de l'Évaluation de l'École* (Avis n° 1). Consulté le 17 janvier 2011 dans <http://www.ladocumentationfrancaise.fr/rapports-publics/024000171/index.shtml>
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2), 497-532.
- Krueger, A. B. & Whitmore, D. M. (2001). *Would smaller classes help close the black-white achievement gap?* (Working paper #451). Princeton University.
- Meuret, D. (2001). *Les recherches sur la réduction de la taille des classes* (Rapport n° 1 établi à la demande du Haut Conseil de l'Évaluation de l'École). Consulté le 23 janvier 2011 dans <http://www.ladocumentationfrancaise.fr/rapports-publics/024000197/index.shtml>
- Ministère de l'Éducation nationale. (2005). Expérimentation d'une réduction des effectifs en cours préparatoires. *Note Évaluation 05.03*.
- Piketty, T. (2004). *L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises. Une estimation à partir du panel primaire 1997*. Paris School of Economics, mimeo.
- Piketty, T. & Valdenaire, M. (2006). L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français. *Les Dossiers Évaluations et Statistique*, 173.
- Slavin, R. (1990). Class Size and Student Achievement: Is smaller better? *Contemporary Education*, 62(1), 6-12.
- Thélot, C. (1993). *L'évaluation du système éducatif*. Paris: Nathan.
- U.S. Department of Education (2003). *Identifying and implementing educational practices supported by rigorous evidence: a user-friendly guide*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. NCEE 2004-3000.
- Wang, M. C. & Finn, J. D. (Ed.). (2000). *How small classes help teachers do their best*. Temple University Center for Research in Human Development and Education, Laboratory for Student Success, and U.S. Department of Education.

Les contradictions organisationnelles de l'évaluation. L'expérience des chefs d'établissement français

**Anne Barrère
Université Paris Descartes, CERLIS¹**

Si le fait d'évaluer n'est pas nouveau dans le monde de l'école, où les enseignants ont toujours jugé et noté leurs élèves, le renouveau d'intérêt qu'il suscite depuis une vingtaine d'années est indéniable, au point de rassembler précisément un ensemble de pratiques, en un seul mot, l'évaluation, qui désigne à la fois le fait de mesurer ou d'objectiver une activité, en général comparativement à d'autres, et de porter un jugement sur elle. À l'école, cette montée de l'évaluation est à replacer dans le contexte spécifique de politiques publiques cherchant à la fois à rendre le service public moins coûteux, plus moderne, plus efficace, et capable de « rendre des comptes » sur ses missions. Elle est désormais inséparable de nouveaux instruments de mesure, enquêtes nationales et internationales d'envergure, qui produisent des indicateurs chiffrés sur divers objets d'études, la plupart du temps centrés sur les résultats et les performances d'un système d'enseignement et de ses établissements et sur lesquels il existe maintenant une abondante littérature de recherche (Demailly, 2001 ; Lessard & Meirieu, 2005 ; Mons, 2007 ; *Revue française de pédagogie*, 2008 ; *Revue internationale d'éducation de Sèvres*, 2008 ; *Sociologie et sociétés*, 2008).

En fait, l'évaluation désigne aujourd'hui des réalités tellement différentes qu'en discuter suppose d'abord de définir de laquelle on parle. En particulier, il existe une différence certaine entre les procédures d'évaluation définies à un niveau macrosocial, et celles qui sont demandées, de plus en plus,

1. Centre de recherche sur les liens sociaux.

aux acteurs de l'école, en supposant qu'ils fassent le lien entre ces savoirs chiffrés et leurs pratiques. Ce qui distingue les deux types d'évaluation, ce ne sont pas forcément les objets de l'analyse (les résultats des élèves ou des établissements) mais le rapport à l'action, indirect et lointain, pour l'évaluation « d'en haut », même si les connaissances informent le politique, et par contre, bien plus immédiat, dans l'évaluation « d'en bas », où c'est en prise directe avec les acteurs concernés par l'évaluation qu'il faut essayer d'agir (Barrère, 2009, 2010).

Mais circonscrire un niveau particulier d'analyse, et même un groupe d'acteurs ne lève pas forcément toutes les difficultés à cerner les contours et les enjeux réels de l'évaluation. Nous voudrions le montrer à partir d'une enquête empirique sur les directions d'établissement de l'enseignement secondaire français². Rappelons qu'en France, les principaux de collège et proviseurs de lycée, dans leur majorité encore d'anciens enseignants, disposent de certaines marges de manœuvre depuis la décentralisation des années 1980, malgré des programmes, un recrutement des enseignants, et une attribution de moyens qui restent nationaux. Ces marges de manœuvre consistent à mettre en œuvre une certaine autonomie, nourrie de la prise en compte des spécificités locales, incarnée dans des projets et politiques d'établissement, dont les collectivités locales (conseil général et régional) et les acteurs associatifs sont aussi de nouveaux partenaires. Or, dans ce cadre évolutif (Pelage, 2003), et si des constats nourris et répétés déplorent que la culture de l'évaluation ne « descende pas » jusqu'aux enseignants, ils ne valent pas pour les chefs d'établissement qui dans leur ensemble y adhèrent bien davantage. L'idée que les indicateurs chiffrés dont disposent les établissements scolaires – principalement les évaluations nationales des acquis des élèves, mais aussi les indicateurs de résultats aux examens, baccalauréat et brevet, ainsi que diverses statistiques comme les flux d'orientation ou les taux de redoublement³ – doivent orienter les politiques d'établissement recueille un large consensus.

2. L'enquête qualitative qui permet d'explorer ces pistes d'analyse s'est déroulée entre janvier 2003 et décembre 2005 dans la région nord de la France. Elle se propose d'appréhender le plus finement possible le travail quotidien des chefs d'établissement et elle s'est appuyée sur deux types de matériaux: tout d'abord 43 entretiens semi-directifs auprès de chefs d'établissement: des principaux de collège (24 entretiens) dont 9 en collège d'éducation prioritaire; des proviseurs de lycée généraux ou polyvalents (13 entretiens) et professionnels (6 entretiens). En deuxième lieu, une enquête monographique d'un an a été menée dans un collège de réseau d'éducation prioritaire, entre 2004 et 2005, anonymisé ici sous le nom de « Salvador Allende ».

3. À partir de 1989, des évaluations systématiques des acquis des élèves en français et en mathématiques sont mises en place, à l'école primaire d'abord (CE 2, 6^e) puis en seconde à partir de 2002. Cette dernière évaluation en classe de seconde étant maintenant abandonnée. Depuis 1994, les lycées ont à leur disposition des indicateurs de pilotage.

Entrer dans la problématique au niveau de l'expérience des chefs d'établissement consiste donc à regarder l'évaluation par un autre biais que celui de la « résistance » au changement⁴. À partir de cette expérience, analysée sous l'angle du travail quotidien, on peut mettre en forme trois types de tensions structurelles, qui font de l'évaluation une réalité assez contradictoire et du coup difficile à cerner. Nous nommerons et expliciterons chaque tension avant de cerner comment les chefs d'établissement s'en accommodent au quotidien.

ENTRE BUREAUCRATIE ET POSTBUREAUCRATIE

L'évaluation est affectée d'une première tension organisationnelle. Elle est prescrite aux chefs d'établissement par leur hiérarchie, sur un mode traditionnellement bureaucratique, tout en étant structurellement porteuse d'intentions critiques à l'égard de ce même mode de fonctionnement.

Une tension structurelle en France

Le « pilotage par les résultats », de mise depuis la décentralisation, signale bien cette alliance prescrite de l'évaluation et de l'action locale, mais il s'exerce de plus dans le cadre d'un accroissement de la légitimité pédagogique des chefs d'établissement français, officiellement « premiers pédagogues de l'établissement » depuis les années 2000⁵ (Blanchet, Wiener & Isambert, 1999).

Depuis 2002, ils doivent établir à l'arrivée dans l'établissement un diagnostic et fixer des objectifs à quatre ans, dont ils discutent ensuite avec leur hiérarchie, avant d'établir la lettre de mission de leur adjoint. Les établissements situés en éducation prioritaire doivent également présenter des contrats d'action à quatre ans, également évalués, cohérents autour d'un projet de zone puis de réseau, et justifiant les moyens supplémentaires qui leur sont alloués. Enfin, la pratique d'audits se répand dans le système éducatif à partir de la fin des années 1980. Réalisés à la demande d'établissements qui sollicitent de l'aide, ou par les inspections générales à la faveur de crises ou de baisses de résultats spectaculaires, ou encore dans des dispositifs de type expérimental, comme dans l'académie de Lille entre

4. Même s'il faut rappeler que Crozier et Friedberg (1977) proposaient de rayer cette notion du vocabulaire sociologique, tant elle induisait de malentendus et, surtout, l'idée d'une passivité intrinsèque et parfois irrationnelle au lieu de stratégies construites; une invitation qui n'a, de toute évidence, guère été suivie d'effets.

5. Une politique quelque peu en contradiction avec l'ouverture théorique du corps et des cadres venus d'autres administrations ou entreprises. Les conseillers principaux d'éducation, de plus en plus nombreux dans leur fonction, éprouvent eux-mêmes cette tension dans la mesure où ils n'ont jamais été enseignants.

1990 et 1993 (Demailly, Deubel & Gadrey, 2002). Globalement, on voit bien que l'évaluation est donc bien une commande institutionnelle forte, faite par l'État à ses propres cadres et personnels.

Et pourtant, la régulation des organisations de travail par l'aval, en fonction des résultats, et non par l'amont en fonction de règles indépendantes des contextes d'action, a pour projet de remédier au manque d'efficacité et de souplesse de bureaucraties jugées incapables de se transformer, en particulier par l'adaptation à des contextes locaux différents et suscitant leurs propres besoins. Il s'agit bien d'infléchir un mode de contrôle, en remettant en question les chaînes traditionnelles du commandement bureaucratique, du sommet vers le bas, en passant par le niveau intermédiaire où se situent précisément les chefs d'établissement. Là où le strict respect des procédures – l'obligation de moyens – n'aboutit pas à des résultats satisfaisants, d'autres démarches, nées d'une plus grande adaptation au terrain, de la collaboration entre acteurs concernés par le même problème, sont légitimes, à condition de prouver leur efficacité. Comme on le voit, l'évaluation ne se suffit pas à elle-même, elle est inséparable de l'idée d'un autre mode de régulation des établissements, largement associée, dans bien des pays, à la décentralisation éducative (Maroy, 2006). Dans les faits, la nécessité de l'évaluation va de pair avec l'idée d'une mobilisation des enseignants et des acteurs de l'établissement qui serait différente du simple « minimum » bureaucratique. Les chefs d'établissement sont d'ailleurs unanimes à critiquer les enseignants qui ne font « que leurs heures de cours », signifiant par là leur adhésion à un élargissement du travail enseignant dans le cadre de l'établissement, en particulier dans le cadre de projets définis localement et portés par du travail en équipe. Mais ils sont également très majoritaires à critiquer le fonctionnement hiérarchique traditionnel des rectorats et des académies, en particulier leur manque de « réactivité » à des situations d'urgence ou leur peu de reconnaissance du travail réel accompli, au nom de grilles de promotion intangibles. Bref, d'une certaine manière, les chefs d'établissement, dans leur grande majorité, adhèrent à des modes de régulation moins bureaucratiques, dont l'évaluation est une des pierres angulaires, alors même qu'elle se présente comme une injonction venue d'en haut.

Solutions quotidiennes

Comment les chefs d'établissement gèrent-ils cette contradiction structurelle? Bien sûr, ni de manière uniforme, ni indépendamment de leurs contextes d'activité qui peuvent être très variés.

Pourtant, ils sont à peu près unanimes à dire subir un renforcement du pôle proprement administratif de leurs tâches, en contradiction avec leurs nouvelles fonctions. Dans ce renforcement, l'évaluation a une part même

si elle n'est pas seule en cause⁶. Les établissements reçoivent une série de chiffres et d'indicateurs, soit annuellement, soit à la faveur d'audits, et réagissent en acteurs bureaucratiques ordinaires. Lorsque je commence l'enquête au Collège Salvador Allende, le premier réflexe de l'équipe de direction est de me munir d'un important dossier comportant d'une part deux rapports d'audit, un, datant de 1994, précédant le classement du collège en « zone violence », un autre, fait par l'inspection générale, suite à une grève d'enseignants et à des problèmes de violence et de sécurité durant l'année 2000⁷. Mais ils me transmettent aussi les chiffres annuels concernant la réussite au brevet, les flux d'orientation. Ces documents sont très riches en chiffres divers sur le contexte de l'établissement, le mode de logement des élèves, leur temps de travail déclaré, mais en un an de présence sur le terrain, il n'y sera jamais fait allusion. Ce traitement bureaucratique des chiffres et indicateurs n'est jamais aussi prononcé que dans le cas des collèges et lycées dont les pourcentages de réussite aux examens confirment le bon fonctionnement. Car, dans ce cas précis, la contradiction apparaît nettement : si l'évaluation ne le demande pas, comment pousser à une régulation moins bureaucratique, comment convaincre de travailler en équipe, ou de s'investir dans des projets ? Certains proviseurs en arrivent alors presque à regretter ces bons chiffres : « Ils n'ont pas tellement envie de bouger non plus... Ici, il y a de bons résultats, donc c'est toujours pareil... » dira l'un d'eux.

Et d'ailleurs, effectivement, d'autres chefs d'établissement ne se résignent pas à cette absorption de la culture d'évaluation par ces procédures bureaucratiques. C'est presque toujours le cas dans les contextes d'éducation prioritaire ou face à des publics marqués par la difficulté scolaire. L'évaluation s'intègre alors dans la recherche d'une meilleure efficacité, qui se légitime en quelque sorte d'elle-même par la lutte contre l'échec. Mais elle va de pair avec deux types de démarches. D'une part, la construction d'indicateurs propres capables d'objectiver un certain nombre de problèmes spécifiques. Par exemple, l'équipe de direction du Collège Salvador Allende a chiffré le nombre de conseils de discipline, de veilles éducatives, (censés prévenir justement l'élève et anticiper les problèmes de discipline) et le nombre de sanctions et d'exclusions de cours qui sont répertoriés⁸ et diffusés lors de bilans de rentrée ou de juin au conseil d'administration. La participation des

6. Les chefs d'établissement font aussi allusion à diverses tâches nouvelles, y compris de nature administrative, confiées depuis la décentralisation.

7. Au-delà du cas de Salvador Allende, et invités à laisser à la fin des entretiens un document écrit de présentation, c'est pour deux tiers d'entre eux un document comprenant des bilans chiffrés, parfois évolutifs, des effectifs et des performances de l'établissement qu'ils m'ont confié, le dernier tiers étant fait de documents, en général plus esthétiques et destinés à la communication extérieure, présentant les lieux et l'offre de formation.

8. Les chiffres en sont impressionnants dans l'ensemble, allant à l'encontre, comme d'autres enquêtes plus larges, de l'idée d'un « laxisme généralisé » en matière de sanctions (Debarbieux, 1996), même s'il faudrait établir des points de comparaison pour se prononcer

parents d'élèves aux réunions a également fait l'objet d'une évaluation de la part de la principale adjointe, qui s'est également livrée au décompte, nettement plus problématique à certains égards, du nombre d'heures mensuelles où les élèves sont libérés de cours, soit pour cause d'absences d'enseignants, soit pour cause de projets ou de sorties, très nombreux à Salvador Allende.

D'autre part, pour faire de ces chiffres une réalité autre qu'administrative, ces établissements tentent d'effectuer ce que l'on pourrait appeler des tâches d'expertise locale, visant à y trouver des repères stables pour l'action d'un chef d'établissement dans un temps donné⁹. Cette situation, bien représentative de l'évaluation au local, contraste avec le sort réservé aux grands constats nationaux (enquêtes PISA, travaux du Haut comité à l'évaluation de l'enseignement) parfois percutants, mais qui s'accompagnent de réelles discussions – et incertitudes – quant à leur interprétation. Dans les établissements, les chefs d'établissement se doivent d'être plus affirmatifs que les experts¹⁰. On peut souligner quelques grandes tendances dans leur manière d'expliquer les chiffres pour guider leur action. Ceux qui ont vu les résultats de l'établissement s'améliorer depuis leur nomination ont tendance à mettre en avant un « effet chef d'établissement » sur les résultats, même s'il est peu relayé de manière aussi affirmative, du moins en France, par les experts¹¹. Ils mettent alors ces progrès sur le compte des actions qu'ils ont mises en place, par exemple l'aide individualisée en seconde¹², l'impulsion d'une meilleure communication entre les enseignants de différentes disciplines ou encore l'incitation au travail en équipe des enseignants, crédité d'efficacité par la quasi-totalité des directions. Un proviseur de lycée se décrit comme efficacement au chevet de son établissement : « J'ai fait tous les conseils de classe de seconde et j'ai préparé tous les conseils avec les profs principaux... comme le médecin, vous savez. Toutes les demi-heures, j'avais un professeur principal qui défilait. »

Pour ceux dont les indicateurs ne sont pas favorables, les rhétoriques sont évidemment différentes et les propos sont plus interrogatifs même s'ils privilégient souvent les facteurs internes de fonctionnement des établissements sur les changements de composition sociale du public. Un proviseur de lycée, qui pour contrecarrer l'élitisme de son établissement, l'ouvre à des familles plus populaires, qu'il « arrache » à l'établissement privé voisin,

plus avant : 812 heures de retenue en 2004, 956 en 2005, 501 exclusions de cours pour l'année scolaire 2004-2005, soit environ 15 exclusions par semaine.

9. N'oublions pas que les chefs d'établissement sont invités à bouger au bout de six ans d'exercice et normalement obligés de le faire au bout de neuf ans.

10. Rappelons par ailleurs que Thélot assumait explicitement que les indicateurs de performance soient un « miroir tendu d'en haut », sans explication propre, où les acteurs locaux se reconnaîtraient et trouveraient des repères pour agir (Pons, 2010).

11. Au sens strict du terme, c'est-à-dire d'une relation claire entre style de direction et performances (Grisay, 2006 ; Meuret, 2003).

12. Alors que l'efficacité générale de cette aide est loin d'être avérée (Duru-Bellat, 2002).

s'interroge longuement sur les facteurs pédagogiques qui ont déclenché, par la suite, une baisse au niveau des performances, alors qu'elle est sans doute imputable en grande partie à la modification du public :

La baisse des taux de réussite au bac, 10% sur l'ensemble des sections... évidemment ça nous a particulièrement interpellés; donc, on a conduit un travail d'analyse pour essayer d'examiner toutes les causes possibles, et ce travail d'analyse ne nous a conduits à rien. C'est-à-dire qu'objectivement, on ne peut pas dire que telle ou telle cause soit à l'origine de cette baisse.

L'équipe de direction du Collège Salvador Allende est fort préoccupée par l'écart de pourcentage d'un point en sa défaveur au brevet des collèges, le séparant de son rival Pablo Picasso, collège d'éducation prioritaire voisin, au management plus vertical et très centré sur la communication avec l'extérieur, alors que l'on peut penser que la petite partie de quartier favorisé présente sur le secteur du second suffit à lui assurer une légère différence de public, suffisante pour expliquer la différence en termes statistiques¹³.

Il ne faudrait pas donner une image fautive de ces perplexités suscitées par la culture de l'évaluation : elles restent saisonnières et circonscrites, lors des phases de résultats, peut-être accentuées par l'enquête sociologique elle-même. Mais malgré tout, la prise au sérieux de l'évaluation comme élément de régulation « postbureaucratique » suppose d'y investir beaucoup de temps, de l'énergie, et de développer certaines compétences, alors que le quotidien des chefs d'établissement est marqué du sceau de l'urgence et du manque de temps. Elle a alors bien des chances de rejoindre un « pôle idéal du travail » comme l'appelle Dujarier (2006), un ensemble de tâches impossibles à réaliser concrètement et qui pourtant sont jugées importantes. C'est aussi un argument pour la laisser se dissoudre dans un fonctionnement plus habituel, au risque de lui enlever la portée que les chefs d'établissement sont, dans leur majorité, prêts à lui donner.

ÉVALUATION ET POUVOIR

La deuxième contradiction se rapporte à la conception des rapports de pouvoir liés à la montée de l'évaluation. Le mot est affecté ici d'une grande labilité et incertitude. D'un côté, l'évaluation se présente comme une objectivation de l'action éducative et comme une promesse de transparence de

13. Pourtant, je n'ai eu de cesse de rappeler à l'équipe de Salvador Allende les résultats experts sur l'effet-public, et le manque de résultats stables sur l'effet-chef d'établissement. Un des points intéressants de l'enquête monographique est le rôle d'experte qui, au fil de l'année, a fini par m'être assigné, alors même que son sujet était loin des performances scolaires, montrant, s'il en était besoin, que l'évaluation produit une forte demande sociale, mais aussi confirmant que les acteurs, par la suite, ne « croient pas les sociologues », lorsque l'expérience fait sens d'une autre manière (Dubet, 2008, chap. 13).

ses résultats, dont tous les acteurs intéressés peuvent avoir connaissance. Elle participe alors à l'idée d'une construction de l'établissement comme démocratie locale (Derouet & Dutercq, 1997) et les textes officiels parlent d'ailleurs de diagnostics partagés et d'action concertée autour des constats réalisés.

Mais d'un autre côté, prévaut l'idée que les chefs d'établissement français n'auraient pas assez de pouvoir formel sur les équipes enseignantes, et que seul un renforcement de ce pouvoir hiérarchique permettrait d'accroître une transformation des pratiques, par exemple dans le sens d'un travail plus collégial, interdisciplinaire, ou encore, plus favorable à l'utilisation des évaluations nationales dans leur travail quotidien, notoirement insuffisante (Thélot, 2008). L'évaluation est alors prise en tenailles entre les deux discours. Il faut ici distinguer entre le pouvoir qui serait accordé aux usagers, principalement les familles, et les relations entre chefs d'établissement et enseignants.

Du côté des familles

En ce qui concerne les familles, force est de constater que l'évaluation participe à l'idée d'une montée de l'importance de l'information et des droits de l'utilisateur comme régulation du service public, même si elle n'en est pas le seul élément. Elles ont droit à des informations sur les performances et la qualité de l'école, d'où l'importance des palmarès chiffrés publiés dans les journaux, mais aussi des réputations plus qualitatives qui, même moins rationnelles, sont aussi des vecteurs puissants d'information (Ballion, 1986). Les choix des familles, plus ou moins informés, pèsent désormais sur les équilibres et l'évolution d'un établissement, surtout dans les contextes les plus difficiles où il peut y avoir des raisons de fuir vers le privé¹⁴ ou de contourner la carte scolaire, même s'ils ne peuvent être considérés de manière monolithique comme freinant la mixité sociale (Van Zanten, 2009; Visier & Zoïa, 2008). Ce nouveau pouvoir est partie prenante de l'évaluation comme mode de régulation postbureaucratique, incluant une plus grande proximité avec les usagers et il est, en ce sens, largement accepté et assumé par les chefs d'établissement. Il est même souvent l'objet d'un véritable travail de communication et d'image, lorsqu'il faut convaincre les parents de classe moyenne de laisser leurs enfants dans l'établissement malgré des incidents ou une réputation défavorable (Van Zanten, 2001). Le principal d'un bon collègue de banlieue résidentielle, dans sa lutte d'influence contre l'établissement privé voisin, plus attractif, raconte comment il doit rassurer,

14. 10% des élèves contournent en moyenne dans le public les établissements de leur secteur, mais le pourcentage peut être beaucoup plus fort dans certains quartiers. L'assouplissement officiel de la carte scolaire à la rentrée 2007, non pris en compte dans l'enquête, n'a pu que renforcer cette donnée.

la plupart du temps lors d'interactions informelles dans l'établissement... ou au supermarché principal du quartier, face à la moindre alerte : incidents entre élèves à la sortie des collèges, rumeurs de présence de drogue douce, ou même conflit entre son adjointe et lui :

Moi, j'habite ici, la gestionnaire habite ici... Et puis, on a de bons rapports avec les parents, les commerçants. Quand il y a quelque chose, ils ne se privent pas de nous le dire, ils disent : « Vous savez ce qu'on raconte, vous savez ce qu'on dit... »

Tel autre principal, toujours dans un contexte favorisé, institue une demi-journée porte ouverte par semaine à l'intention des parents, pour garder un meilleur contact avec eux. Mais les réputations sont aussi objets de réactions plus vives. Une proviseure de lycée professionnel, révoltée par son classement au bas d'un palmarès, voulait même regrouper tous les derniers des académies pour initier un mouvement de lutte national : « Je me suis faite écorcher, cartonner, placarder, stigmatiser par toutes les presses du quartier. »¹⁵ Le principal du Collège Salvador Allende, en tournée dans les écoles primaires pour convaincre de l'apaisement du climat de son établissement, se désespère de voir les parents d'élèves interrompre son discours et se lever brutalement pour vérifier qu'ils sont « dans les bonnes rues », celles qui correspondent au Collège voisin Pablo Neruda. Pourtant les rébellions des cadres face à l'évaluation, si elles sont loin d'être inexistantes, y compris dans le privé (Courpasson & Thoenig, 2008) sont pour l'instant individuelles et hésitantes. C'est d'autant plus compréhensible dans le cas des chefs d'établissement que ce nouveau pouvoir conféré aux usagers peut aussi être pour eux fortement légitimant de certaines actions à l'interne. Des contournements massifs d'un établissement ou une baisse des effectifs d'élèves peuvent être des leviers forts d'incitation au changement. De plus, il reste malgré tout une source de pouvoir ponctuel et limité en France. Si les parents d'élèves siègent au conseil d'administration, la situation où ils le dirigeraient, fréquente dans le monde anglo-saxon, est pour le moment bien éloignée du contexte français.

Du côté des enseignants

Tout autre est évidemment la situation face aux enseignants, dans un contexte où les chefs d'établissement, encadrant des équipes de professionnels toujours dotés d'une large autonomie théorique, doivent assurer des tâches pédagogiques transversales d'animation et d'impulsion, sans pour autant avoir accès véritablement au cœur de l'action dans la classe.

15. On rejoint ici la grogne de certains cadres contre l'évolution de certaines de leurs conditions de travail, où la culture des résultats joue un grand rôle (Courpasson & Thoenig, 2008; Dupuy, 2006).

Comment s'inscrit l'évaluation dans cette nouvelle donne ? Elle complexifie sans aucun doute les relations entre chefs d'établissement et enseignants, et ouvre la possibilité de nouveaux conflits de pouvoir, pour le moment seulement en émergence.

Pour certains, la nouvelle culture de l'évaluation constitue indéniablement un moyen de consolider un pouvoir formel trop défaillant, surtout en matière de contrôle des enseignants. L'évaluation réalisée au niveau des établissements et des équipes est alors, en fait, un moyen de faire avancer le délicat dossier de l'évaluation des enseignants¹⁶. Rappelons que l'évaluation reste partagée entre une évaluation administrative, que beaucoup de chefs d'établissement dénoncent précisément comme « affolante d'absurdité » tant elle est encadrée par des critères précisément purement bureaucratiques, et l'évaluation pédagogique des inspecteurs, beaucoup plus ponctuelle¹⁷. Mais la plupart des chefs d'établissement, observateurs quotidiens et parfois critiques de l'action enseignante, estiment que la situation actuelle ne leur permet ni de récompenser, ni de sanctionner suffisamment les enseignants. Par le biais des indicateurs chiffrés, même s'ils sont centrés sur les résultats des élèves, c'est alors l'efficacité enseignante « ordinaire » qui est visée. Bien sûr, pour le moment, ce n'est pas explicitement la classe qui est concernée, mais de manière plus courante les projets dans l'établissement, qui font pleinement partie des prérogatives des chefs d'établissement. Une principale mettra ainsi en cause l'opération phare de son équipe d'enseignants de français depuis quelques années, « Livres en fête », au vu du constat d'une « moins-value » des élèves de troisième en français, malgré de bons résultats en sixième : « Les professeurs qui avaient une plus-value de 6 % se sont reposés pendant quatre ans. » Un projet de structure-relais à l'intérieur d'un collège très difficile, qui aux dires de la principale, ne faisait descendre ni le nombre de conseils de discipline ni le nombre d'exclusions sera lui aussi supprimé. Ces deux épisodes seront fortement conflictuels, la deuxième principale finissant même, après d'autres problèmes de la même teneur, par quitter l'établissement. Mais d'autres pratiques, plus discrètes et d'une certaine manière tout à fait antérieures à la culture de l'évaluation, trouvent dans la nouvelle donne une légitimation supplémentaire : « ne pas faire intervenir certains enseignants dans certaines classes », une décision que ce principal assume « en off » dans l'entretien a, selon lui, largement permis d'améliorer les résultats de l'établissement. Les chefs d'établissement se centrent aussi sur un autre objet pédagogique : l'évaluation des élèves. La culture de l'évaluation conduit à comparer à l'interne les moyennes des classes et des enseignants, à analyser des écarts. Les indicateurs chiffrés globaux peuvent mettre en avant des écarts aux moyennes académiques ou nationales en matière de redoublement, d'orientation. Dans les collèges, les

16. Les enseignants ne sont pas favorables, rappelons-le, à une augmentation du pouvoir des chefs d'établissement en la matière (Do & Serra, 1996).

17. Les enseignants du secondaire sont inspectés une fois tous les sept ans en moyenne.

écarts entre le contrôle continu et le contrôle réalisé par l'examen servent à mettre en évidence des équipes « trop sévères », ou « trop laxistes », le premier défaut étant bien davantage dénoncé comme susceptible de démotiver les élèves: « Quand un professeur est un peu dur dans l'évaluation, je fais en sorte que dans une autre discipline, il y ait quelqu'un qui sera au contraire un peu plus généreux dans l'annotation pour compenser », dit ce principal de collège.

Ici, et comme dans bien d'autres pays (Mons, 2007), l'évaluation est alors synonyme d'un renforcement du contrôle sur une professionnalité enseignante en proie à des méfiances et à des doutes. La gouvernance postbureaucratique, fondée sur des mobilisations concertées, où les résultats sont des éléments de réflexion partagée, s'efface derrière l'idée que mesurer des résultats pourrait sanctionner des incompétences, même si pour l'instant les conséquences salariales ne sont pas à l'ordre du jour. Dans cette nouvelle perspective, les chefs d'établissement sont confrontés à deux tensions majeures.

D'une part, remonte à leur niveau, dans un effet de miroir, la tension entre « arbitre » et « entraîneur » décrite par Merle (1996) à propos des élèves. Comment à la fois impliquer et évaluer, si les évaluations sont défavorables? Que faire lorsque des équipes très mobilisées n'améliorent guère leurs résultats? On le sait, il n'y a pas forcément de cercle vertueux entre l'efficacité pédagogique et la mobilisation dans l'établissement (Felouzis & Perroton, 2007). Face aux constats négatifs établis lors d'une évaluation (et donc, ne l'oublions pas, sujette à révision ou discussion), une équipe de direction peut alors prendre des mesures qui sont susceptibles de décourager l'implication des enseignants, voire d'engendrer des conflits inédits, comme nous l'avons vu dans le cas de la décision d'arrêt de certains projets. D'autre part, les enseignants n'étant naturellement pas inactifs dans ces nouveaux rapports de force, s'ouvre un nouvel espace potentiel de conflits, où bien des initiatives des chefs d'établissement peuvent également être contestées, à l'heure même où la recherche met en évidence des différences dans l'efficacité de leur régulation (Dumay, 2009). Pourtant, il reste assez peu investi pour l'instant. Les conflits – parfois très lourds – entre enseignants et équipes de direction ne concernent que très minoritairement cette expertise locale et bien plus fréquemment le style d'autorité des directions ou telle ou telle décision prise dans le monde « bureaucratique »¹⁸ (Barrère, 2006). On peut y voir une conséquence du manque d'implication de la majorité des enseignants dans la culture de l'évaluation. On peut aussi le lire d'ores et déjà comme un succès des directions dans la mainmise sur ce qui apparaît de plus en plus comme un mode de légitimation de leur action locale. L'évaluation sert actuellement bien davantage aux chefs d'établissement à contester l'action

18. Pour ne citer que quelques exemples faisant conflit dans l'enquête: la date de fermeture d'un lycée pour l'examen, l'ordre du jour d'un conseil d'administration, les dépenses de l'établissement, le nombre des devoirs communs, etc.

enseignante qu'aux enseignants à contester les chefs d'établissement, qui pourtant sont eux-mêmes, de plus en plus considérés théoriquement comme « responsables » des synergies collectives et de leurs résultats¹⁹.

Du coup, analysée du côté des rapports de pouvoir, et sans cesser de porter des rhétoriques démocratisantes, en particulier en ce qui concerne l'intervention des usagers, c'est bien plutôt l'hypothèse générale d'une euphémisation de rapports de pouvoir toujours aussi réels et verticaux qui semble prévaloir (Courpasson, 2000). C'est aussi l'extension du domaine de contrôle possible des enseignants par les chefs d'établissement qui est à l'ordre du jour, plus ou moins nuancée par leur capacité et volonté de mettre en délibération les dispositifs pédagogiques adoptés au niveau de l'établissement.

L'ÉVALUATION : RATIONALITÉ EN MOYENS OU EN VALEUR ?

Enfin, une dernière tension porte sur la nature réelle du processus de rationalisation proposé par l'évaluation des établissements. Si elle se présente plutôt modestement comme un instrument, une procédure visant à orienter l'action, elle se leste de fait d'autres dimensions, dans un contexte où l'institution scolaire doit bon gré mal gré faire le deuil d'un projet éducatif global. Quel est le projet de l'enseignement secondaire aujourd'hui ? Il combine de forts accents modernisateurs, faisant le constat implicite d'une école inadaptée et en retard par rapport à la société, avec le maintien d'idéaux historiques de justice scolaire et sociale, malmenés depuis la massification par les constats sociologiques. Mais il peine à reconstruire une visée cohérente, que ce soit au travers de ses évolutions curriculaires, ou de la définition d'une visée globale, comme l'était la formation du citoyen pour l'école primaire et la formation d'une élite généraliste et cultivée dans l'enseignement secondaire (Isambert-Jamati, 1970), dans un système éducatif à deux vitesses. La rhétorique de « la réussite scolaire pour tous » en tient sans doute lieu depuis la massification, mais est-elle crédible dans une école où la compétition scolaire ne cesse de se durcir (Broccolichi, Ben Ayed & Trancart, 2010) ? Dans ce contexte, l'évaluation, dans sa capacité à mesurer précisément les différentiels de réussite scolaire connaît parfois une promotion, et finit par prendre quelques accents utopiques, que permettent aussi ses aspects flous ou contradictoires.

19. Pour donner un contre-exemple, en marge de l'enquête, un lycée de l'est de la France où j'ai été amenée à intervenir a connu malgré tout une discussion autour de ce type d'enjeux, les enseignants ayant fini par refuser, contre l'avis du proviseur, la constitution d'une classe de seconde « de faible niveau », au vu des résultats experts. Il faut préciser que le leader du Syndicat national des enseignants du secondaire (SNES) de l'établissement est fortement impliqué dans les débats pédagogiques.

Une solution à tout faire

C'est sans doute d'abord à un autre niveau que celui des établissements eux-mêmes que ce processus est visible, dans bien des discours officiels, projets de réformes, ou propositions de formation, visant en particulier les directions d'établissement. Bien sûr, et d'autant plus qu'on en maîtrise bien les outils, l'évaluation est compatible avec de multiples références possibles en termes de valeur. Mais du coup, et précisément parce qu'elle est plastique et compatible avec des visions différentes du système, l'évaluation devient une « solution à tout faire » (Alter, 2000), et par raccourci l'objectif lui-même. Le rapport de 2009 sur l'école primaire, après avoir pointé du doigt un certain nombre d'échecs et de résultats préoccupants, conclut qu'une des solutions à apporter est précisément... d'évaluer l'école primaire (Emin, 2009). Le consensus autour de l'instrument finit par dispenser de la formulation du problème, suivant une démarche bien repérée par la sociologie des organisations. Ou par rendre invisible ses aspérités : Bottani (2009) montre que le concept de résultats scolaires évolue sans aucune discussion entre 1992, où ils étaient définis comme les produits d'un système politique contextualisé, qui avait ses forces et ses faiblesses, et la diffusion de PISA où ils renvoient bien davantage aux pratiques et à la responsabilisation des acteurs par rapport à une efficacité individuelle. Dans la réforme récente du socle commun des compétences, la construction de nouveaux dispositifs d'évaluation, en partie à cause de leur complexité (les compétences étant diffractées en 72 capacités, dont certaines sont interdisciplinaires [Raulin, 2008]), fait passer au second plan les débats scientifiques fort nourris sur le sens du mot compétences (Crahay, 2006 ; Rey, 1998), ou l'ambivalence du projet lui-même (Del Rey, 2010). Pour paraphraser Mac Luhan²⁰, on peut alors dire que l'évaluation est le projet !

Un projet unidimensionnel ?

Dans les établissements scolaires, et même avant la mise en place du socle commun, la promotion possible de la rationalité portée par l'évaluation est aussi perceptible. Au local, les projets sont désormais débiteurs de contextes particuliers et de divers arrangements locaux (Derouet, 1992). Face à cette complexité, la culture de l'évaluation propose une réduction de taille : elle diffuse une conviction latente, exprimée telle quelle par le coordinateur du réseau auquel participe le Collège Salvador Allende, lors d'une réunion précisément consacrée à l'évaluation du contrat d'objectifs : « Chaque action doit produire des effets ». Dans un sens, finit par prévaloir l'idée que « la valeur ajoutée » de son action est toujours quantifiable. « On travaille, on travaille, on n'a pas les résultats », soupirera le principal de Salvador Allende

20. Qui, rappelons-le, avait affirmé que « Le média, c'est le message ».

devant ses mauvais résultats au brevet, et malgré de très bons résultats « climatiques ». Et à la fois, bien sûr, cette conviction se heurte à de multiples impasses au quotidien. Comment évaluer, comme y incite le contrat du réseau d'éducation prioritaire, l'évolution de la relation avec les familles, le succès d'une opération de prévention routière, qui a beaucoup plu aux élèves, l'action « La tête et les jambes », qui prévoit des défis lecture et des défis sportifs, ou même, le dynamisme du réseau? Si l'évaluation invite à rendre visible des résultats de l'action, cela veut-il dire que ce qui n'est pas chiffrable n'a pas de valeur? La réponse raisonnable est négative, mais il n'est pas si simple de renoncer à la légitimation par les chiffres.

Face à ces tensions, les directions d'établissement mises en difficulté par les indicateurs chiffrés font appel dans le discours à des résultats qualitatifs, reposant par exemple sur des considérations en termes d'ambiance de travail, de dynamisme, de respect des personnes. Comme disait un proviseur interrogé: « Moi finalement, mon indicateur, c'est le degré de bonheur dans l'établissement. » Mais il est difficile de franchir le pas suivant, c'est-à-dire de critiquer le principe de la quantification, et de renoncer à l'envie de construire ou de trouver d'autres indicateurs susceptibles de contrer une image peu valorisante en termes de performance. Comment faire réapparaître autrement, sinon, la partie du travail et de l'énergie dépensés qui a disparu avec l'évaluation (Dejours, 2003)? Le proviseur soucieux du bonheur de l'établissement est aussi celui qui fait passer des questionnaires aux familles sur leur degré de satisfaction; la principale adjointe de Salvador Allende affirme que les élèves rejoignent les classes en 5 minutes, au lieu de 15 au préalable, ce qui est une manière de chiffrer l'amélioration de l'ambiance, qui ne se voit pas pour l'instant dans la baisse des conseils de discipline ou des sanctions. L'évaluation devient alors un moyen trop complexe à manier au vu de certains objectifs, mais du coup fait peser une double menace: soit la réduction des finalités de l'éducation à la seule efficacité en termes de réussite scolaire et d'apprentissage, comme le mesurent les examens nationaux ou les enquêtes expertes, soit au contraire la perspective d'une spirale inflationniste qui viserait à évaluer autre chose, autrement... Mais quand et comment? L'évaluation au local est alors le lieu d'une autre forme d'utopie, plus techniciste, qui verrait proliférer des indicateurs capables de rendre justice au travail réel effectué.

CONCLUSION

Si l'évaluation est désormais un horizon normatif de bien des secteurs de la société, et en particulier de l'école, elle est aussi, comme nous l'avons décrit, un ensemble de tâches et de problèmes de plus en plus présents au quotidien.

Dans cette mutation qui est aussi sémantique, diverses pratiques et différents niveaux font désormais amalgame mais c'est aussi cet aspect polymorphe qui permet à l'évaluation d'être dans un sens à la fois incontournable et imperméable aux critiques : ce n'est jamais l'évaluation en elle-même qui est à remettre en cause, mais tel ou tel de ses usages, la « bonne » évaluation conseil n'est pas la « mauvaise » évaluation sanction ou contrôle. Ainsi, pour critiquer l'évaluation, il s'agirait aussi de l'évaluer, comme le titrent plusieurs journées d'étude ou colloques, dans une démarche structurellement inflationniste. C'est donc aussi sa capacité englobante, ses contradictions, ses déplacements qu'il faut appréhender si on veut avoir une chance d'en saisir toutes les dimensions (Martuccelli, 2010). La prolifération de contre-évaluations est peut-être le seul moyen de lutter contre des évaluations jugées injustes ou contraignantes (Vacarmes, 2008) et c'est ce dont rêvent effectivement bien des établissements scolaires mal traités par leurs indicateurs. Mais cette perspective ne peut qu'alimenter une sorte de prolifération, qui amène aussi à s'interroger sur la part de l'énergie et de l'activité sociale qui est d'ores et déjà et sera, à l'avenir, dépensée dans les pratiques d'évaluation, alors même que leur efficacité peut être questionnée en miroir (Harris & Herrington, 2006).

Dans le cas des directions d'établissement, l'évaluation s'inscrit dans trois types de choix professionnels, plus ou moins contraints par les contextes. Le premier consiste à prendre plus ou moins au sérieux l'évaluation comme instrument de régulation locale, alors que prévaut la tentation de l'associer à une gestion traditionnellement bureaucratique. Le deuxième s'inscrit dans le fait de partager plus ou moins avec les usagers et surtout les enseignants le nouveau pouvoir d'expertise procuré par la production ou l'interprétation de chiffres. Le troisième est le choix de la place, plus ou moins centrale, qu'occupe l'évaluation dans la légitimation que les chefs d'établissement font de leur action locale. Due en partie à leur inégal investissement dans le maniement des indicateurs quantitatifs – certains raisonnent en termes de « valeur ajoutée » alors que d'autres en critiquent son caractère réducteur – cette place est aussi débitrice de l'image, plus ou moins favorable, que renvoient les chiffres de leur action, à un moment où eux-mêmes sont de plus en plus censés « rendre des comptes ». Ainsi, l'évaluation apparaît-elle, aux yeux des chefs d'établissement français, un instrument aussi obligatoire dans son principe qu'ambivalent dans ses conséquences, ce qui les pousse, selon les contextes, à intensifier ou tempérer ses usages quotidiens.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Alter, N. (2000). *L'innovation ordinaire*. Paris : PUF.
- Ballion, R. (1986). *La bonne école*. Paris : Hatier.
- Barrère, A. (2006). Les chefs d'établissement face aux enseignants : enjeux et conflits de l'autonomie pédagogique. *Revue française de pédagogie*, 156, 89-100.
- Barrère, A. (2009). Les directions d'établissement scolaire à l'épreuve de l'évaluation locale. *Carrefours de l'éducation*, 28, 199-214.
- Barrère, A. (2010). Ce que l'évaluation fait aux établissements scolaires. Une année ordinaire au collège Salvador Allende. *Ethnologie française*, 1, 141-149.
- Blanchet, R., Wiener, C. & Isambert, J.-P. (1999). *La revalorisation des chefs d'établissements de l'enseignement secondaire*. Paris : MENRT.
- Bottani, N. (2009). Nouveaux savoirs sur l'enseignement, enjeux sur le plan international. In J.-C. Emin & J.-L. Villeneuve (Éd.), *Évaluer l'évaluation. Actes du colloque IREA-CFDT* (pp. 33-49). Paris : Le Manuscrit.
- Broccolichi, S., Ben Ayed, C. & Trancart, D. (2010). *École : les pièges de la concurrence. Comprendre le déclin de l'école française*. Paris : La Découverte.
- Courpasson, D. (2000). *L'action contrainte*. Paris : PUF.
- Courpasson, D. & Thoenig, J.-C. (2008). *Quand les cadres se rebellent*. Paris : Vuibert.
- Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.
- Crozier, M. & Friedberg, E. (1977). *L'acteur et le système*. Paris : Seuil.
- Debarbieux, E. (1996). *La violence en milieu scolaire* (Vol. 2). Paris : ESF.
- Dejours, C. (2003). *L'évaluation du travail à l'épreuve du réel. Critique des fondements de l'évaluation*. Paris : INRA Éditions.
- Del Rey, A. (2010). *À l'école des compétences : de l'éducation à la fabrique de l'élève performant*. Paris : La Découverte.
- Demailly, L. (Éd.). (2001). *Évaluer les politiques éducatives*. Bruxelles : De Boeck.
- Demailly, L., Deubel, P. & Gadrey, N. (2002). *Évaluer les établissements scolaires. Enjeux, expériences, débats*. Paris : L'Harmattan.
- Derouet, J.-L. (1992). *École et justice*. Paris : Métailié.
- Derouet, J.-L. & Dutercq, Y. (1997). *L'établissement scolaire, autonomie locale et service public*. Paris : ESF.
- Do, C.-L. & Serra, N. (1996). Les enseignants du second degré et l'évaluation individuelle de leur activité. *Éducation et formations*, 46, 61-70.
- Dubet, F. (2008). *Faits d'école*. Paris : EHESS.
- Dujarier, M.-A. (2006). *L'idéal au travail*. Paris : PUF.
- Dumay, X. (2009). Efficacité des modes locaux de coordination et de gestion des établissements. *Revue française de pédagogie*, 167, 101-129.
- Dupuy, F. (2006). *La fatigue des élites*. Paris : Seuil.

- Duru-Bellat, M. (2002). *Les inégalités sociales à l'école. Genèse et mythes*. Paris : PUF.
- Emin, J.-C. (2009). Problématique du colloque. In J.-C. Emin & J.-L. Villeneuve (Éd.), *Évaluer l'évaluation. Actes du colloque IREA-CFDT* (pp. 19-25). Paris : Le Manuscrit.
- Felouzis, G. & Perroton, J. (2007). Repenser les effets établissements : marchés scolaires et mobilisations. *Revue française de pédagogie*, 159, 103-119.
- Grisay, A. (2006). Que savons-nous de l'effet établissement ? In G. Chappelle & D. Meuret (Éd.), *Améliorer l'école* (pp. 215-230). Paris : PUF.
- Harris, D. & Herrington, C. (2006). La régulation par les résultats contribue-t-elle à l'amélioration des écoles ? In G. Chappelle & D. Meuret (Éd.), *Améliorer l'école* (pp. 203-215). Paris : PUF.
- Isambert-Jamati, V. (1970). *Crises de l'enseignement, crises de la société*. Paris : PUF.
- Lessard, C. & Meirieu, P. (Éd.). (2005). *L'obligation de résultats en éducation*. Bruxelles : De Boeck.
- Maroy, C. (2006). *École, régulation et marché*. Paris : PUF.
- Martuccelli, D. (2010). Critique de la philosophie de l'évaluation. *Cahiers internationaux de sociologie*, 128-129, 27-52.
- Merle, P. (1996). *L'évaluation des élèves*. Paris : PUF.
- Meuret, D. (2003). Efficacité et équité des collèges. In J.-L. Derouet (Éd.), *Le collège unique en question* (pp. 46-67). Paris : PUF.
- Mons, N. (2007). *Les nouvelles politiques éducatives*. Paris : PUF.
- Pelage, A. (2003). La redéfinition du métier de chef d'établissement secondaire : changement statutaire, construction de l'engagement professionnel et épreuves pratiques. *Revue française de pédagogie*, 145, 21-36.
- Pons, X. (2010). *Évaluer l'action éducative*. Paris : PUF.
- Raulin, D. (2008). *Le socle commun des connaissances et des compétences*. Paris : Hachette Éducation.
- Revue française de pédagogie*. (2008). Dossier : Évaluation des politiques éducatives et comparaisons internationales, 64.
- Revue internationale d'éducation de Sèvres*. (2008). Dossier : L'école et son contrôle, 48.
- Rey, B. (1998). *Les compétences transversales en question*. Paris : ESF.
- Sociologie et sociétés*. (2008). Les nouvelles politiques d'éducation et de formation, 40(1).
- Thélot, C. (2008). La principale faiblesse de l'évaluation et du contrôle de l'école en France. *Revue internationale d'éducation*, 48, 97-109.
- Vacarmes*. (2008). En finir avec l'évaluation, 44.
- Van Zanten, A. (2001). *L'école de la périphérie*. Paris : PUF.
- Van Zanten, A. (2009). *Le choix de l'école*. Paris : La Découverte.
- Visier, L. & Zoïa, G. (2008). *La carte scolaire et le territoire urbain*. Paris : PUF.

Développement de l'évaluation externe et restructuration du métier de direction d'établissement scolaire au Canada

Branka Cattonar* & Claude Lessard**
***Université de Louvain, CIRTES et GIRSEF¹**
****Université de Montréal, CRIFPE²**

ÉVOLUTION DU MODE DE RÉGULATION SCOLAIRE ET REDÉFINITION NORMATIVE DU RÔLE DES DIRECTIONS D'ÉTABLISSEMENT

Vers une régulation par les « résultats »

Ces dix dernières années, les systèmes éducatifs des différentes provinces au Canada ont connu plusieurs transformations importantes dans leur mode de régulation. Au-delà des politiques et des évolutions scolaires spécifiques à chaque province, certaines tendances générales sont communes aux différentes régions³. En particulier, Lessard et Brassard (2005) identifient une

1. Respectivement « Centre interdisciplinaire de recherche travail, État et société » et « Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation ».
2. Centre de recherche interuniversitaire sur la formation et la profession enseignante.
3. Le Canada est un État fédéral qui regroupe dix provinces et trois territoires. Comme le montrent Bernard et Saint-Arnaud (2004), les provinces canadiennes se caractérisent par des politiques sociales spécifiques et une articulation particulière entre l'État, le marché et la famille même si elles relèvent toutes d'un même régime providentiel de type libéral. L'éducation au primaire et au secondaire est une compétence provinciale et la gouvernance scolaire se base sur une structure à trois niveaux: les autorités provinciales (parlement provincial ou ministère de l'Éducation), les autorités intermédiaires (commissions scolaires ou

« stratégie pancanadienne », portée par une volonté d'harmoniser l'éducation au Canada, qui est axée sur l'obligation de résultats et l'imputabilité des acteurs éducatifs en mettant l'emphase sur la production de résultats quantifiés et d'indicateurs de performance scolaire. Plus précisément, ces auteurs soulignent deux évolutions majeures dans le mode de gouvernance éducative au Canada. La première est celle d'un renforcement de l'autorité centrale du système éducatif dans chacune des provinces, reflété dans diverses mesures comme un contrôle accru de la qualité, le développement d'évaluations standardisées, la production de standards de performance, mais aussi un curriculum et un financement davantage centralisés. La deuxième est celle d'une intensification de « l'axe horizontal de gouvernance » qui manifeste un relatif mouvement de décentralisation des pouvoirs vers les établissements et se traduit par la promotion d'une mobilisation des acteurs scolaires locaux, l'élaboration de « plans de performance » produits localement par chaque école, le développement d'une logique de compétition entre les établissements scolaires (Lessard & Brassard, 2005). Cette évolution met en particulier l'accent sur un pilotage du système éducatif par les résultats basé sur une logique de reddition de comptes et le développement de l'évaluation externe à différents paliers (système éducatif et établissements) que les différents acteurs scolaires sont supposés utiliser pour améliorer leurs performances à leur niveau⁴.

L'établissement scolaire comme communauté éducative mobilisée et responsable

Ces évolutions dans la régulation du système éducatif ont des effets à la fois au niveau des organisations et à celui de la professionnalité des personnels de l'éducation. Elles s'accompagnent en effet d'une redéfinition normative de la conception et du rôle de l'établissement scolaire. Selon Lessard et Brassard (2005), au Canada, la conception actuellement dominante de l'établissement est celle d'une « école communautaire » qui doit à la fois former une véritable « communauté éducative » en son sein et établir des liens avec la communauté qu'elle dessert. Dans cette logique, la mobilisation des acteurs locaux est fortement encouragée : les enseignants sont enjoins à travailler en équipe et les parents d'élèves se sont vus accorder une place plus importante dans le fonctionnement du système éducatif (notamment via

districts scolaires, qui sont des entités administrées par un conseil de commissaires élus par la population et qui exercent des responsabilités sur un territoire donné) et les établissements scolaires.

4. Ces tendances sont également relevées, à des degrés divers, dans de nombreux pays européens où l'on assiste à la montée d'un « État évaluateur » qui tend à s'attribuer le contrôle sur les orientations d'ensemble du système éducatif (définition des missions et des critères d'opérationnalisation des « résultats » du système), tout en déléguant (à des degrés variables) la gestion des moyens aux acteurs locaux (Maroy & Dupriez, 2000).

leur participation aux conseils d'établissement). Cette tendance va de pair avec une plus forte responsabilisation des établissements à l'égard de la réussite scolaire de leurs élèves dans le cadre d'une « autonomisation cadrée » du personnel enseignant et administratif. L'autonomie professionnelle du personnel éducatif est en effet valorisée par la rhétorique politique, mais elle reste encadrée, de manière diverse selon les provinces, par des curriculums centralisés et des évaluations standardisées et rendues publiques. Selon Ben Jaafar et Anderson (2004), l'établissement est ainsi devenu une « unité de reddition de comptes », tenue publiquement responsable de la performance des élèves. En même temps, on assiste à diverses formes de « déssectorialisation » de l'éducation qui augmentent les possibilités de choix de l'école par les parents, instaurent une forme de concurrence entre les écoles et les obligent à se positionner sur un « quasi-marché scolaire » (Lessard & Brassard, 2005). En définitive, l'établissement scolaire apparaît de plus en plus comme un lieu où les différents acteurs doivent se mobiliser et travailler ensemble pour produire de bons résultats, notamment à partir des indicateurs et analyses issues des évaluations externes, et sont tenus publiquement responsables de la performance des élèves.

Le directeur d'établissement comme acteur central du changement

Cette redéfinition de l'établissement scolaire, qui lui confère une importance plus grande dans la mise en œuvre des politiques scolaires et une responsabilité accrue dans la qualité et l'efficacité de l'enseignement, s'accompagne d'une redéfinition institutionnelle, normative et prescriptive, du rôle et du travail des personnels éducatifs. En ce qui concerne les directeurs d'école, de nouvelles prescriptions institutionnelles les enjoignent à devenir des « leaders pédagogiques » qui animent l'activité éducative au sein de leur école et assurent le pilotage des réformes scolaires en mobilisant leur équipe éducative (Bergeron, Massé & Rathé, 2005 ; Brassard *et al.*, 2004 ; Corriveau, 2004). Les directeurs d'école sont ainsi pensés comme des acteurs centraux du changement, à la fois responsables des résultats de leur école et de la réussite des autres réformes scolaires décrétées (Lapointe *et al.*, 2009). En Europe, des tendances comparables sont relevées et plusieurs études montrent que les directeurs ne doivent plus seulement être des administrateurs et des gestionnaires, mais aussi des « animateurs pédagogiques » initiateurs de la politique pédagogique de leur école et des « agents de changement » du système éducatif (Dutercq, 2006). En Communauté française de Belgique, Dupriez (2002, 2005) montre ainsi que, dans une logique de pilotage du système éducatif par les résultats, les récentes politiques scolaires valorisent un nouveau modèle idéal de « l'établissement mobilisé », où les différents acteurs sont censés construire collectivement des projets afin de produire

un enseignement de qualité et adapté aux spécificités de son environnement. Le directeur d'école voit en même temps son rôle redéfini : il est appelé à devenir un « animateur pédagogique », non seulement capable de gérer des personnes et des ressources, mais qui s'implique aussi dans la conduite du projet pédagogique de son école et garantit la concertation entre les enseignants. En France, Pelage (2003) met en évidence que le chef d'établissement est aujourd'hui tenu de devenir l'acteur dynamique de la transformation du système éducatif en combinant une pluralité de compétences et d'attributions : rigueur gestionnaire et mobilisation des ressources humaines, efficacité managériale et obligation des résultats, responsabilités et innovations pédagogiques. Le rôle du directeur y est redéfini autour de la figure du « chef d'établissement pédagogue », qui doit soutenir la modernisation du système éducatif en favorisant l'émergence de pratiques pédagogiques innovantes et en encourageant la réflexion collective et le travail en équipe des enseignants. En Angleterre, Osborn (2002) relève que les directeurs d'école sont désormais perçus avant tout comme des « agents de changement » qui doivent prendre la responsabilité de la mise en œuvre des réformes scolaires, alors qu'avant les années 1980, les *headteachers* étaient d'abord perçus comme des membres du personnel enseignant dont la principale fonction était de diriger les autres enseignants. Leur nouveau rôle les éloigne alors de l'enseignement, des élèves et des enseignants. De manière plus générale, Leclercq (2005) constate que dans de nombreux pays d'Europe et d'Amérique du Nord, l'autonomie et la responsabilisation accrues des établissements scolaires se sont accompagnées d'une redéfinition de la fonction de directeur d'école, lequel n'est plus considéré comme un simple administrateur chargé d'observer les règlements, d'exécuter un budget et de gérer le personnel, mais simultanément comme un manager, un animateur pédagogique et un initiateur, responsable des résultats qu'il obtient avec un collectif qu'il doit connaître et animer.

Comme le souligne Barrère (2009), pour la France, parmi les nouvelles missions attribuées aux directeurs d'école, celle qui relève de l'évaluation occupe de plus en plus une place centrale. C'est en effet à partir des constats issus des divers dispositifs d'évaluation externe que les directeurs d'école sont supposés établir des « diagnostics partagés » avec leur équipe éducative pour élaborer ensuite leur travail d'animation pédagogique et construire localement une politique éducative efficace. Au Canada également, les directeurs sont tenus d'élaborer des « plans de performance, de réussite ou d'amélioration » notamment à partir des nombreuses évaluations externes à leur disposition (les évaluations standardisées des élèves dans le cadre des épreuves provinciales annuelles, le *Bulletin* annuel des écoles secondaires de l'Institut Fraser, celui du Conseil canadien de l'apprentissage, les épreuves et analyses des commissions scolaires incorporées à leurs planifications stratégiques, les programmes PISA, TIMSS et PIRLS). Les directeurs sont de la sorte

encouragés par les politiques à devenir un nouvel « échelon de régulation pertinent », intermédiaire entre le niveau national et local (Barrère, 2008).

Dans cet article, nous allons nous pencher sur les effets qu'ont ces évolutions du mode de régulation du système éducatif sur les conditions et la teneur du travail des directions d'établissement. Comment les directeurs ressentent-ils les récentes évolutions scolaires? Dans un tel contexte, comment vivent-ils, conçoivent-ils et exercent-ils leur métier? Nous tâcherons de répondre à ces questions en nous appuyant sur les résultats d'une enquête par questionnaire réalisée auprès de directeurs d'école primaire et secondaire dans l'ensemble du Canada⁵. Notre hypothèse est qu'à travers le tourbillon des changements (réels et imaginaires), qui prennent sens dans le contexte de la mise en place d'une nouvelle régulation de l'éducation (Maroy & Dupriez, 2000), s'installe peu à peu une nouvelle conception du rôle de direction, perçue plutôt favorablement par les principaux intéressés. De nouvelles dimensions de la fonction prennent de l'importance, d'autres, plus anciennes, deviennent secondaires, le travail se diversifie et s'enrichit en même temps qu'il s'alourdit. Cette transformation de la fonction n'est ni achevée, ni complètement stabilisée; le temps nous dira si elle s'institutionnalise durablement, mais la plupart des directions interrogées, en dépit de la surcharge qu'elle occasionne, la voient plutôt d'un bon œil.

Nous commencerons par présenter brièvement l'enquête, l'échantillon et les principes de l'analyse statistique mise en œuvre. Nous exposerons ensuite les principaux résultats de l'enquête qui éclairent, d'une part, la manière dont les directeurs d'école vivent les récentes évolutions du système éducatif et, d'autre part, la manière dont ils conçoivent et disent exercer leur travail. Nous verrons que la plupart se déclarent confrontés à de nombreuses évolutions et ont une vision relativement nuancée de leurs impacts. Nous verrons aussi que, de leurs points de vue, leur métier implique une multitude de responsabilités et de rôles et qu'il leur apporte une satisfaction professionnelle très diversifiée, même si une certaine surcharge se révèle et les confronte à des processus d'intensification du travail. Nous terminerons en émettant quelques hypothèses sur la manière dont l'évolution du mode de régulation scolaire engage une restructuration du métier de directeur au Canada. Nous supposerons en particulier qu'elle s'accompagne d'un accroissement de leur implication dans les relations externes à l'école et dans la supervision des enseignants.

5. Cette enquête a été réalisée dans le cadre d'une recherche subventionnée par le Conseil de recherches en sciences humaines du Canada (programme des Grands travaux de recherche concertée), sous la direction d'abord de M. Tardif, puis de C. Lessard (tous deux de l'Université de Montréal).

L'ENQUÊTE

L'enquête par questionnaire a été administrée, en 2005, à 4800 directeurs selon un plan d'échantillonnage stratifié selon les provinces et l'ordre d'enseignement⁶. 2144 directeurs ont répondu, pour un taux de réponse de 44,6%. Parmi ceux-ci, 63,1% dirigent une école primaire, 23,7% une école secondaire et 13,2% une école « mixte » (offrant à la fois un enseignement primaire et secondaire). La majorité travaille dans une école publique (90%) et en région urbaine (70%). Enfin, l'échantillon est composé de 25,5% de directeurs provenant de l'Ontario, 22,5% des Prairies⁷, 19,6% du Québec, 15,4% de l'Atlantique⁸, 14,4% de la Colombie-Britannique et 2,6% des Territoires⁹.

Le questionnaire qui leur a été adressé comporte une cinquantaine de questions poursuivant plusieurs objectifs : dresser un portrait du profil socioprofessionnel des directeurs, éclairer leurs parcours d'insertion professionnelle, saisir leurs perceptions des changements scolaires, analyser leur rapport au métier (tâches accomplies, conceptions idéales du travail et satisfaction professionnelle), décrire leurs conditions de travail (les caractéristiques sociales et scolaires des élèves de l'école qu'ils dirigent, les rapports sociaux au sein de leur établissement).

L'analyse des données ainsi recueillies vise à étudier le métier de directeur en partant des représentations subjectives que les chefs d'établissement ont de leur travail tout en tenant compte des conditions objectives dans lesquelles se déroulent leurs expériences professionnelles. Elle s'inscrit dans une perspective sociologique « constructiviste » (Corcuff, 1995) de l'identité professionnelle qui la considère comme une construction dynamique à la fois sociale et individuelle, résultat d'un processus de socialisation biographique et relationnel lié au contexte (socio-historique, éducatif et professionnel) particulier dans lequel il s'inscrit (Cattonar, 2001). Concrètement, l'analyse statistique s'est attachée à réaliser une « description compréhensive » de la réalité actuelle du métier des directeurs. Nous étudions la manière dont leurs perceptions des changements scolaires, leurs conceptions du métier, leurs pratiques et leur vécu varient, d'une part, selon leur profil social et professionnel (genre, âge, diplôme, etc.), et d'autre part, selon le contexte dans lequel ils travaillent (niveau d'enseignement, localisation urbaine ou

6. L'échantillon des répondants est partiellement représentatif au seuil de signification de 5% au niveau de la région (il sous-représente l'Ontario, sur-représente l'Atlantique et les Territoires) et au niveau de l'ordre d'enseignement (il sur-représente les écoles secondaires), sans que cela ait toutefois justifié de redressement de l'échantillon.

7. La région des Prairies regroupe les provinces de l'Alberta, de la Saskatchewan et du Manitoba.

8. La région de l'Atlantique regroupe la Nouvelle-Écosse, le Nouveau-Brunswick, Terre-Neuve-et-Labrador et l'Île-du-Prince-Édouard.

9. La région des Territoires regroupe les Territoires du Nord-Ouest, le Yukon et le Nunavut.

rurale de l'école, province, caractéristiques des élèves et du personnel, etc.) (voir Cattonar *et al.*, 2007). Dans cet article cependant, nous mettrons principalement l'accent, d'une part, sur les perceptions du métier et les expériences professionnelles qui sont communes à l'ensemble des enquêtés; d'autre part, sur les items relatifs au nouveau mode de régulation scolaire.

LE PILOTAGE PAR LES RÉSULTATS: DES EFFETS NÉGATIFS, MAIS UNE ADHÉSION AU PRINCIPE

Comme nous pouvons le voir au tableau 1, la majorité des directeurs enquêtés estime que de nombreuses évolutions, de nature diverse, ont eu des effets importants sur leur école. En particulier, la majorité des directeurs juge *très important* ou *important* l'impact de plusieurs changements dans le mode de régulation scolaire qui renvoient au pilotage par les résultats et au développement de l'évaluation: les nouvelles politiques d'imputabilité (79%, en cumulant les deux premières positions de l'échelle), l'évaluation standardisée des élèves (71,7%) et l'évaluation formelle des enseignants (61,6%). Par contre, seuls 39,7% des directeurs estiment important l'accroissement de la compétition entre les écoles. On remarque également que ce sont surtout les changements d'ordre pédagogique (l'introduction de nouvelles approches éducatives et des technologies de l'information et de la communication) et les restrictions budgétaires (la diminution des ressources humaines et autres) qui sont mises en avant par la plupart des directeurs (plus de 85%), alors que les changements socioculturels du public scolaire ou de l'environnement paraissent moins prégnants¹⁰.

Aux yeux des directeurs, ces évolutions ont de nombreuses conséquences sur leur fonction et leur école¹¹. Parmi les effets que l'on peut qualifier de « neutres », une majorité de directeurs cite le développement de nouvelles capacités d'adaptation (92,1%), une plus forte demande de formation et d'accompagnement institutionnel (81,6%), une conscience plus grande des relations avec l'environnement de l'école (80,1%) et une modification de leur approche de la gestion (79,6%). Parmi les principaux effets négatifs mis en avant, la majorité évoque une augmentation de leur charge de travail (96%), le développement d'un climat de méfiance dans les écoles (66,7%), une perte des points de repère habituels (62,6%), une

10. Les perceptions de l'importance des différents changements sont toutes corrélées ensemble positivement. Les corrélations les plus fortes s'observent notamment entre les nouvelles politiques d'imputabilité, l'évaluation standardisée des élèves, l'évaluation formelle des enseignants et la nouvelle répartition des pouvoirs entre les instances centrales et locales (les valeurs des Tau de Kendall sont supérieures à 0,3 et sont statistiquement significatives au seuil de 1%).

11. Quatre questions ont été posées aux directeurs pour saisir leurs perceptions d'une trentaine d'effets possibles.

Tableau 1 : La perception qu'ont les directeurs de l'impact des évolutions scolaires sur leur école

	Très important	Important	Peu important	Pas important du tout	Total	N
Changements d'ordre pédagogique						
Nouvelles approches éducatives (curriculum)	44,2 %	47,4 %	7,0 %	1,4 %	100 %	2082
Introduction des technologies de l'information et de la communication	38,1 %	50,5 %	8,5 %	2,9 %	100 %	2087
Restrictions budgétaires						
Diminution des ressources humaines	58,2 %	27,3 %	6,9 %	7,7 %	100 %	2080
Diminution des autres ressources	51,8 %	35,3 %	7,4 %	5,6 %	100 %	2079
Évolution du mode de régulation scolaire						
Nouvelles politiques d'imputabilité	33,5 %	45,5 %	15,5 %	5,5 %	100 %	2081
Évaluation standardisée des élèves	23,3 %	48,4 %	19,7 %	8,6 %	100 %	2081
Évaluation formelle des enseignants	19,9 %	41,7 %	26,8 %	11,6 %	100 %	2079
Nouvelle répartition des responsabilités et pouvoirs entre instances centrales, locales et à l'intérieur des établissements	34,2 %	38,0 %	18,2 %	9,6 %	100 %	2080
Réorganisation et fusion des conseils scolaires	28,1 %	25,4 %	20,3 %	26,1 %	100 %	2079
Accroissement de la compétition entre les écoles	16,1 %	23,6 %	34,1 %	26,1 %	100 %	2077
Changements d'ordre « démographique »						
Mouvements au sein du personnel scolaire (retraite, redéploiement ou renouvellement)	24,1 %	44,5 %	24,9 %	6,4 %	100 %	2072
Variation du nombre d'élèves	31,8 %	35,8 %	24,6 %	7,9 %	100 %	2084
Changements socioculturels						
Changements socio-économiques dans l'environnement	20,2 %	39,4 %	28,9 %	11,5 %	100 %	2077
Diversification culturelle des élèves	12,9 %	30,0 %	37,3 %	19,8 %	100 %	2082

baisse de la qualité des services aux élèves (60,3 %) et le développement d'un sentiment d'inefficacité (58,5 %). En particulier, on peut relever que plus les directeurs pensent que les nouvelles politiques d'imputabilité et le développement des évaluations des élèves et des enseignants sont importants, plus ils disent avoir dû modifier leur approche de la gestion et suivre une formation additionnelle, mais aussi avoir le sentiment que leur maîtrise de la situation a diminué, que leur charge de travail a augmenté, qu'il y a au sein des écoles une perte des points de repère habituels, le développement d'un climat de méfiance, le développement d'un sentiment d'inefficacité, un renforcement des rivalités, une détérioration du milieu de vie, une diminution de la qualité des services rendus aux élèves, ainsi qu'une plus forte demande de formation et une implication plus grande des parents¹². Le pilotage par les résultats semble ainsi avoir des effets plutôt négatifs sur le vécu professionnel des directeurs et le fonctionnement des écoles. Deux autres évolutions sont également corrélées à des conséquences négatives : il s'agit des diminutions des ressources et de la diversification culturelle des élèves¹³.

En même temps, la plupart des directeurs paraissent cependant rester optimistes face aux différentes évolutions scolaires dans la mesure où ils pensent qu'elles *auront*, dans le futur, de nombreux effets positifs : sur l'apprentissage des élèves (81,7 %), leur fonction de directeur (74,4 %), l'intégration sociale des élèves (68,5 %), les relations avec les parents (67,7 %), la professionnalisation des enseignants (67,4 %), la reconnaissance de la mission de l'école (67,2 %) et l'efficacité du système scolaire (66,9 %). C'est en particulier le cas des politiques d'imputabilité et des évaluations standardisées qui sont corrélées à plusieurs effets *anticipés* positifs : sur l'apprentissage des élèves, la professionnalisation des enseignants, l'efficacité du système scolaire et les relations avec les parents¹⁴. Cet optimisme quant à des bénéfices escomptés pourrait refléter une certaine adhésion de la part des directeurs aux évolutions en cours. Il pourrait aussi se comprendre par des motivations professionnelles particulières, distinctes notamment de celles des enseignants (Lessard, Kamanzi & Larochelle, 2008). Au Québec, Corriveau *et al.* (2001), à partir d'une enquête menée auprès de directeurs d'établissement du primaire et du secondaire portant sur le changement en éducation, affirment qu'« en moyenne les répondants sont favorables, voire même très favorables aux changements » (p. 7), et qu'ils se définissent comme des « facilitateurs » de la mise en œuvre dans leur établissement de

12. Les corrélations entre les différents items ont été mesurées par des Tau de Kendall dont les valeurs varient entre 0,10 et 0,25 et sont significatives au seuil de 1 %.

13. Les valeurs des Tau de Kendall sont comprises entre 0,16 et 0,32 et sont significatives au seuil de 1 %.

14. Les valeurs des Tau de Kendall sont comprises entre 0,08 et 0,15 et sont significatives au seuil de 1 %. La nouvelle répartition des pouvoirs entre instances centrales et locales est une autre évolution scolaire fortement associée à des effets anticipés positifs.

la réforme scolaire. Dans le même ordre d'idées, Deniger, Kamanzi, Chabot et Hébert (2006), dans une étude quantitative de la perception du nouveau programme de formation de l'école québécoise, qualifient les directions d'école interrogées de « partisans enthousiastes » du nouveau programme et estiment que les enseignants sont plutôt des « artisans sceptiques » faisant « contrepoids à l'enthousiasme des directions d'école » (p. 57). Plusieurs études ont aussi montré que les directeurs apprécient les défis posés par les changements et le fait de se sentir « au cœur de l'action » malgré l'exigence et la complexité de la tâche (Barrère, 2006; Corriveau, 2004). Nous verrons plus loin qu'une majorité des directeurs valorise d'ailleurs leur rôle « d'agent de changement ». On pourrait supposer que cette attitude face aux évolutions scolaires est relativement récente et qu'elle traduit une certaine adhésion à l'injonction qui leur est faite aujourd'hui de devenir des acteurs responsables de la transformation du système éducatif. Une étude réalisée au début des années 1990, dans l'ensemble du Canada, montrait en effet que les directeurs avaient au contraire une opinion négative à l'égard des politiques gouvernementales (King & Peart, 1992).

Il faut souligner que les évolutions scolaires et leurs effets sont diversement ressentis selon les contextes où travaillent les directeurs d'école. En premier lieu, leurs perceptions varient fortement selon les provinces, lesquelles ont effectivement connu des évolutions spécifiques (Lessard & Brassard, 2005). Par ailleurs, l'enquête montre que leurs perceptions des évolutions scolaires diffèrent également, moins fortement cependant, selon d'autres variables contextuelles, comme le niveau d'enseignement, le profil du public scolaire et la localisation urbaine ou rurale de l'école qu'ils dirigent (voir Cattonar *et al.*, 2007). En ce qui concerne le pilotage par les résultats, on observe ainsi que les nouvelles politiques d'imputabilité sont plus souvent jugées importantes par les directeurs d'école primaire¹⁵.

LE RAPPORT AU MÉTIER

Une multitude de rôles et de responsabilités à assumer, une forte valorisation du travail pédagogique

À l'instar de plusieurs études réalisées dans divers pays européens (Barrère, 2006; Boissinot, 2005; Dupriez, 2002; Osborn, 2002), les données de notre enquête indiquent qu'aux yeux des directeurs, leur travail implique l'exercice d'une multitude de rôles et de responsabilités.

Tout d'abord, en ce qui concerne les rôles qu'ils déclarent *effectivement* remplir dans le cadre de leur travail (voir la colonne « rôles remplis » du

15. La valeur du V de Cramer est de 0,12 et est significative au seuil de 1 %.

tableau 2), plus de 96 % des enquêtés jugent *très importants* ou *importants* les rôles traditionnels qui renvoient à la gestion et à l'administration : gestionnaire des urgences, chef d'orchestre et administrateur général de l'école. Arrivent ensuite des rôles davantage liés à l'administration pédagogique (agent de changement des politiques et des pratiques de l'école, planificateur du projet éducatif de l'école, superviseur et évaluateur du travail des enseignants), ainsi qu'à la gestion des relations externes à l'établissement (interlocuteur des parents et médiateur, agent de liaison auprès des autorités, promoteur de l'école dans la communauté). Les rôles les moins souvent jugés importants dans l'exercice de leur travail sont ceux qui concernent l'animation pédagogique (leader pédagogique) et l'éducation des élèves.

Tableau 2: Les rôles remplis et idéaux

	Rôles remplis				Rôles idéaux			
	Très important ou important*	Peu ou pas important du tout*	Total	N	Très important ou important*	Peu ou pas important du tout*	Total	N
Gestion et administration								
Gestionnaire des urgences	97,9 %	2,1 %	100,0 %	2067	92,7 %	7,3 %	100,0 %	1978
Chef d'orchestre, coordonnateur, rassembleur, chef d'équipe	97,9 %	2,1 %	100,0 %	2062	97,0 %	3,0 %	100,0 %	1976
Administrateur général de l'école (budget, équipements)	96,1 %	3,9 %	100,0 %	2064	90,9 %	9,1 %	100,0 %	1981
Travail pédagogique								
Agent de changement des politiques et pratiques de l'école	94,3 %	5,7 %	100,0 %	2058	97,1 %	2,9 %	100,0 %	1985
Planificateur du projet éducatif de l'école	93,8 %	6,2 %	100,0 %	2055	95,3 %	4,7 %	100,0 %	1976
Superviseur et évaluateur du travail des enseignants	93,7 %	6,3 %	100,0 %	2060	96,7 %	3,3 %	100,0 %	1978
Leader pédagogique	89,1 %	10,9 %	100,0 %	2060	97,6 %	2,4 %	100,0 %	1988
Éducateur des élèves	85,7 %	14,3 %	100,0 %	2059	91,3 %	8,7 %	100,0 %	1974
Gestion des relations externes								
Interlocuteur des parents et médiateur	92,5 %	7,5 %	100,0 %	2060	88,9 %	11,1 %	100,0 %	1976
Agent de liaison auprès des autorités	91,5 %	8,5 %	100,0 %	2060	90,7 %	9,3 %	100,0 %	1975
Promoteur de l'école dans la communauté	89,6 %	10,4 %	100,0 %	2064	94,6 %	5,4 %	100,0 %	1987

*Afin d'alléger la présentation du tableau, nous avons regroupé, d'une part, les réponses « très important » et « important », et, d'autre part, « peu important » et « pas important du tout ».

Si l'on considère les rôles qu'ils *souhaiteraient* exercer (voir la colonne « rôles idéaux » du tableau 2), on observe que la majorité des directeurs (plus de 88 %) estime que leur travail implique aussi, *idéalement*, une grande variété de rôles. En particulier, on observe que plus de 95 % des répondants valorisent les rôles qui renvoient au travail pédagogique (leader pédagogique, agent de changement des politiques et pratiques de l'école, superviseur du travail des enseignants, planificateur du projet éducatif de l'école) et à la coordination (chef d'orchestre). Les rôles les moins souvent valorisés sont ceux d'interlocuteur des parents, d'agent de liaison auprès des autorités et d'éducateur des élèves.

La redéfinition institutionnelle du rôle des directeurs qui est promue actuellement par les politiques scolaires et qui les enjoint à devenir des administrateurs pédagogiques et des agents de transformation éducative semble ainsi posée en tant qu'idéal par les directeurs eux-mêmes. Comme le souligne Barrère (2006), appelés à être des « prescripteurs d'initiative et de dynamisme » pédagogique (notamment en mobilisant leur personnel), ils semblent vivre cette injonction davantage comme un « principe fort de leur réalisation personnelle » que comme une contrainte (p. 159). Il faut souligner qu'au Canada, les programmes obligatoires et universitaires de formation des directions d'établissement diffusent cette vision idéale du métier. Nombre d'écrits relatifs à l'administration scolaire (par exemple, Hargreaves, Lieberman, Fullan & Hopkins, 1998; Murphy & Seashore-Louis, 1999; Philips, Raham & Reniham, 2003) valorisent l'exercice du leadership par les directions, tant du point de vue pédagogique que communautaire, ou encore dans la mise en valeur de l'établissement dans un environnement de plus en plus concurrentiel. Notre étude rejoint celle réalisée par Lapointe *et al.* (2009) au Québec, dont les résultats « tendent à corroborer le modèle théorique prédominant en administration de l'éducation selon lequel le directeur d'école, en tant qu'agent de changement, joue un rôle déterminant lors de l'implantation d'une réforme » (p. xi). Il fait donc sens de penser que les chefs d'établissement formés à l'université en administration de l'éducation sont amenés à adhérer à cette idéologie du manager agent de changement et leader pédagogique et que cette idéologie professionnelle est leur réponse et manière de tirer leur épingle du jeu d'une nouvelle régulation de l'éducation par les résultats (Lessard, Kamanzi & Larochelle, 2008). Le fait que les représentations *idéales* des responsabilités ou des rôles varient faiblement selon le contexte de travail donne à penser que nous sommes ici en présence de représentations construites et diffusées dans la formation initiale et continue des directions.

Si l'on examine plus en détail les différentes responsabilités qu'ils disent assumer (voir la colonne « responsabilités assumées » du tableau 3), leur travail apparaît de nouveau comme fortement varié. En particulier, on observe que la plupart (plus de 70 %) déclarent *être entièrement responsables* ou *jouer un rôle majeur* dans des tâches qui concernent l'administration péda-

gogique (élaboration du règlement et de la mission de l'école, affectation des tâches d'enseignement, répartition des élèves, évaluation et définition des programmes et méthodes pédagogiques) ainsi que l'évaluation (analyse des statistiques scolaires) et la reddition de comptes. Ces tâches, qui sont aujourd'hui fortement encouragées par les politiques scolaires, paraissent donc être considérées par une grande partie des directeurs comme relevant de fait de leur responsabilité. Elles semblent en outre être valorisées par la plupart d'entre eux, puisqu'ils les considèrent également comme des tâches qu'ils souhaitent accomplir (voir la colonne « responsabilités idéalisées » du tableau 3). Barrère (2006, 2009), en France, dans une étude qualitative, observe une même adhésion à la culture de l'évaluation qui, selon elle, permet aux directeurs une certaine revalorisation professionnelle. Ceux-ci tendent ainsi aujourd'hui à se définir comme des « managers (de la République) » qui considèrent légitimes les nécessités de transparence et la recherche d'efficacité.

Tableau 3 : Les responsabilités assumées et idéalisées

	Responsabilités assumées				Responsabilités idéalisées			
	Entièrement responsable, rôle majeur*	Rôle mineur, aucunement responsable*	Total	N	Entièrement responsable, rôle majeur*	Rôle mineur, aucunement responsable*	Total	N
Administration et travail pédagogiques								
Élaboration du règlement	97,1 %	2,9 %	100,0 %	2056	94,2 %	5,8 %	100,0 %	1960
Élaboration de la mission de l'école	96,3 %	3,7 %	100,0 %	2058	94,3 %	5,7 %	100,0 %	1961
Affectation des tâches d'enseignement	90,7 %	9,3 %	100,0 %	2045	95,1 %	4,9 %	100,0 %	1960
Répartition des élèves	85,4 %	14,6 %	100,0 %	2050	85,8 %	14,2 %	100,0 %	1959
Définition des programmes et méthodes pédagogiques	75,3 %	24,7 %	100,0 %	2039	82,0 %	18,0 %	100,0 %	1960
Évaluation des programmes et méthodes pédagogiques	75,4 %	24,6 %	100,0 %	2042	84,8 %	15,2 %	100,0 %	1957
Choix du matériel didactique	52,5 %	47,5 %	100,0 %	2058	62,1 %	37,9 %	100,0 %	1974
Évaluation du matériel didactique	35,1 %	64,9 %	100,0 %	2052	46,6 %	53,4 %	100,0 %	1972
Évaluation et reddition de comptes								
Reddition de comptes	95,6 %	4,4 %	100,0 %	2044	89,5 %	10,5 %	100,0 %	1958
Analyse statistiques scolaires	78,0 %	22,0 %	100,0 %	2050	68,6 %	31,4 %	100,0 %	1966

	Responsabilités assumées				Responsabilités idéalisées			
	Entièrement responsable, rôle majeur*	Rôle mineur, aucunement responsable*	Total	N	Entièrement responsable, rôle majeur*	Rôle mineur, aucunement responsable*	Total	N
Gestion des ressources								
Allocation du budget	91,9%	8,1%	100,0%	2051	93,5%	6,5%	100,0%	1970
Gestion des fonds générés	85,6%	14,4%	100,0%	2050	79,5%	20,5%	100,0%	1966
Gestion des ressources matérielles	85,0%	15,0%	100,0%	2058	74,2%	25,8%	100,0%	1973
Élaboration du budget	77,9%	22,1%	100,0%	2052	87,5%	12,5%	100,0%	1970
Cueillette de fonds privés	50,9%	49,1%	100,0%	2038	35,4%	64,6%	100,0%	1954
Gestion des relations internes								
<i>Relations avec les enseignants</i>								
Supervision des enseignants	93,9%	6,1%	100,0%	2053	93,9%	6,1%	100,0%	1962
Recrutement des enseignants	65,9%	34,1%	100,0%	2036	93,2%	6,8%	100,0%	1976
Développement pédagogique des enseignants	62,1%	37,9%	100,0%	2043	73,0%	27,0%	100,0%	1967
<i>Relations avec les élèves</i>								
Sanction disciplinaire des élèves	92,0%	8,0%	100,0%	2054	81,1%	18,9%	100,0%	1959
Encadrement des élèves	79,6%	20,4%	100,0%	2049	67,4%	32,6%	100,0%	1950
Recrutement des élèves	25,7%	74,3%	100,0%	2010	35,2%	64,8%	100,0%	1934
<i>Gestion du personnel</i>								
Supervision du personnel professionnel	65,7%	34,3%	100,0%	2032	74,1%	25,9%	100,0%	1957
Supervision du personnel technique	60,2%	39,8%	100,0%	2028	63,8%	36,2%	100,0%	1945
Recrutement du personnel professionnel	29,6%	70,4%	100,0%	2025	65,3%	34,7%	100,0%	1989
Recrutement du personnel technique	35,9%	64,1%	100,0%	2012	65,7%	34,3%	100,0%	1972
Gestion des relations externes								
Participation aux comités de gestion	55,9%	44,1%	100,0%	2044	57,0%	43,0%	100,0%	1962
<i>Relations avec l'environnement</i>								
Sensibilisation du milieu	90,7%	9,3%	100,0%	2051	87,6%	12,4%	100,0%	1970
Développement de partenariats	74,9%	25,1%	100,0%	2050	72,6%	27,4%	100,0%	1969
<i>Relations avec les parents</i>								
Résolution des conflits avec les familles	85,8%	14,2%	100,0%	2046	76,1%	23,9%	100,0%	1966
Animation de la participation des parents	83,5%	16,5%	100,0%	2054	77,9%	22,1%	100,0%	1970
Formation des parents	55,0%	45,0%	100,0%	2050	49,6%	50,4%	100,0%	1968

* Afin d'alléger la présentation du tableau, nous avons regroupé, d'une part, les réponses « entièrement responsable » et « rôle majeur », et, d'autre part, « rôle mineur » et « aucunement responsable ».

La majorité des directeurs dit également assurer de nombreuses autres responsabilités, plus traditionnelles, telles que la supervision des enseignants, la gestion de l'ordre interne (sanction disciplinaire et encadrement des élèves), la gestion des ressources (allocation et élaboration du budget, gestion des fonds générés et des ressources matérielles) et la gestion des relations externes (sensibilisation du milieu, résolution des conflits avec les familles, animation de la participation des parents, développement de partenariats avec l'environnement de l'école). L'étude menée en France par Barrère (2006) met en évidence que cette multitude de tâches à effectuer peut être usante, d'autant plus qu'elles sont brèves, dispersées et fragmentées. En même temps, elle est également vécue par les directeurs d'école comme stimulante car elle favorise une « sensation grisante d'action » appréciée et renforce leur sentiment d'utilité. Notre enquête montre d'ailleurs que les directeurs eux-mêmes souhaiteraient, *idéalement*, exercer une multitude de tâches (voir la colonne « responsabilités idéalisées » du tableau 3). Cependant, on observe qu'ils valorisent surtout des responsabilités qui renvoient à un travail d'administration pédagogique (affectation des tâches d'enseignement, élaboration de la mission et du règlement de l'école), aux relations avec les enseignants (supervision et recrutement des enseignants) et à la gestion des ressources (allocation et élaboration du budget).

Une satisfaction professionnelle globale

Malgré la complexité et la lourdeur de leur tâche, notre enquête montre que la majorité des directeurs se déclare *entièrement* ou *plutôt satisfaits* de la plupart des aspects de leur métier (voir tableau 4). En particulier, plus de 80 % se disent satisfaits de leur développement professionnel (87,9 % en cumulant les deux premières positions de l'échelle), de l'appui de leurs supérieurs (84,6 %), de leur imputabilité (ou redevabilité) (82,3 %) et de leur autonomie professionnelle (81,0 %). À nouveau, la régulation par les résultats et la responsabilisation des directions d'établissement qu'elle implique ne paraissent pas être mal vécues par ceux-ci : en tout cas, la plupart d'entre eux se déclarent satisfaits de leur imputabilité. Ils sont par contre plus partagés quant à leur rémunération (52,6 % se disent satisfaits contre 47,4 % insatisfaits) et la plupart sont insatisfaits (entièrement ou plutôt) de leur charge de travail (53,3 %) et de son incidence sur leur vie familiale (63,9 %).

Tableau 4: La satisfaction professionnelle

	Entièrement satisfait	Plutôt satisfait	Plutôt insatisfait	Entièrement insatisfait	Total	N
Développement professionnel	29,6%	58,6%	10,0%	1,8%	100,0%	2082
Appui des supérieurs	43,3%	42,7%	10,5%	3,5%	100,0%	2073
Imputabilité	19,2%	63,7%	14,3%	2,8%	100,0%	2074
Autonomie professionnelle	23,5%	58,0%	15,3%	3,2%	100,0%	2077
Reconnaissance de la fonction	15,4%	44,8%	29,1%	10,7%	100,0%	2075
Encadrement du travail par des normes juridiques	7,0%	54,6%	30,9%	7,5%	100,0%	2063
Rémunération	11,9%	41,0%	32,8%	14,4%	100,0%	2074
Charge de travail	9,5%	37,3%	36,8%	16,3%	100,0%	2080
Incidence du travail sur la vie familiale	6,2%	30,5%	45,5%	17,8%	100,0%	2072

La surcharge de travail et la relation problématique entre vie privée et professionnelle sont fréquemment relevées dans la littérature sur le métier de directeur (Barrère, 2006; Bouchamma, 2004; Corriveau, 2004; King & Peart, 1992; Leclercq, 2005; Osborn, 2002) et elles peuvent aisément se comprendre si l'on considère la multitude de tâches à remplir. La satisfaction professionnelle globale qu'ils affichent néanmoins peut par contre paraître plus étonnante à première vue si on la compare à l'insatisfaction professionnelle exprimée en général par les enseignants (Cattonar, 2006). La satisfaction professionnelle se comprend cependant au regard des attentes à l'égard du métier et des motivations professionnelles qui interviennent dans le choix de la carrière (Michel, 1998). En nous appuyant sur d'autres études qui ont analysé les raisons qui guident le choix de quitter l'enseignement pour entamer la carrière de directeur, on pourrait alors supposer que leur satisfaction professionnelle globale est notamment liée à l'exercice de responsabilités et à la variété de leurs tâches, aspects du métier souvent évoqués parmi les motivations professionnelles les plus importantes (Barrère, 2006; Cacouault-Bitaud, 1998).

La surcharge de travail, même si elle pèse, notamment sur l'équilibre travail-famille, ne remet donc pas en cause la satisfaction professionnelle globale. Le changement voulu de statut et de fonction semble combler les directeurs et les installe dans une situation qu'ils perçoivent comme gagnante, même si celle-ci exige beaucoup de temps et d'énergie. L'intensification du travail semble vécue sur un mode assez positif, mais il faudrait poursuivre plus en profondeur l'analyse afin de mieux cerner les tensions entre la surcharge de travail et la satisfaction professionnelle.

QUELLE RESTRUCTURATION DU MÉTIER DE DIRECTION D'ÉTABLISSEMENT ?

En définitive, dans quelle mesure les données de notre enquête permettent-elles de conclure à une restructuration du métier de direction engagée par les récentes évolutions du mode de régulation scolaire ? Nous avons vu que lorsque l'on pose directement la question aux directeurs, la plupart affirment que les diverses évolutions scolaires, et en particulier le développement des pratiques d'évaluation, ont un impact sur la charge et la teneur de leur travail (accroissement de la charge de travail, modification de leur approche de la gestion...) et le développement de leurs compétences (exigence de nouvelles capacités d'adaptation). On pourrait aussi supposer, au vu des nombreuses responsabilités qu'ils disent exercer, qu'elles s'accompagnent d'une complexification de leur rôle.

Par ailleurs, on pourrait faire l'hypothèse que le développement de la régulation par les résultats renforce la supervision des enseignants par les directeurs et accroît leur implication dans les relations externes à l'école. En effet, on observe que plus les directeurs considèrent important l'impact des nouvelles politiques d'imputabilité et des évaluations standardisées des élèves et des enseignants, plus ils déclarent prendre en charge la supervision des enseignants, les relations avec les parents (notamment leur formation), la promotion de l'école dans la communauté et les liaisons avec les autorités (voir tableau 5). La régulation par les résultats semble ainsi renforcer le rôle des directions en tant qu'acteurs intermédiaires entre les différents paliers et acteurs scolaires (enseignants, parents, communauté, autorités). D'autres études ont également mis en évidence que le développement des évaluations externes s'accompagne d'une recomposition des relations entre directeurs et enseignants qui pose la question de l'autonomie professionnelle de ces derniers (Baluteau, 2009 ; Barrère, 2006). La préoccupation des directeurs pour les relations avec les parents et la communauté peut aussi se comprendre dans la mesure où, comme le souligne Barrère (2010), l'un des principaux enjeux du développement des évaluations externes est la constitution de la réputation de l'école dans un contexte où les établissements sont mis en concurrence pour le maintien du recrutement des élèves. De manière

schématique, on pourrait dire que les directeurs d'établissement, dans le contexte de la nouvelle régulation scolaire qui se met en place, sont appelés à développer une « professionnalité entrepreneuriale » (Demailly, 1998) caractérisée par une nouvelle autonomie et une capacité à se « positionner » sur un marché concurrentiel.

Tableau 5: Corrélations entre la perception de l'importance de la régulation par les résultats, les rôles et responsabilités assumés et la satisfaction professionnelle des directions d'établissement scolaire

	Nouvelles politiques d'imputabilité	Évaluation standardisée des élèves	Évaluation formelle des enseignants
Rôles assumés	+ superviseur du travail enseignant (0,101***), interlocuteur des parents (0,104***), promoteur de l'école dans la communauté (0,073***), agent de liaison auprès des autorités (0,087***)	+ superviseur du travail enseignant (0,155***), interlocuteur des parents (0,072***), promoteur de l'école dans la communauté (0,097***), agent de liaison auprès des autorités (0,065***)	+ superviseur du travail enseignant (0,180***), interlocuteur des parents (0,103***), promoteur de l'école dans la communauté (0,085***), agent de liaison auprès des autorités (0,074***)
Responsabilités assumées	+ supervision des enseignants (0,106***), formation des parents (0,111***), reddition comptes (0,110***), analyse des statistiques scolaires (0,092***)	+ supervision des enseignants (0,164***), formation des parents (0,107***), analyse des statistiques scolaires (0,073***)	+ supervision des enseignants (0,113***), analyse des statistiques scolaires (0,070***)
Satisfaction professionnelle	+ insatisfaits de leur charge de travail (- 0,159***), incidence sur vie familiale (- 0,106***)	+ insatisfaits de leur charge de travail (- 0,088***), incidence sur vie familiale (- 0,069***)	

Les valeurs dans le tableau représentent des Tau de Kendall. ***p < .001

Enfin, notre étude montre que la plupart des directions d'école semblent adhérer au principe de la régulation par les résultats et plus largement à la redéfinition de leur rôle, souhaitée par les pouvoirs publics et leurs formateurs, autour du leadership pédagogique et du pilotage des changements. Elles disent exercer et souhaiter exercer de multiples responsabilités liées à ces nouveaux rôles et pensent que les récentes évolutions scolaires auront, dans le futur, de nombreux effets positifs. On pourrait donc avancer que les directions d'école, au Canada, sont porteuses de la nouvelle régulation

qui se met en place. En même temps, il faut nuancer ce tableau positif. En effet, on observe que les nouvelles politiques d'imputabilité et l'évaluation standardisée des élèves tendent à être associées à une insatisfaction professionnelle plus forte concernant la charge de travail (voir tableau 5)¹⁶. En outre, comme nous l'avons vu, les directions d'école mettent aussi en avant plusieurs effets négatifs : diminution du sentiment de maîtrise des situations, augmentation de la méfiance et des rivalités, etc. Les récentes évolutions paraissent donc générer certaines tensions au sein des écoles, dont l'analyse reste à approfondir.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Baluteau, F. (2009). La division du travail pédagogique dans les établissements secondaires français. Le rôle des chefs d'établissement. *Sociologos. Revue de l'association française de sociologie*, 4, 1-20.
- Barrère, A. (2006). *Sociologie des chefs d'établissement. Les managers de la République*. Paris : PUF.
- Barrère, A. (2008). Les chefs d'établissement au travail : hétérogénéité des tâches et logiques d'action. *Travail et formation en éducation*, 2, 1-13.
- Barrère, A. (2009). Les directions d'établissement scolaire à l'épreuve de l'évaluation locale. *Carrefours de l'éducation*, 2(28), 199-214.
- Barrère, A. (2010). Ce que l'évaluation fait aux établissements scolaires. *Ethnologie française*, 40(1), 141-149.
- Ben Jaafar, S. & Anderson, S. E. (2004, mai-juin). *Accountability in Education: Trends in Canada from 1990-2003*. Paper presented at the Annual Meeting of the Canadian Society for Studies in Education, Winnipeg.
- Bergeron, D., Massé, D. & Rathé, J.-P. (2005). Réforme, décentralisation et professionnalisation des chefs d'établissement au Québec. *Politiques d'éducation et de formation. Analyses et comparaisons internationales*, 13(1), 65-78.
- Bernard, P. & Saint-Arnaud, S. (2004). Du pareil au même? La position des quatre principales provinces canadiennes dans l'univers des régimes providentiels. *Cahiers canadiens de sociologie*, 29(2), 209-239.
- Boissinot, M.-M. (2005). Chefs d'établissement en France : une situation paradoxale au centre de tensions multiples. *Politiques d'éducation et de formation. Analyses et comparaisons internationales*, 13(1), 53-64.
- Bouchamma, Y. (2004). Gestion de l'éducation et construction identitaire sur le plan professionnel. *Éducation et francophonie*, 32(2), 62-78.

16. D'autres évolutions sont également associées à une insatisfaction professionnelle plus grande concernant la charge de travail : il s'agit de la diminution des ressources et de la nouvelle répartition des responsabilités (les corrélations ont été mesurées par des Tau de Kendal dont les valeurs varient entre 0,10 et 0,20 et sont significatives au seuil de 1 %).

- Brassard, A., Cloutier, M., De Saedeleer, S., Corriveau, L., Fortin, R., Gélinas, A. & Savoie-Zajc, L. (2004). Rapport à l'activité éducative et identité professionnelle chez les directeurs d'établissement des ordres d'enseignement préscolaire et primaire. *Revue des sciences de l'éducation*, 30(3), 487-508.
- Cacouault-Bitaud, M. (1998). Les personnels de direction du second degré dans trois académies: déterminants d'accès aux fonctions et représentations du rôle. In R. Bourdoncle & L. Demailly (Éd.), *Les professions de l'éducation et de la formation* (pp. 339-350). Paris: Presses universitaires du Septentrion.
- Cattonar, B. (2001). Les identités professionnelles enseignantes. Ébauche d'un cadre d'analyse. *Les Cahiers de Recherche du Girsef*, 10, 1-35.
- Cattonar, B. (2006). Comment les enseignants du secondaire conçoivent-ils et vivent-ils leur métier? In G. Chapelle & D. Meuret (Éd.), *Améliorer l'école* (pp. 153-163). Paris: PUF.
- Cattonar, B., Lessard, C., Blais, J.-G., Larose, F., Riopel, M.-C., Tardif, M., Bourque, J. & Wright, A. (2007). *Les directeurs et les directrices d'école au Canada: contexte, profil et travail* (Rapport de recherche, Grands travaux de recherche du CRSH). Université de Montréal, CRIFPE-Montréal.
- Corcuff, P. (1995). *Les nouvelles sociologies. Constructions de la réalité sociale*. Paris: Nathan Université.
- Corriveau, L. (2004). Identification professionnelle ou suridentification à la profession? *Éducation et francophonie*, 32(2), 95-110.
- Corriveau, L., Fortin, R., Gélinas, A., Savoie-Zajc, L., Héon, L. & De Saedeleer, S. (avec Chevrier, E.). (2001). *Les directions d'établissement et les changements en éducation* (Rapport d'étape). Université de Montréal, Groupe de recherche interuniversitaire sur les directions d'établissement scolaire (GRIDES).
- Demailly, L. (1998). Les métiers relationnels de service public: approche gestionnaire, approche politique. *Lien social et Politiques*, 40, 17-24.
- Deniger, M.-A., Kamanzi, C., Chabot, N. & Hébert, C. (2006). *Évaluation du nouveau programme de formation de l'école québécoise: la qualité de sa mise en œuvre et ses effets perçus à ce jour. Enquête auprès des directions d'école, du personnel enseignant, des professionnels non enseignants et des parents des écoles primaires* (Collection Études et recherches, vol. 6, n° 1). Québec, Université Laval.
- Dupriez, V. (2002). Entre tradition et transformations, que révèle l'analyse des établissements scolaires? In C. Maroy (Éd.), *Les écoles d'enseignement secondaire et leurs enseignants* (pp. 55-90). Bruxelles: De Boeck Université.
- Dupriez, V. (2005). La transformation du métier de chef d'établissement en Communauté française de Belgique. *Politiques d'éducation et de formation. Analyses et comparaisons internationales*, 13(1), 11-25.

- Dutercq, Y. (2006). Les nouveaux pouvoirs des chefs d'établissements. In G. Chappelle & D. Meuret (Éd.), *Améliorer l'école* (pp. 141-152). Paris: PUF.
- Hargreaves, A., Lieberman, A., Fullan, M. & Hopkins, D. (1998). *International Handbook of Educational Change*. London: Kluwer Academic Press.
- King, A. J. C. & Peart, M. J. (1992). *Le personnel enseignant au Canada. Travail et qualité de vie* (Étude nationale réalisée pour le compte de la Fédération canadienne des enseignantes et des enseignants). Groupe d'évaluation des programmes sociaux, Université Queen's à Kingston, Ottawa.
- Lapointe, P., Garon, R., Brassard, A., Dupuis, P., Chapel, C. & Brunet, L. (2009). *La gestion des activités éducatives des directeurs et des directrices d'écoles primaires dans le contexte de la réforme en éducation au Québec* (Rapport de recherche présenté au Fonds québécois de la recherche sur la société et la culture et au ministère de l'Éducation, du Loisir et du Sport, projet2005-AC-103493). Consulté le 15 janvier 2011 dans [http://www.fqrcs.gouv.qc.ca/upload/editeur/RF-PierreLapointe103493\(1\).pdf](http://www.fqrcs.gouv.qc.ca/upload/editeur/RF-PierreLapointe103493(1).pdf)
- Leclercq, J.-M. (2005). Perspective internationale: quelle direction pour les établissements scolaires? *Politiques d'éducation et de formation. Analyses et comparaisons internationales*, 13(1), 79-100.
- Lessard, C. & Brassard, A. (2005). La « gouvernance » de l'éducation au Canada: tendances et significations. *Éducation et Sociétés*, 18, 181-201.
- Lessard, C., Kamanzi, C. & Larochelle, M. (2008). La perception des politiques éducatives chez les directions d'école et les enseignants canadiens: l'influence de l'idéologie professionnelle. *Sociologie et Sociétés*, 40(1), 93-118.
- Maroy, C. & Dupriez, V. (2000). La régulation dans les systèmes scolaires. Proposition théorique et analyse du cadre structurel en Belgique francophone. *Revue française de pédagogie*, 13, 73-87.
- Michel, S. (1998). Motivation et implications professionnelles. In M. De Coster & F. Pichault (Éd.), *Traité de sociologie du travail* (pp. 403-421). Bruxelles: De Boeck Université.
- Murphy, J. & Seashore-Louis, K. (Ed.). (1999). *Handbook of Research on Educational Administration*. San Francisco: Jossey-Bass.
- Osborn, M. (2002). Le nouveau rôle des chefs d'établissements: le cas de l'Angleterre. *Politiques d'éducation et de formation. Analyses et comparaisons internationales*, 5(2), 67-80.
- Pelage, A. (2003). La redéfinition du métier de chef d'établissement secondaire: changement statutaire, construction de l'engagement professionnel et épreuves pratiques. *Revue française de pédagogie*, 145, 21-36.
- Philips, S., Raham, H. & Reniham, P. (2003). *The Role of the School Principal, Present Status and Future Challenges in Managing Effective Schools*. Kelowna: Society for the Advancement of Excellence in Education.

Évaluation, concertation, décision: quelle régulation pour le système éducatif? Le cas de l'enseignement primaire genevois

Georges Solaux*, Jean-Marc Huguenin,
Jean-Paul Payet*** & José V. Ramirez******

***Université de Bourgogne, **IDHEAP¹, Lausanne**

*****Université de Genève ****HEG² et**

Leading House en économie de l'éducation, Genève

Suite à trois années de concertation, le Département de l'instruction publique du canton de Genève a mis en place un « nouveau fonctionnement » de l'enseignement primaire à la rentrée scolaire de septembre 2008. La mise en œuvre du nouveau dispositif était accompagnée de la création d'un outil d'évaluation externe appelé « Observatoire du fonctionnement de l'enseignement primaire »³ (dans la suite du texte, dénommé Observatoire). L'analyse de cet outil d'évaluation fait l'objet du présent article. La problématique développée vise à montrer que, si l'expertise en matière d'évaluation de l'éducation n'est pas la science, elle peut toutefois être modestement fondée et sans « grandiloquence conceptuelle » (Lallement, 1993) sur les valeurs de la production des connaissances, voire contribuer à quelque production de connaissances limitées aux « hypothèses mineures qui jaillissent chaque jour à foison dans le travail quotidien de la recherche » (Merton, 1965, pp. 13-14).

1. Institut de hautes études en administration publique.

2. Haute école de gestion.

3. Voir le Mandat relatif à l'Observatoire du fonctionnement de l'enseignement primaire pour les années 2009-2012 du 27 août 2009, consultable à l'adresse : <http://www.geneve.ch/dip/GestionContenu/detail.asp?mod=communiquer.html&id=1303>

Notre propos tentera de tisser quelques liens entre les notions de science, d'expertise et de décision démocratique, nous inscrivant ainsi dans la problématique développée par Weber dans deux conférences prononcées en 1917 et 1919 (*Wissenschaft als Beruf* le 7 novembre 1917, et *Politik als Beruf* le 28 janvier 1919) lorsqu'il pose la question du sens de la science dans sa relation entre « la profession et la vocation de savant » (Weber, 2003). Il y adopte la réponse donnée par Tolstoï : « Elle n'a pas de sens parce qu'elle ne peut répondre à la seule question qui importe pour nous, laquelle est de savoir ce que nous devons faire et comment nous devons vivre »⁴ (cité par Weber, 2003, p. 90). Sans avoir la prétention d'apporter la contradiction à la réponse de Tolstoï, notre ambition consiste à montrer qu'il a été possible, dans le cadre de cette évaluation de l'enseignement primaire genevois, d'apporter des informations relevant du champ scientifique aux acteurs sociaux en vue de les aider à déterminer des choix ou des pistes d'action pour mieux vivre l'école.

Pour développer cette proposition, une première partie contextualisera la place de l'Observatoire dans le cadre du nouveau fonctionnement de l'enseignement primaire. Dans une deuxième partie, les caractéristiques de l'Observatoire seront rapportées à celles de la réforme dont il est censé évaluer la mise en œuvre. Dans une troisième partie, quelques éléments de justification de la dimension pluri et interdisciplinaire des méthodes d'investigation mobilisées et des analyses réalisées seront proposés. Enfin la quatrième partie illustrera, par un exemple, la mise en œuvre de l'approche interdisciplinaire.

LE NOUVEAU FONCTIONNEMENT DE L'ENSEIGNEMENT PRIMAIRE DU CANTON DE GENÈVE

Le nouveau fonctionnement de l'enseignement primaire du canton de Genève repose sur un long processus de concertation⁵. En septembre 2005, le conseiller d'État⁶ chargé du Département de l'instruction publique du

4. Voir note de bas de page 90 (Weber, 2003), « la référence va à la critique que fait Tolstoï dans la première partie de son essai de critique sociale : *Que faire?* »

5. Le nouveau fonctionnement de l'enseignement primaire s'inscrit dans une longue temporalité de concertation au niveau politique. Il ne saurait être question d'isoler la période récente retenue ici (de 2003 à aujourd'hui), qui correspond à la nomination d'un nouveau conseiller d'État chargé du Département de l'instruction publique, des périodes précédentes, notamment de celle dite de la « rénovation ».

6. En Suisse, l'éducation est l'un des nombreux domaines dont la compétence relève principalement du niveau cantonal. Il faut cependant noter un mouvement actuel vers une certaine centralisation, appelé « harmonisation », au niveau fédéral. Il n'est pas inutile d'avoir à l'esprit que ce processus d'harmonisation entre les politiques éducatives cantonales constitue un des arrière-plans normatifs de la réforme du fonctionnement ici analysée.

canton de Genève réunit les représentants des enseignants de l'enseignement primaire (SPG)⁷, des inspecteurs de l'enseignement primaire (AIG)⁸, des formateurs (CEFEP)⁹, des secrétaires et des différents niveaux de l'administration du Département de l'instruction publique (DIP) et de la Direction générale de l'enseignement primaire (DGEP). Le conseiller d'État annonce aux « partenaires » son intention de faire évoluer le « fonctionnement » de l'enseignement primaire sur la base de propositions émanant du groupe de travail présent, qui deviendra rapidement la « Commission de fonctionnement ». Cette dernière, présidée par le Directeur général de l'enseignement primaire, reçoit un premier mandat le 8 décembre 2005. Elle se réunira 25 fois entre septembre 2005 et juin 2008, soit plus de six fois par année. Le nouveau fonctionnement est mis en place en août 2008. Trois années de concertation furent donc nécessaires pour adopter un système répondant aux souhaits et revendications parfois contradictoires des différents acteurs.

La justification de la réforme est inspirée d'études, analyses et évaluations tombées dans le domaine public. De manière générale, elle s'inscrit dans le cadre du plan d'action décidé par la Conférence suisse des directeurs cantonaux de l'instruction publique suite aux résultats obtenus par les élèves suisses au Programme international pour le suivi des acquis des élèves (PISA) de 2000 (Conférence suisse des directeurs cantonaux de l'instruction publique, 2003)¹⁰. Les élèves et leurs performances sont prioritairement ciblés par la recherche d'un système produisant davantage d'efficacité et d'équité. En effet, l'école primaire genevoise se caractérise par une performance de ses élèves qui se situe en dessous de la moyenne suisse dans les trois disciplines – mathématiques, sciences et lecture – testées par PISA (Huguenin, 2011). Les deux objectifs de recherche de plus grande efficacité et équité conduisent alors à proposer un système moins uniforme et davantage centré sur les besoins locaux, sur les caractéristiques spécifiques des populations scolaires. Dans le cadre de la recherche d'une plus grande équité, un Réseau d'enseignement prioritaire (REP) prévoyant des allocations de ressources plus importantes aux écoles présentant les profils socioculturels les plus faibles est créé. Il s'ensuivra la décision de donner aux établissements une autonomie partielle.

Une autre justification est inspirée d'enquêtes réalisées auprès du personnel enseignant en 2002 et 2003. L'enquête Santé 2002¹¹ mettait en évi-

7. Société pédagogique genevoise.

8. Association des inspecteurs genevois.

9. Centre de formation continue de l'enseignement primaire.

10. Ce plan d'action prévoit notamment l'autonomisation des écoles avec l'introduction ou le renforcement d'organes de direction associés à l'examen périodique des objectifs de performance.

11. «Évaluation des relations entre les conditions de travail et la santé du personnel de l'enseignement primaire du Département de l'instruction publique», juin 2003, Action en santé publique, Versoix, Suisse.

dence la faiblesse du soutien hiérarchique. Il s'ensuit la décision d'installer une hiérarchie de proximité dans les écoles avec la création d'un corps de directeurs d'établissement primaire¹².

Un troisième axe de motivation est constitué par la volonté de mener une réflexion en lien avec les *13 priorités pour l'Instruction publique*¹³ énoncées par le DIP en 2003, et notamment la priorité numéro 13 qui est centrée sur le souci de promouvoir de nouveaux modes de régulation au niveau cantonal comme suit :

Les décisions politiques feront l'objet d'évaluations régulières en matière d'efficacité, de résultats et de coûts du système. Ces évaluations donneront ensuite lieu systématiquement à des concertations avec l'ensemble des milieux intéressés, tout particulièrement les collaborateurs du DIP, les associations professionnelles et syndicales ; celles de parents, d'élèves ou d'étudiants¹⁴.

À la rentrée scolaire d'août 2008, le nouveau fonctionnement propose donc la mise en œuvre d'actions destinées à atteindre les objectifs d'efficacité, d'efficience et d'équité, supposant un soutien hiérarchique de proximité et la modernisation de la régulation du système tant au niveau local qu'au niveau cantonal. Institutionnellement, le regroupement des 163 écoles du canton de Genève en 91 établissements scolaires répartis en cinq régions géographiques repose sur la mise en place des projets d'établissement, sur un dispositif de concertation avec la création des Conseils d'établissement, et enfin sur la création du corps de directeurs d'établissement primaire, autorité hiérarchique locale responsable de l'autonomie de l'établissement dans le cadre des lois et règlements existants, mais aussi hiérarchie de proximité pour les enseignants.

Au niveau cantonal, la nouvelle régulation repose sur l'inflexion de la fonction administrative de la DGEP, qui doit évoluer d'une direction de gestion à une direction d'objectifs, sur la définition d'un nouveau mandat¹⁵ délivré à la Commission de fonctionnement jusqu'en 2012 et sur l'installation d'un Observatoire du fonctionnement de l'enseignement primaire¹⁶, dont la création était par ailleurs revendiquée depuis 1999 par le syndicat majori-

12. Voir le site de la Direction générale de l'enseignement primaire: <http://icp.ge.ch/ep/etidep>

13. Voir http://www.geneve.ch/dip/13_priorites.asp

14. Voir http://www.geneve.ch/dip/13_priorites.asp

15. Voir le Mandat relatif à la Commission de fonctionnement de l'enseignement primaire pour les années scolaires 2009-2012 du 2 décembre 2009, consultable à l'adresse: <http://www.geneve.ch/dip/GestionContenu/detail.asp?mod=communique.html&id=1303>

16. En 2002 déjà, le Département de l'instruction publique avait mis en place une « Commission scientifique de suivi et de conseil relative à la phase d'extension de la rénovation de l'enseignement primaire » composée « d'universitaires et de chercheurs aux compétences reconnues » (voir le Mandat relatif à cette Commission scientifique du 15 mars 2002, p. 2, à consulter à la DGEP, rue Calvin, Genève).

taire des enseignants du primaire (SPG). Les associations de parents d'élèves et les communes du canton intègrent la Commission de fonctionnement. On retrouve ainsi au niveau local, dans les Conseils d'établissement, et au niveau cantonal, dans la Commission de fonctionnement, les mêmes acteurs : l'administration, les enseignants, les parents d'élèves et les communes. En tant que nouveaux modes de régulation, les Conseils d'établissement et la Commission de fonctionnement peuvent être considérés comme

des modes d'ajustement permanents d'une pluralité d'actions et de leurs effets, permettant d'assurer l'équilibre dynamique de systèmes instables [...]. S'il y a régulation, c'est parce que les règles ne peuvent tout prévoir, doivent être interprétées, remises en cause (en adaptation perpétuelle en fonction des situations et objectifs). La régulation de tout groupe social correspond aux interactions entre les intérêts particuliers de chaque composante du groupe et l'intérêt commun ou général de celui-ci. (Bauby, 2002, p. 15)

Le nouveau fonctionnement de l'enseignement primaire se caractérise ainsi par une double action qui vise à la fois les structures et les acteurs. Les structures évoluent avec la création des établissements, des régions et des Conseils d'établissement; les acteurs sont visés par l'autonomie partielle donnée aux établissements qu'ils sont chargés de faire vivre. Ce caractère particulier et exigeant explique l'importance prise par les termes *fonctions* et *processus* dans le suivi de la mise en œuvre de la réforme.

Conceptuellement, le nouveau fonctionnement s'inscrit dès lors dans la décentralisation du management des systèmes scolaires que connaissent depuis un certain nombre d'années la plupart des pays anglo-saxons, et plus particulièrement l'Australie où le concept de *School-based management* a été développé dans les années 1960 (sous la pression du *Australian Capital Territory* [ACT]) comme alternative à un modèle de management centralisé au niveau national (Zajda & Gamage, 2009). L'idée d'un conseil d'établissement, regroupant des représentants des enseignants, des parents et de la communauté locale, constitue le principal fondement de ce modèle de management. Bien que cette idée soit commune à l'ensemble des modèles visant une plus grande autonomie des établissements scolaires, il existe bel et bien un continuum entre un système scolaire totalement centralisé et celui où prévaut une totale décentralisation de l'autorité pouvant aller jusqu'à la privatisation de la production scolaire. Dans le cas de Genève, l'objectif déclaré est de faire évoluer le système d'une position intermédiaire mais proche d'une forte centralisation (au niveau cantonal) à une position toujours intermédiaire mais se rapprochant de l'autre pôle du continuum. Bien que la réforme ne vise aucunement une privatisation du système scolaire, telle qu'elle est pratiquée aux États-Unis ou en Angleterre, le mouvement engagé à Genève est forcément accompagné de tensions, notamment liées à l'instauration d'une certaine forme de concurrence entre les établissements.

L'OBSERVATOIRE DU FONCTIONNEMENT DE L'ENSEIGNEMENT PRIMAIRE

Le mandat délivré à l'Observatoire, qui engage les experts de 2009 à 2012, comprend les missions suivantes :

Mener des observations, dans une perspective opérationnelle, sur les fonctions, processus et impacts générés par le nouveau fonctionnement de l'enseignement primaire, à tous les échelons du système [...], évaluer la portée des changements par rapport aux objectifs du nouveau fonctionnement, proposer des ajustements, organiser des observations en concertation avec la Commission de fonctionnement de l'enseignement primaire. (Mandat de l'Observatoire pour les années 2009-2012, 2009, p. 1)

Le mandat délivré à la Commission de fonctionnement prévoit, quant à lui, deux missions essentielles : d'une part, analyser les « fonctions, les processus et impacts générés par la nouvelle organisation » (Mandat de l'Observatoire pour les années 2009-2012, 2009, p. 1), d'autre part « proposer à la DGEP les ajustements qu'il est nécessaire d'apporter en s'appuyant sur les rapports de l'Observatoire et sur les constats effectués sur le terrain » (*ibid.*). On note une forte articulation entre les mandats des deux organismes. Si le premier évalue les fonctions, processus et impacts générés par le nouveau fonctionnement au regard des objectifs politiques définis, la Commission de fonctionnement est chargée d'analyser ces mêmes fonctions, processus et impacts en s'appuyant sur les rapports de l'Observatoire. On se trouve donc en présence « d'un processus unique et original de co-construction des savoirs dont le résultat vise à soutenir l'élaboration de politiques publiques en éducation » (Saint-Pierre, 2006, p. 73).

Autrement dit, les savoirs « savants » produits par l'Observatoire doivent comporter une dimension opérationnelle d'expertise et s'inscrire dans le cadre de l'action en proposant des ajustements. Ces savoirs sont confrontés ensuite aux savoirs pratiques des acteurs du système éducatif, confrontation qui génère des avis, des réactions, des contradictions et *in fine* des propositions dont se saisit ou non l'administration scolaire.

Le mandat de l'Observatoire présente une triple exigence d'objectifs déjà citée plus haut : « Évaluer la portée des changements par rapport aux objectifs du nouveau fonctionnement, proposer des ajustements, organiser des observations en concertation avec la Commission de fonctionnement de l'enseignement primaire ». L'objectif « organiser des observations en concertation avec la Commission de fonctionnement de l'enseignement primaire » répond à la volonté de transparence et de démocratisation des prises de décision. Afin de s'inscrire dans une démarche qui associe le caractère nécessairement externe de toute évaluation avec la concertation exigée, la Commission de fonctionnement a souhaité, sur proposition de

l'Observatoire, que la méthode de travail de ce dernier respecte les standards d'évaluation de la Société suisse d'évaluation (Widmer, Landert & Bachmann, 2000). Ceux-ci précisent notamment que

la planification, l'exécution et la présentation d'une évaluation encouragent les parties prenantes et groupes concernés à suivre le processus d'évaluation et à tirer parti de l'opération [... car] le degré d'application des résultats et des recommandations de l'évaluation dépend étroitement de la mesure dans laquelle les parties prenantes et groupes concernés attendent que l'évaluation ait une utilité pour eux-mêmes et pour leurs projets. Une condition importante pour encourager et satisfaire cette attente consiste à inclure les différents protagonistes dans la planification et l'organisation du projet d'évaluation. (p. 6)

Les missions d'évaluation de l'Observatoire sont ainsi définies en Commission de fonctionnement, les acteurs sont auditionnés pour donner leurs avis sur le projet du rapport d'évaluation, avis intégrés au rapport présenté à la Commission de fonctionnement. En outre, une phase supplémentaire d'échanges entre l'Observatoire et chaque partenaire peut avoir lieu, en cas d'incompréhension ou de désaccord.

L'objectif « évaluer la portée des changements par rapport aux objectifs du nouveau fonctionnement » inscrit l'évaluation dans une perspective normative, dans la mesure où les fonctions, processus et impacts doivent être évalués en fonction des principes fondateurs du nouveau fonctionnement et non en fonction de principes portés par les évaluateurs. C'est ainsi que le référent commun de jugement de valeur des experts consiste à déterminer si les actions évaluées concourent à :

1. développer une autonomie partielle des établissements susceptible de promouvoir l'efficacité et l'équité de l'enseignement primaire;
2. installer une hiérarchie de proximité porteuse d'effets sur les équipes enseignantes mais aussi responsable administrative et pédagogique de l'autonomie partielle des établissements;
3. ancrer la concertation dans le fonctionnement des établissements, notamment par les Conseils d'établissement qui associent les parents, les communes et les enseignants.

Chaque action est évaluée à l'aune de ces critères d'autonomie, d'efficacité, d'efficience, d'équité, d'encadrement hiérarchique et de concertation, et il s'agit d'apporter des éléments de réponse à la question suivante inspirée de Chevallier (2005, p. 389) : intrinsèquement porteuse d'innovation, une politique est-elle pour autant porteuse de changement ou, mieux encore, des changements attendus ?

L'objectif « proposer des ajustements » suppose que les experts, au-delà du jugement porté sur l'action évaluée, recommandent d'éventuelles

modifications organisationnelles. Ce travail suppose à la fois une culture des systèmes éducatifs et la capacité d'entendre les suggestions des partenaires. La culture des systèmes éducatifs existants permet de puiser dans les innovations introduites dans les enseignements primaires hors du contexte local du canton de Genève les actions qui ont exercé l'impact recherché. C'est ainsi que les travaux réalisés par exemple sur la décentralisation des systèmes éducatifs dans le monde et en Europe (Mons, 2004) fournissent un certain nombre de mises en garde qui ne peuvent être ignorées lorsqu'on travaille sur la dimension de l'autonomie de l'établissement. Les suggestions d'évolution peuvent également émaner des partenaires qu'il faut savoir écouter tout en faisant la part des choses entre revendication catégorielle et recherche de l'intérêt général. La difficulté pour l'expert réside dans le fait que ces deux aspects ne sont d'ailleurs pas toujours contradictoires et qu'une revendication catégorielle peut parfois servir l'intérêt général. La culture des systèmes et la capacité d'écoute permettent aux experts de délibérer sur ce qu'il conviendrait de faire pour atteindre plus rapidement les objectifs fixés.

Ces contraintes exercées sur l'évaluation limitent-elles l'action des experts et interdisent-elles d'utiliser la notion d'expertise scientifique pour qualifier les jugements portés ?

UNE ÉQUIPE D'EXPERTS

Les systèmes éducatifs sont des organisations complexes qui sont étudiées par un champ disciplinaire lui-même empreint de cette complexité dans la mesure où, en général, il est institutionnellement appelé « sciences de l'éducation ». Le pluriel des « sciences » montre que de nombreuses sciences humaines et sociales peuvent être mobilisées pour produire des connaissances sur le système éducatif. En fonction des objets d'évaluation proposés par le mandat de l'Observatoire et des objectifs qui lui sont assignés, il était difficile de se limiter à un seul champ disciplinaire. En effet, lorsqu'il s'agit d'évaluer des « fonctions », on pénètre le champ de la gestion des ressources humaines, du management, de la sociologie. L'évaluation des « processus » suppose une approche par les sciences de gestion, la sociologie des organisations. L'évaluation des « impacts » sur les mentalités, sur les représentations, sur l'efficacité, l'équité et l'efficience nécessite la mobilisation de techniques d'enquêtes et de mesure de la performance. Le champ cognitif proposé à l'Observatoire est donc très large par les objets et par les méthodes d'investigation à mobiliser. Pour cette raison, chacun des experts est reconnu comme spécialiste dans un des champs d'investigation suivants : la sociologie, l'économie politique et le management.

La sociologie politique permet de fonder l'action d'évaluation sur le fait que la « culture est un enjeu, un ensemble de ressources et de modèles que

les acteurs sociaux cherchent à gérer, à contrôler, qu'ils s'approprient ou dont ils négocient entre eux la transformation en organisation sociale » (Touraine, 1984, p. 32). Les entreprises de réforme des systèmes éducatifs constituent des redéfinitions de la culture scolaire et donc de la culture vivante qui sera partagée par les générations à venir. Dans le cadre qui nous intéresse, l'enjeu est de taille et le moment opportun pour les acteurs sociaux de tenter de renforcer leur identité lors des évaluations et concertations afin d'introduire dans le dispositif la part de culture qu'ils jugent indispensable pour l'avenir. La connaissance sociologique permet d'une part de maîtriser « les mécanismes à travers lesquels des orientations culturelles sont transformées en pratiques sociales » (Touraine, 1984, p. 98) et d'autre part d'identifier les stratégies d'acteurs. Les évaluateurs sont l'objet de la pression des acteurs dominants qui, sous le couvert de l'intérêt général, tentent d'obtenir des avantages. La professionnalité enseignante constitue par exemple un point constant de crispation, aussi bien chez les enseignants que chez les parents d'élèves, entre les défenseurs d'un enseignant traditionnellement seul dans sa classe, et compris fondamentalement comme un agent de transmission des connaissances, et les promoteurs d'un enseignant travaillant en équipe et davantage tourné vers la transmission de méthodes de travail ou de technologies intellectuelles. Le social pénètre ainsi l'école par les conceptions et idéologies que véhiculent les acteurs intéressés à, et par son fonctionnement. Le travail de l'Observatoire confirme que le système éducatif apparaît comme un lieu de rencontre d'acteurs fréquemment opposés, comme un lieu de gestion de conflits. Les experts peuvent maîtriser le fonctionnement de cette organisation sociale au moyen des connaissances produites par la sociologie politique qui permet notamment de comprendre les rapports de force et les conflits générés par les stratégies d'acteurs (Solaulx, 2000).

Les rapports de force inhérents à toute organisation sont de plus en plus étudiés par les économistes par le biais du modèle « principal-agent » (Laffont & Martimort, 2002), d'inspiration hobbesienne (Bowles, 1985). De manière schématique, le modèle suppose que l'objectif du principal est de maximiser l'écart entre le bénéfice qu'il retire du niveau de performance atteint par l'agent et le coût nécessaire pour atteindre un tel niveau. Dans cette logique, l'agent fera en sorte d'atteindre le niveau de performance attendu par le principal tout en minimisant son coût. Le tissu des multiples relations principal-agent existant à chaque niveau hiérarchique du système (dans le cas qui nous intéresse ici, de la relation DIP-DGEP à celle entre enseignant et élève) permet ainsi d'évaluer ce dernier en termes de structuration de l'information et des incitations auxquelles font face l'ensemble des acteurs le composant. Dans un système scolaire définissant clairement les objectifs, l'une des questions abordées par cette littérature concerne le lien entre le système de rémunération et de compensation des enseignants et les indicateurs sur lesquels ces derniers peuvent avoir une influence (Lazear, 2003), que ce soit de manière individualisée ou plus collectivement, comme par exemple au niveau des

équipes d'enseignants des établissements scolaires. Bien que ces recherches sur la structure incitative caractérisant les systèmes scolaires ne s'inscrivent pas *stricto sensu* dans l'économie de l'éducation, elles émanent en partie des principaux résultats empiriques liés à ce champ d'investigation, à savoir que les systèmes d'éducation doivent être évalués à l'aune des *outputs* produits (Hanushek, 2003). Au-delà de l'idéologie néo-libérale généralement associée à ce type de modèle, l'économie de l'éducation en tant que champ d'étude en plein essor n'en propose pas moins un certain regard critique et souvent davantage contextualisant que les autres champs d'étude en économie, notamment sur les tensions généralement associées au processus de décentralisation dans les systèmes scolaires. Weiler (1993) par exemple propose de distinguer trois modèles de décentralisation : le modèle *redistributif* qui discute la distribution du pouvoir du haut vers le bas ; le modèle d'*efficacité* qui se focalise sur les aspects financiers et l'efficacité de la décentralisation ; et le modèle des *cultures apprenantes* qui se centre sur la diversité culturelle et l'acceptabilité des programmes (ou normes) au niveau local. Dans certains cas, il peut être argué que la décentralisation des systèmes scolaires, contrairement à l'objectif généralement recherché, peut conduire à un accroissement des inégalités locales, et ce quand bien même le système n'a pas été privatisé. L'analyse des inégalités et des différentes formes de concurrence dans la production éducationnelle est sans aucun doute d'une grande importance dans le cadre de la réforme actuelle du système d'enseignement primaire dans le canton de Genève.

Les résultats des principales recherches menées ces dernières années en économie de l'éducation font clairement ressortir l'importance d'orienter l'évaluation des systèmes scolaires en fonction des *outputs* – ce qui n'implique pas forcément que la quantité et la qualité des ressources allouées aux systèmes ne jouent aucun rôle. Ils permettent de souligner l'importance d'un management des systèmes scolaires qui soit en mesure d'identifier, de mesurer et d'évaluer la performance à l'aune des objectifs qu'ils visent, quand bien même ceux-ci sont de nature multidimensionnelle (Gerard, 2001).

Le management est l'« action ou art ou manière de conduire une organisation, de la diriger, de planifier son développement, de la contrôler » (Thiéart, 2010, p. 7). Cette définition situe l'action d'évaluation dans un cadre cohérent plus large composé de quatre fonctions génériques (planification, organisation, activation – des ressources – et évaluation) qui visent à organiser et à piloter les ressources mises en œuvre pour favoriser la performance du système et des entités qui le composent. Par essence interdisciplinaire, le management public intègre les développements, les pratiques, les outils ou les contraintes issus de différentes disciplines comme l'administration publique, l'économie politique, la science politique, la sociologie, la psychologie, le droit ou la

pédagogie¹⁷. Trois dimensions du management peuvent être identifiées : le management stratégique, le management opérationnel et le contrôle de gestion (*controlling*). Tant le management stratégique que le management opérationnel ont une portée décisionnelle que le contrôle de gestion n'a pas. Par contre, ce dernier joue un rôle de pourvoyeur d'informations et d'analyses au profit de la prise de décision, notamment en matière d'évaluation de la performance. Schöni (2006) évoque ainsi la notion de *controlling* éducationnel (*Bildungskontrolling*). Le développement de tableaux de bord équilibrés (Kaplan & Norton, 2005), notamment dans les établissements scolaires genevois, fait partie intégrante du contrôle de gestion (Weber & Schäffer, 1999). La performance en management se définit en fonction des objectifs fixés, des résultats obtenus – *outputs*, impacts, *outcomes* – et des moyens mis en œuvre – ces trois éléments formant les sommets du triangle de gestion développé par Gilbert (1980). L'évaluation de la performance est ainsi basée sur les critères d'efficacité (lien entre objectifs et résultats), d'efficience (relation entre moyens et résultats) mais également de pertinence (articulation entre moyens et objectifs). Dans le cadre des travaux de l'Observatoire, la fonction d'évaluation du management doit pouvoir s'appuyer sur des objectifs clairement identifiables, spécifiques et mesurables. De manière générale, elle pose les questions suivantes : le cadre budgétaire de la réforme (moyens constants) est-il adapté aux objectifs fixés – critère de pertinence? Les résultats obtenus correspondent-ils aux objectifs – critère d'efficacité? Le rapport entre les résultats et les ressources est-il optimisé – critère d'efficience?

La sociologie assume une fonction traditionnellement critique à l'égard des pouvoirs institués. On note, a priori, une contradiction entre posture sociologique et posture évaluative, qui s'est communément traduite par une forte prise de distance de la sociologie académique à l'égard de l'expertise, tout au moins au niveau des discours. Dans la pratique, l'opposition est moins évidente. Les travaux sociologiques sont bien connus des experts, et ceux-ci sont également dans nombre de cas des sociologues dits académiques, adhérant au postulat critique de leur discipline, mais le mettant dès lors à l'épreuve dans une activité pratique d'évaluation et de conseil aux décideurs, politiques ou économiques. La critique radicale du pouvoir, qui serait systématiquement suspecté d'exercer une violence légitime (Weber, 2003) ou symbolique (Bourdieu & Passeron, 1970), ne peut constituer la

17. En Suisse, des Assises romandes de l'éducation sont organisées annuellement par les syndicats d'enseignants. En 2010, elles ont abordé le thème : « Veut-on des pilotes dans les écoles? Les directions d'établissement en question ». Dans leur texte de présentation, elles relèvent ainsi que « pour de nombreux cantons, comme pour l'OCDE, l'effet direction est un élément clé de l'efficacité d'un système éducatif. La directrice ou le directeur d'établissement devrait être pédagogue, expert en évaluation, rassembleur, décideur, expert en relations publiques et communication, gestionnaire, garant de la légalité et de l'équité, gentil organisateur, médiateur... » (Syndicat des enseignants romands [SER], 2010, p. 18).

posture de l'expert, mais celle-ci se fonde néanmoins sur l'adhésion à des valeurs qui légitiment la pertinence générale de la réforme faisant l'objet de l'évaluation. Dans le cas de la réforme actuelle du système d'enseignement primaire dans le canton de Genève, les valeurs de l'égalité des élèves devant l'école en dépit des inégalités liées au milieu social, de l'utilité et de l'utilisation optimale du bien public, de la maîtrise démocratique des processus de changement fondent cette adhésion. Dans le processus de mise en œuvre d'une politique, les valeurs générales sont cependant altérées, voire dévoyées. L'apport de la sociologie est d'être attentif à ce hiatus, inévitable mais explicitable – et donc maîtrisable – entre valeurs et réalisations (Becker, 1983), de repérer les effets pervers qui dénaturent l'esprit d'une réforme. Il y a donc une forme de vigilance, en quelque sorte « éthique », constitutive du travail de l'Observatoire, qui s'inquiète de ce que les intentions politiques soient encore présentes dans leurs effets – par effets, nous entendons également des effets au niveau de la perception de la réforme par les acteurs, qui de fait se traduit par des attitudes qui permettent ou empêchent le changement (Merton, 1965). La méthodologie mise en œuvre détermine *in fine* si la pluralité des points de vue est prise en compte par l'évaluation, et dans quelle mesure les voix moins légitimes, plus faibles (Payet, Rostaing & Giuliani, 2010) sont perçues et traduites. Une posture attentive aux rapports de pouvoir, aux dissymétries dans l'activité de légitimation (vs de délégitimation), aux formes et aux stratégies de construction de la réalité participe du caractère démocratique de l'évaluation.

C'est le croisement de ces quatre regards théoriques et méthodologiques complémentaires sur le même objet qui permet de tendre vers une compréhension aussi complète que possible du phénomène observé. Les quatre champs disciplinaires en proposent des interprétations plurielles, et leur articulation constitue en soi une forme de délibération « démocratique ». Mais l'articulation est rendue d'autant plus ardue que le champ de l'évaluation est lui-même complexe et que les savoirs savants ne se déploient pas sans opposition.

APPROCHE INTERDISCIPLINAIRE APPLIQUÉE À LA RECONFIGURATION DES ÉTABLISSEMENTS

L'une des missions de l'Observatoire a été d'évaluer un projet de reconfiguration des établissements primaires développé par la Commission de fonctionnement. Ce projet visait la réduction du nombre d'établissements et, par conséquent, l'augmentation du nombre moyen d'élèves par établissement¹⁸.

18. En 2008, 91 établissements sont créés; le nombre d'établissements actuel est de 90. En moyenne, chacun des établissements accueille 381 élèves (écart type = 117). L'établissement médian compte quant à lui 370 élèves. Cela signifie qu'une moitié des établissements

Les ressources totales allouées aux établissements demeurant constantes, une réduction du nombre d'établissements devait permettre de dégager des moyens à réallouer aux grands établissements, dont la dotation en personnel administratif se révèle insuffisante.

La première étape du travail a consisté à construire l'objet à évaluer. Quelle signification donner au terme « reconfiguration » ? Cette dernière comporte une dimension économique, sociologique, organisationnelle, financière et, par le caractère pratique de sa mise en œuvre, managériale. De manière générale, les champs disciplinaires représentés par l'économie politique, la sociologie et le management s'interrogent chacun sur les conditions nécessaires à l'amélioration de l'efficacité, de l'équité et de l'efficience des systèmes éducatifs. Dans le cas présent d'une reconfiguration, ils s'interrogent plus spécifiquement sur l'impact de la taille des établissements sur ces trois principes.

Les méthodes et techniques d'investigation des différents champs disciplinaires varient cependant pour apporter une réponse à la même problématique. Celles-ci comportent des dimensions quantitatives et qualitatives à appréhender dans leur complémentarité. Ainsi, l'économie politique permet de quantifier l'impact d'une augmentation de la taille des établissements (nombre d'élèves) sur la performance des élèves, sur l'efficience des écoles ou sur l'équité. La sociologie, par ses techniques d'enquête et d'entretien, attire l'attention sur les conditions des prises de position favorables ou défavorables des acteurs face à un tel projet. Le management propose des méthodes issues de la recherche opérationnelle pour évaluer l'efficience des établissements et identifier leur situation en termes de rendements d'échelle. Il permet également, dans le cadre de la gestion financière, de développer une méthode d'allocation des ressources permettant d'atteindre une plus grande équité et de favoriser l'efficience.

Dans le cas présent, l'Observatoire a mobilisé les méthodes d'investigation suivantes :

- exploitation des résultats d'une enquête sur le nouveau fonctionnement réalisée en 2010 auprès des enseignants, des directeurs d'établissement et des parents d'élèves (sociologie, sociométrie);
- auditions d'un panel de directeurs et d'enseignants issus d'établissements de configurations différentes (sociologie);
- auditions des partenaires, notamment les organisations professionnelles et les parents (sociologie);
- analyse comparative de la méthode d'allocation des ressources en personnel administratif entre les cantons de Genève, de Vaud et de Bâle-Ville (management);

accueil plus de 370 élèves. Le plus petit établissement compte 157 élèves, le plus grand 726.

- étude de deux revues de la littérature scientifique portant sur les effets de la taille des établissements (Leithwood & Jantzi, 2009; Slate & Jones, 2005) (économie politique, sociologie, management);
- analyse d'efficience¹⁹; identification des situations de rendements d'échelle croissants et décroissants (économie politique, management).

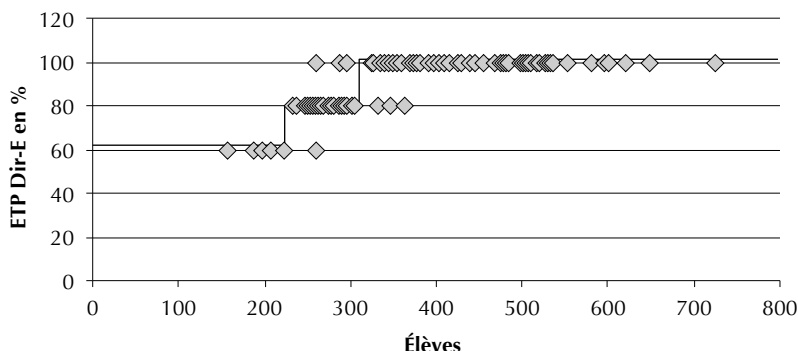
L'analyse d'efficience a été réalisée sur les 90 établissements du canton. Elle met en relation les ressources consommées par les établissements (*inputs*: personnel enseignant et personnel administratif – directeurs et secrétaires – en équivalents temps plein) avec les effectifs et les résultats des élèves de chaque établissement (*outputs*: nombre d'élèves et résultats aux épreuves cantonales de référence). L'analyse s'est concentrée sur l'efficience d'échelle, autrement dit l'efficience liée à la taille des établissements. Elle montre qu'en moyenne, l'efficience d'échelle des établissements atteint 98%. La consommation moyenne de ressources associée à la taille des établissements est ainsi calibrée de manière adéquate dans les établissements. Une reconfiguration des établissements se traduirait par une économie de ressources de l'ordre de 2% seulement²⁰.

Une analyse comparative effectuée entre les cantons de Genève, Vaud et Bâle-Ville montre des différences importantes dans les dotations en personnel administratif et les méthodes d'allocation des ressources. À Genève, les ressources en personnel de direction sont allouées en fonction du nombre de classes réparties en trois catégories, alors qu'elles sont fondées sur le nombre d'élèves dans les cantons de Vaud et de Bâle-Ville. Les grands établissements genevois disposent ainsi de dotations de direction (par élève) nettement inférieures à celles des petits établissements d'une part et des deux autres cantons d'autre part. Le graphique 1 représente l'allocation en personnel de direction (équivalents temps plein) selon le nombre d'élèves par établissement dans le canton de Genève. Nous y identifions deux effets de seuil dus au fait que les ressources sont allouées en fonction de trois catégories de nombres de classes. Ces effets de seuil nuisent à l'égalité et à l'équité entre établissements.

19. L'analyse d'efficience est réalisée à l'aide de la méthode *Data Envelopment Analysis*. Pour un aperçu synthétique de la méthode, voir Taylor (2010). Pour un descriptif complet, voir Cooper, Seiford & Tone (2006).

20. Pour ce faire, il faudrait que la taille des établissements identifiés comme étant en situation de rendements d'échelle croissants soit augmentée, tandis que la taille des établissements évoluant en situation de rendements d'échelle décroissants soit réduite.

Graphique 1 : Allocation des ressources et effets de seuil



ETP Dir-E en% : allocation en personnel de direction en équivalents temps plein, selon le nombre d'élèves par établissement.

Les auditions menées révèlent que les différents acteurs concernés regrettent le manque d'information, le flou et le refus de négociation émanant de l'administration centrale, initiatrice du projet de reconfiguration des établissements. Elles suggèrent que l'ensemble des critères susceptibles d'être pris en compte pour les reconfigurations soit diffusé. Ce sentiment d'absence d'information conduit à la propagation de rumeurs sur la stratégie du management soupçonné d'utiliser ce projet comme levier d'économie. Par ailleurs, les acteurs souhaitent être associés à la prise de décision concernant leur établissement et participer à la définition des modalités pratiques de regroupement de leurs établissements.

Une analyse de la littérature scientifique sur l'effet taille des établissements complète ces entretiens. Elle permet de mettre en évidence le fait que les petits établissements sont plus efficaces pour atteindre les objectifs fixés en permettant notamment d'améliorer les résultats d'élèves provenant de milieux défavorisés sans pour autant péjorer les résultats d'élèves venant de milieux plus favorisés. Enfin, il est à relever que la taille des établissements accueillant des populations défavorisées ne devrait pas aller au-delà de 300 élèves, tandis que celle des établissements accueillant des populations hétérogènes ou favorisées ne devrait pas aller au-delà de 500 élèves²¹ (Leithwood & Jantzi, 2009). Parmi les explications données concernant l'efficacité et l'efficience supérieures des petits établissements, les auteurs citent une culture d'établissement plus affirmée et fondée sur le sentiment d'appartenance à une communauté locale, une connaissance plus grande des élèves par les enseignants et une personnalisation de leur intervention plus adaptée à leurs besoins.

21. Il s'agit d'effectifs pour des établissements d'enseignement primaire.

Sur la base de l'évaluation effectuée, l'Observatoire a proposé un certain nombre de mesures aux partenaires de la Commission de fonctionnement. Il attire l'attention sur le fait que la reconfiguration éventuelle doit être fondée prioritairement sur les principes d'efficacité, d'équité et d'efficience des établissements. Parmi les recommandations proposées, retenons ici²² :

- une réduction du nombre d'établissements ne se justifie pas pour des raisons d'efficience d'échelle dans le cas présent ;
- la nécessité d'accompagner tout projet de reconfiguration de la diffusion aux enseignants, parents et communes à la fois de l'exposé des motifs par établissement concerné et de la planification des reconfigurations (voir résistances et manque d'information des enseignants) ;
- la révision de la mécanique de l'allocation des ressources en personnel administratif pour qu'elle soit fondée sur le nombre d'élèves et leurs caractéristiques socio-économiques.

Ces informations communiquées aux acteurs puis discutées en Commission de fonctionnement ont parfois généré un malaise, car elles ne correspondent pas toujours aux attentes des acteurs et questionnent leurs actions et revendications. Les quelques savoirs nouveaux ou « savants » se heurtent alors aux savoirs « pratiques ». Il en est ainsi du système d'allocation des moyens qui interroge l'administration centrale sur ses pratiques de gestion, qui demeurent essentiellement caractérisées par le centralisme. Par exemple, l'allocation des ressources aux établissements ne permet pas la mise en œuvre de leur autonomie. Les dotations sont effectuées sans que des objectifs explicites soient clairement identifiables, ce qui empêche toute évaluation a posteriori.

L'ensemble de ces observations, communiquées aux acteurs, est alors débattu avec parfois, pour certains d'entre eux, une mise en question de l'objectivité des méthodes d'investigation de l'Observatoire, tant les résultats sont contraires à leurs attentes. Pour les uns (la DGEP), il paraissait évident que la réduction du nombre d'établissements s'imposait, à la fois pour des raisons de cohérence, d'efficacité et de coût. Pour les autres (la SPG), toute réduction du nombre des établissements apparaissait comme une concession par rapport à une position au moment de la négociation de la réforme, où le syndicat des enseignants avait ferraillé pour un nombre d'établissements équivalent au nombre d'écoles. En convoquant des données issues de la littérature internationale, le propos de l'Observatoire déplaçait l'enjeu d'un face-à-face local entre acteurs vers une recherche d'objectivité en lien avec les valeurs et les objectifs de la réforme. En écoutant et en relayant les frustrations des acteurs face à un mode de décision perçu comme unilatéral, il soulignait la contradiction entre la survivance d'un mode traditionnel de management par

22. Pour la totalité des propositions, se référer à : http://icp.ge.ch/ep/etidep/IMG/pdf/observatoire_note_synthetique_novembre_2010.pdf

le haut et l'objectif d'accroître l'autonomie des établissements. En articulant la question de la commande (reconfigurer la taille des établissements) avec celle de l'allocation des moyens, il mettait l'accent sur l'enjeu d'équité sociale, lié à la concurrence entre établissements. Ces différents déplacements de la question, opérés par l'évaluation, ont créé une situation dans laquelle chaque acteur était amené à se positionner par rapport à un objet commun mais redéfini par un regard externe. D'une certaine manière, et idéalement, le travail de l'Observatoire consiste à la fois à donner une légitime place au conflit tel qu'il est vécu par les acteurs tout en proposant de le traiter par une lecture nouvelle d'un problème, dont les acteurs peuvent faire un objet commun, en dépassant les clivages traditionnels.

L'expérience montre que trois catégories de situations peuvent se présenter. Dans une première configuration, l'acteur concerné par l'évaluation est en accord avec les conclusions de l'Observatoire et envisage avec les autres acteurs des solutions. Dans une seconde configuration, l'acteur concerné par l'évaluation est en désaccord avec les conclusions de l'Observatoire, mais les autres acteurs acceptent les conclusions et font pression pour que des solutions soient envisagées. Dans un troisième cas de figure, l'acteur concerné par l'évaluation et les autres acteurs sont en désaccord avec les conclusions de l'Observatoire, et il est alors difficile d'envisager des solutions.

CONCLUSION

En articulant au plus près approche scientifique et discussion démocratique, le processus retenu évite que ne s'installe une « république des experts » négligeant le fait que les acteurs doivent être mobilisés pour qu'une politique ait quelque chance de réussir. Ce parti pris d'efficacité dans le processus de construction des décisions de gestion de l'enseignement primaire garantit-il pour autant l'absence de l'idéologie ou des idéologies dans la prise de décision? Le fait d'avoir entendu une pluralité d'opinions dans les auditions et les enquêtes menées par l'Observatoire et dans le cadre de la Commission de fonctionnement est-il le gage d'une plus grande objectivité? La prise de décision politique sur la base d'opinions ainsi largement sollicitées est-elle plus objectivement et plus scientifiquement fondée que la prise de décision sur la base de quelques opinions? Compte tenu du fait que les discussions relatives aux choix à effectuer sont fondées sur les données de l'Observatoire dont les analyses externes reposent sur la mobilisation de schèmes explicatifs issus de quatre champs disciplinaires, on peut faire l'hypothèse que si l'idéologie persiste, son impact sur la décision est plus faible que si les données de l'Observatoire n'existaient pas.

En effet, si chacun des acteurs s'exprime avec ses convictions, ces dernières sont de fait confrontées aux résultats tangibles de l'évaluation préalable.

Grâce aux rapports des experts, le fonctionnement du système éducatif perd de son opacité, et il devient possible d'identifier en toute rigueur les variables pertinentes qui agissent effectivement sur l'objet du débat. Ce mode de gouvernement laisse alors moins de place à l'épanouissement des stratégies d'acteurs, et à des prises de décision qui seraient d'abord fonction de l'opinion des acteurs dominants.

Cet ensemble d'observations amène à poser le problème des relations entre science et politique. Faut-il soumettre le politique au scientifique au point de développer une « intelligentsia » technocratique à tous les niveaux du pouvoir d'État? Des exemples récents concernant la politique de scolarisation pour tous dans les pays en développement ont montré les limites de planifications prétendument fondées sur une base scientifique (Akkari & Payet, 2010). Par ailleurs, la transformation d'une société ne repose pas seulement sur la gestion des moyens actuels et à venir, mais également sur des projets, des visions et des rêves. Les projets et les rêves dépassent le scientifique et concernent beaucoup plus directement le politique. Enfin, la science ne risquerait-elle pas de devenir à son tour un acteur parmi d'autres acteurs en évoluant vers ce qu'il est convenu d'appeler une « science de commande » qui perdrait toute liberté dans la détermination des champs, objets et résultats de recherche pour être limitée à l'immédiatement et socialement utile?

RÉFÉRENCES BIBLIOGRAPHIQUES

- Akkari, A. & Payet, J.-P. (Éd.). (2010). *Transformations des systèmes éducatifs dans les pays du Sud. Entre globalisation et diversification* (Collection Raisons éducatives). Bruxelles: De Boeck.
- Bauby, P. (2002). L'Europe des services publics: entre libéralisation, modernisation, régulation, évaluation. *Politique et management public*, 1, 15-30.
- Becker, H. (1983). Studying urban schools. *Anthropology and Education Quarterly*, 14(2), 99-108.
- Bourdieu, P. & Passeron, J.-C. (1970). *La Reproduction. Éléments pour une théorie du système d'enseignement*. Paris: Minuit.
- Bowles, S. (1985). The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models. *American Economic Review*, 76(1), 16-36.
- Chevallier, J. (2005). Politiques publiques et changement social. *Revue Française d'Administration Publique*, 115, 383-390.
- Conférence suisse des directeurs cantonaux de l'instruction publique. (2003). *Mesures consécutives à PISA 2000: Plan d'action*. Consulté le 10 mars 2011 dans www.edudoc.ch/static/web/arbeiten/pisa2000_aktplan_f.pdf
- Cooper, W. W., Seiford, L. M. & Tone, K. (2006). *Introduction to Data Envelopment Analysis and its Uses*. New York: Springer.

- Gerard, F.-M. (2001). L'évaluation de la qualité des systèmes de formation. *Mesure et évaluation en éducation*, 24(2-3), 53-77.
- Gilbert, P. (1980). *Le contrôle de gestion dans les organisations publiques*. Paris : Éditions d'Organisation.
- Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, 113, 64-98.
- Huguenin, J.-M. (2011). Politique de l'éducation et conjoncture en Suisse. In N. Soguel (Éd.), *Des politiques au chevet de la conjoncture* (chapitre 3). Lausanne : Presses polytechniques et universitaires romandes.
- Kaplan, R. S. & Norton, D. P. (2005). *Le tableau de bord prospectif*. Paris : Éditions d'Organisation.
- Laffont, J.-J. & Martimort, D. (2002). *The Theory of Incentives : The Principal-Agent Model*. Princeton, NJ : Princeton University Press.
- Lallement, M. (1993). *Histoire des idées sociologiques*. Paris : Nathan.
- Lazear, E. P. (2003). Teachers Incentives. *Swedish Economic Policy Review*, 10, 179-214.
- Leithwood, K. & Jantzi, D. (2009). A Review of Empirical Evidence About School Size Effects : A Policy Perspective. *Review of Educational Research*, 79(1), 464-490.
- Mandat de l'Observatoire du fonctionnement de l'enseignement primaire pour les années 2009-2012, Secrétariat Général du Département de l'Instruction Publique, Genève le 27 août 2009, consultable à l'adresse : <http://www.geneve.ch/dip/GestionContenu/detail.asp?mod=communiquer.html&id=1303>
- Merton, R. K. (1965). *Éléments de théorie et de méthode sociologique*. Paris : Plon.
- Mons, N. (2004). Politiques de décentralisation : diversité internationale, légitimations théoriques et justifications empiriques. *Revue française de pédagogie*, 146, 41-52.
- Payet, J.-P., Rostaing, C. & Giuliani, F. (Éd.). (2010). *La relation d'enquête. La sociologie au défi des acteurs faibles*. Rennes : Presses Universitaires de Rennes.
- Saint-Pierre, C. (2006). Entre recherche pratique et politique : le Conseil supérieur de l'éducation du Québec, une démarche originale. *Revue française de pédagogie*, 154, 73-84.
- Schöni, W. (2006). *Handbuch Bildungscontrolling : Steuerung von Bildungsprozessen in Unternehmen und Bildungsinstitutionen*. Zürich : Verlag Rüegger.
- Slate, J. R. & Jones, C. H. (2005). Effects of school size : A review of the literature with recommendations. *Essays in Education*, 13.
- Solau, G. (Éd.). (2000). *L'évaluation des politiques d'éducation*. Paris : CNDP, Centre national de documentation pédagogique.
- Syndicat des enseignants romands. (2010). Veut-on des pilotes dans les écoles ? Les directions d'établissement en question. *Éducateur*, 10, 18.

- Taylor, L. L. (2010). Economic Approaches to School Efficiency. In D. J. Brewer & P. J. McEwan (Ed.), *Economics of Education* (pp. 208-213). Oxford and San Diego: Elsevier.
- Thiétart, R. A. (2010). *Le management*. Paris: PUF.
- Touraine, A. (1984). *Le retour de l'acteur*. Paris: Fayard.
- Weber, J. & Schäffer, U. (1999). *Balanced Scorecard & Controlling: Implementierung – Nutzen für Manager und Controller – Erfahrungen in deutschen Unternehmen*. Wiesbaden: Gabler.
- Weber, M. (2003). *Le savant et le politique*. Paris: La Découverte.
- Weiler, H. (1993). Control vs Legitimation: The Politics of Ambivalence. In J. Hannaway & M. Carnoy (Ed.), *Decentralization and School Improvement: Can We Fulfill The Promise?* (pp. 55-83). San Francisco: Jossey-Bass.
- Widmer, T., Landert, C. & Bachmann, N. (2000). *Standards d'Évaluation de la Société Suisse d'Évaluation (Standards SEVAL)*. Consulté le 10 janvier 2011 dans <http://www.seval.ch/fr/>
- Zajda, J. & Gamage, D. T. (Ed.). (2009). *Decentralisation, School-Based Management and Quality*. New York: Springer.

L'évaluation des élèves : externalisation et hétérogénéité¹

José V. Ramirez

**HEG² et Leading House en économie de l'éducation,
Genève**

La publication des résultats de l'enquête PISA 2000, centrée sur les compétences en littéracie, a soulevé en Suisse et dans d'autres pays ayant obtenu un « mauvais » classement, comme l'Allemagne et la France, des réactions mêlant surprise et populisme. Les conclusions hâtivement tirées ont réduit l'explication du mauvais classement aux médiocres résultats des élèves issus de l'immigration. Simple comme un stéréotype. Toutefois, les résultats de PISA 2006, centrée cette fois sur les mathématiques et les sciences, montrent que malgré un score moyen en mathématiques parmi les plus élevés des pays de l'OCDE, les « écarts relevés en Suisse pour les enfants de milieux socio-économiques modestes ou issus de l'immigration ne sont pas vraiment à la hauteur de la qualité d'ensemble du système d'enseignement » (Organisation de coopération et de développement économiques [OCDE], 2011, p. 2). En récoltant des informations parfaitement comparables dans le temps et entre pays, les évaluations réalisées par le biais des grandes enquêtes internationales telles que TIMSS, PISA ou encore PIRLS, permettent d'enrichir considérablement les données « locales » pour l'étude du lien

1. L'auteur tient à remercier le Département de l'instruction publique (DIP) du canton de Genève et le Service de la recherche en éducation (SRED) pour lui avoir facilité l'accès aux données de la base de données scolaires (BDS) et de la base de données des évaluations communes (EVACOM) dans le cadre du projet sur les « Déterminants de la transition entre le secondaire I et le secondaire II » dans le canton de Genève mené au sein de la Leading House en économie de l'éducation de Genève et financé par l'Office fédéral de la formation et des technologies (OFFT).

2. Haute école de gestion.

entre les résultats scolaires et les caractéristiques familiales des élèves tout en considérant les effets de « contexte ». Car si la question des différences de performance scolaire selon l'origine familiale des enfants est depuis longtemps explorée (Bourdieu & Passeron, 1970; Coleman, 1966), elle est aujourd'hui d'autant plus incontournable dans un monde caractérisé par d'importants flux migratoires qui ne cessent d'accroître l'hétérogénéité des populations, principalement dans les pays à relativement hauts revenus. Un phénomène d'hétérogénéité qui concerne davantage la population d'élèves que celle des enseignants, ce qui n'est probablement pas sans effet sur l'interaction enseignant-élève.

Cet article s'intéresse au lien entre les performances scolaires et les caractéristiques socio-économiques et sociodémographiques des élèves dans le contexte de l'externalisation (et la standardisation) des évaluations. Nous discutons ce lien en partant de l'hypothèse que l'information produite par les évaluations externes peut modifier la distribution des résultats scolaires. Cette information, lorsqu'elle est considérée à l'externe du système scolaire, génère des signaux créateurs d'incitations pouvant accentuer les inégalités de résultats généralement observées en fonction de l'origine familiale des élèves (Wössmann, 2005). En connaissance de cause, les établissements peuvent toutefois réduire ces effets par l'instauration de pratiques s'inscrivant dans une politique de discrimination positive. Néanmoins, en postulant que les évaluations faites par les enseignants ne sont qu'une mesure imparfaite de la qualité éducative d'un enfant (Rosenthal & Jacobson, 1968), on peut raisonnablement se demander si, indépendamment d'une politique de discrimination positive, les évaluations externes ne constituent pas aussi en elles-mêmes un outil efficace pour diminuer les inégalités. La question principale est de savoir si l'usage des évaluations externes (ou partiellement externes) peut renforcer l'équité sociale, du fait qu'il est moins affecté par les inefficiences sociales liées à l'interaction enseignant-élève.

Dans la section qui suit, nous discutons de l'information généralement produite par les évaluations externes, et présentons les principales caractéristiques de ce type d'évaluations en Suisse et plus particulièrement dans le canton de Genève. Dans la deuxième section, nous discutons dans le cadre du modèle principal-agent les hypothèses faites quant au comportement des agents en lien avec les signaux et incitations associés aux évaluations externes. La section suivante présente une discussion sur les effets négatifs générés par l'interaction enseignant-élève sur l'évaluation des performances scolaires. La quatrième section présente une illustration empirique à partir des différences entre le résultat à l'évaluation annuelle standardisée (et partiellement externalisée) et la moyenne annuelle en mathématiques au secondaire inférieur du canton de Genève en 2005-2006. Un canton dont près de la moitié de la population d'élèves fréquentant l'école publique obligatoire déclare une autre langue que le français comme première langue

(Service de la recherche en éducation [SRED], 2009), et dont le système de scolarité obligatoire fait partie des 26 systèmes en vigueur en Suisse. La dernière section conclut.

L'EXTERNALISATION DE L'ÉVALUATION DES ÉLÈVES

Aux extrêmes du continuum représentant les logiques de l'évaluation se trouvent l'évaluation formative et les grandes enquêtes internationales : d'un côté, il s'agit d'évaluer les acquis des élèves de manière interne, de l'autre d'évaluer le système scolaire dans son ensemble (Wirthner & Ntamakiliro, 2008). En tant qu'outil, les évaluations externes remplissent de multiples fonctions, notamment celle d'informer, car elles sont génératrices « de données quantitatives aisément comparables » (Mons, 2009, p. 10). Elles permettent dès lors un certain contrôle, vers l'interne par les directions d'établissement et les autorités politiques, et vers l'externe par les parents (voire les médias). Les modalités de communication des résultats peuvent accentuer les réticences à la culture d'évaluation centralisée, qui plus est lorsqu'il s'agit d'évaluations externes à fort enjeu³.

En Suisse, les évaluations externes, communément appelées épreuves ou évaluations cantonales (EC), ne sont pas tout à fait externes, car si elles sont bien élaborées par des personnes externes à la classe et à l'établissement, elles sont administrées et généralement corrigées par les enseignants titulaires de la matière dans la classe. Elles sont toutefois totalement standardisées, de la conception à la correction en passant par les conditions de passation. Ces EC existent dans tous les cantons romands et ont pour objectif premier d'évaluer les acquis des élèves (notamment aux divers points de transition entre cycles scolaires), mais sont également exploitées pour le monitoring du système, sur la base essentiellement d'une comparaison des résultats de chaque établissement scolaire par rapport à la moyenne cantonale (Institut de recherche et de documentation pédagogique, 2010). Les nouvelles épreuves romandes, prévues dans le cadre du concordat HarmoS⁴, entreront en vigueur progressivement dès la rentrée scolaire de 2011 (voire 2012)

3. La manière de communiquer les résultats aux évaluations externes peut grandement varier d'un pays à l'autre. Si l'objectif est d'accroître l'efficacité du système scolaire par une plus grande concurrence entre établissements, il y a de fortes chances que la communication résulte en un classement des établissements. Un classement qui peut générer de nombreux effets pervers à l'interne du système (démotivation des équipes d'enseignants dans les établissements accueillant majoritairement des élèves provenant de milieux socio-économiques défavorisés, réduction du curriculum, bachotage pour préparer les élèves), mais également à l'externe, en augmentant le risque de ghettoïsation urbaine. Qui plus est si les parents peuvent librement choisir l'établissement scolaire de leurs enfants et qu'il y a peu de freins à la mobilité sur le marché du logement (Tiebout, 1956).

4. «L'accord intercantonal sur l'harmonisation de la scolarité obligatoire (concordat HarmoS) est un nouveau concordat scolaire suisse. Il harmonise pour la première fois au

dans tous les cantons de Suisse romande. Elles s'inscrivent davantage dans le contexte des enquêtes internationales, à savoir le pilotage du système. Il est en effet prévu que la conception et l'administration, voire la correction, soient externalisées. De cette manière, la Suisse (notamment sa partie francophone) déplacera clairement sa position sur le continuum des pratiques de l'évaluation dans le but d'évaluer et de piloter le système scolaire à un niveau plus centralisé. Autrement dit, l'ensemble du continuum des pratiques de l'évaluation pourrait dès lors être couvert.

Seul le canton de Genève pratique toutefois les EC durant les trois degrés du secondaire inférieur. D'autres cantons romands (Fribourg, Vaud et Valais) y recourent uniquement pour le dernier degré du secondaire inférieur. À l'instar des autres cantons romands, la communication des résultats des EC à Genève se fait essentiellement vers l'intérieur du système. Les parents et les élèves n'ont généralement accès qu'aux résultats individuels, les enseignants aux statistiques des résultats de l'établissement. Les établissements et autorités reçoivent les taux de réussite de l'ensemble des cohortes. Enfin, les écoles et associations professionnelles engageant des apprentis ont également accès aux résultats cantonaux, et bien entendu à ceux de leurs candidats à l'apprentissage.

En termes d'organisation et de déroulement, les EC en mathématiques au secondaire inférieur sont préparées dans des commissions qui sont essentiellement formées par des enseignants, puis validées par la direction générale du secondaire inférieur (Soussi, Guilley, Guignard & Nidegger, 2009). La passation et la correction sont généralement faites par les enseignants de la matière. Elles portent en principe sur les deux tiers voire les trois quarts du programme annuel et, à l'instar des EC dans les autres matières, leurs poids varient entre 20% et 25% dans la moyenne de la troisième et dernière période de l'année scolaire. La valeur effective des EC dans l'évaluation annuelle des élèves est par conséquent relativement faible, quand bien même leur poids symbolique est certainement plus important, du moins pour l'administration, comme l'atteste le fait qu'elles figurent séparément dans le livret scolaire.

INSTITUTIONNALISATION DE L'ÉVALUATION EXTERNE: SIGNAUX ET INCITATIONS

L'hypothèse communément admise dans les modèles principal-agent est que les multiples parties composant un système scolaire répondent à des incitations par des choix et des comportements pouvant générer des externalités, négatives ou positives (e.g. effets de pairs). Dès lors, qu'elle soit

niveau suisse la durée des degrés d'enseignement, leurs principaux objectifs et le passage de l'un à l'autre » (Conférence suisse des directeurs de l'Instruction publique, 2009).

donnée par une école publique ou privée, la production éducationnelle a généralement une nature de bien public, voire toujours lorsqu'il s'agit de production éducationnelle dans une classe. Une autre caractéristique commune importante à ces modèles est de considérer le temps et l'effort alloués par les élèves comme les *inputs* les plus importants dans le processus de production éducationnelle.

Costrell (1994) propose un modèle théorique dans lequel les autorités fixent les standards éducationnels et les élèves choisissent leur niveau d'effort en fonction de ces standards de manière à maximiser leur niveau d'utilité. La principale conclusion du modèle est que l'institutionnalisation d'évaluations centralisées (que ce soit au niveau national ou régional) dans un système scolaire génère une hausse des standards, des résultats et procure des gains sociaux plus importants. Sous certaines conditions, le modèle prévoit également que l'imperfection de l'information sur la qualité éducationnelle des élèves puisse résulter en des situations sous-optimales, tant au niveau individuel que collectif, notamment lorsque la variance dans la distribution des performances scolaires est élevée⁵. À partir d'un modèle de production éducationnelle dans lequel la relation principal-agent se limite à l'État et à l'élève (les choix et les comportements des autres agents – enseignants, parents, directions d'établissement – étant considérés comme exogènes), Bishop et Wössmann (2004) confirment les résultats théoriques de Costrell. L'évaluation centralisée génère un signal sur la qualité éducationnelle produite à l'intérieur de chaque classe de nature plus homogène que celui généralement associé aux évaluations internes, permettant une comparaison de la performance scolaire des élèves dans des classes et établissements différents et un accroissement de la qualité du signal envoyé tant aux degrés ultérieurs qu'aux employeurs potentiels vers lesquels les élèves transitent par la suite. C'est donc essentiellement parce qu'elles modifient la structure incitative à laquelle font face les élèves que les évaluations externes augmentent le rendement externe de l'effort qu'ils fournissent durant leur cursus.

L'une des faiblesses de ces modèles réside néanmoins dans l'hypothèse d'homogénéité quant aux effets générés par l'existence d'évaluations externes sur la performance scolaire des élèves. Wössmann (2005) suggère que cette hypothèse s'avère restrictive si l'on prend explicitement en compte dans le modèle le comportement des parents et des établissements, tout en considérant le facteur temps. L'hétérogénéité potentielle des effets liés à l'existence d'évaluations externes réside dans la nature et l'importance du

5. Le modèle prévoit qu'une information parfaite augmente l'effort et les performances des élèves se situant dans le haut et le bas de la distribution, mais pas de ceux se trouvant au centre de la distribution. Pour un certain nombre de niveaux d'hétérogénéité (en termes de qualité éducationnelle des élèves), l'amélioration de l'information augmente le bien-être social, notamment dans un système scolaire décentralisé.

signal envoyé à l'extérieur du système. Ces effets seraient plus importants en fin de scolarité obligatoire lorsque la probabilité d'entrer sur le marché du travail augmente : plus la proportion d'employeurs potentiels prenant en compte les résultats de ces évaluations dans leur sélection est importante, plus grande sera la différence entre les bons élèves et les autres. Ces derniers ayant a priori une distribution de futurs employeurs potentiels moins étendue que les élèves aux performances scolaires relativement plus élevées. De plus, dans un système comportant des évaluations externes, les parents de classes socio-économiques favorisées peuvent avoir une plus grande influence sur la fonction d'effort de leur enfant, par une intensification de leurs incitations, notamment lorsqu'il doit passer des évaluations externes. Ces motivations parentales ne se limitent toutefois pas aux transitions entre cycles scolaires ou vers le marché du travail, mais concernent également l'évaluation du système scolaire dans son ensemble, de la qualité de la classe et de l'établissement, en passant par celle de l'enseignant. Elles sont logiquement plus importantes dans un système scolaire décentralisé, où l'information sur les résultats des évaluations externes permet aux parents de mieux situer l'établissement fréquenté par leur progéniture par rapport à l'ensemble du système. D'un autre côté, si les établissements et les enseignants pratiquent une discrimination positive (plus intensive autour des évaluations externes), ils peuvent faire en sorte de compenser ce manque d'attention et d'incitation dans le milieu familial, réduisant dès lors l'écart de résultats produit par les différences en termes de caractéristiques familiales des élèves.

Les tests empiriques réalisés par Wössmann (2005), à partir de plusieurs vagues des enquêtes internationales TIMSS et PISA, confirment l'existence d'effets liés aux évaluations externes qui augmentent en fonction du niveau de performance scolaire des élèves. Ces effets varient également en fonction du degré de décentralisation du système et en fonction du temps, dans le sens où leur hétérogénéité est plus forte en dernière année du secondaire inférieur. Le lien entre l'hétérogénéité de ces effets et les caractéristiques familiales reste toutefois peu clair, bien que l'existence d'évaluations externes semble plutôt diminuer les désavantages initiaux des élèves de parents immigrés ou moins éduqués.

En résumé, ces différents modèles principal-agent prévoient que l'existence d'évaluations externes produit des effets hétérogènes résultant essentiellement des incitations générées dans le système par les signaux envoyés à l'externe du système. En termes de distribution des résultats scolaires, la principale prédiction est que, plus les parents et le marché du travail donnent de l'importance à ces signaux, plus le potentiel d'accroissement des inégalités sera grand. Une politique de discrimination positive envers les élèves issus de milieux défavorisés pouvant toutefois freiner cette tendance, comme le suggèrent les résultats empiriques de Wössmann (2005).

EXTERNALISATION ET INTERACTION ENSEIGNANT-ÉLÈVE

Indépendamment de l'existence d'une volonté des établissements à mener une politique de discrimination positive, l'évaluation externe peut générer des effets qui permettent de réduire l'impact des caractéristiques socio-économiques et sociodémographiques sur la performance des élèves. Pour comprendre cela, il suffit de considérer les principales caractéristiques des évaluations externes par rapport aux évaluations internes. Dans le cas extrême d'évaluations totalement externalisées (e.g. PISA), l'examineur n'a aucune forme d'interaction avec l'élève. À l'inverse, dans les évaluations réalisées exclusivement en interne à la classe, l'existence de stéréotypes négatifs et les préjugés et attentes de l'enseignant peuvent potentiellement modifier la distribution des résultats scolaires.

Depuis les travaux de Rosenthal et Jacobson (1968) sur la question de l'influence de l'enseignant sur le comportement de l'élève et de ses acquis scolaires, de nombreuses recherches ont clairement démontré comment les préjugés et les attentes de l'enseignant agissent non seulement sur l'attitude de l'élève, mais également sur la mesure de sa performance et donc sur l'inégalité scolaire. Dee (2005) suggère deux types d'inefficience sur la mesure de la performance scolaire qui marquent l'interaction enseignant-élève. Dans le premier type, l'enseignant a un rôle passif. Si ses caractéristiques, notamment sociodémographiques, sont identiques à celles de l'élève, ce dernier sera davantage motivé à fournir l'effort et à lui envoyer des signaux positifs. À l'inverse, des caractéristiques fortement dissemblables auront pour effet de nuire à la performance scolaire de l'élève, qui plus est lorsqu'il possède une caractéristique sur laquelle pèse un stéréotype négatif. Steele et Aronson (1995) ont empiriquement démontré cette intuition. Leurs résultats prouvent l'existence de ce qu'il est devenu courant d'appeler en psychologie sociale la « menace du stéréotype » : le fait d'être la cible d'un stéréotype négatif génère une pression psychologique sur l'élève qui biaise ses performances scolaires. Dans l'autre forme d'inefficience, l'enseignant a par contre un rôle actif, par le biais des préjugés et attentes qu'il associe à chacune des caractéristiques socio-économiques et sociodémographiques de l'élève⁶.

De manière simplifiée, on peut supposer que l'interaction enseignant-élève est empreinte d'une discrimination à la fois « subjective », basée sur les croyances de l'enseignant, et « objective », basée sur les signaux produits par le système (e.g. parcours, résultats précédents dans la matière, résultats dans les autres matières). Si l'on se réfère à la littérature économique, la discrimination associée à la relation enseignant-élève peut être statistiquement traduite de deux manières. Selon la première (Arrow, 1973), on

6. Dee (2004), par exemple, a démontré empiriquement qu'un appariement aléatoire des élèves avec des enseignants ayant la même couleur de peau augmentait le score moyen à la fois des élèves noirs et blancs.

considère que l'enseignant part du principe que la distribution de la qualité éducationnelle des élèves est fonction de leurs caractéristiques socio-économiques et sociodémographiques : par exemple, la distribution de la qualité éducationnelle des élèves allogènes est, à parité d'autres choses, caractérisée par une moyenne inférieure à celle des élèves indigènes. Autrement dit, la probabilité qu'un élève allogène tiré au hasard fasse partie des « bons » élèves est moindre que celle associée à un élève indigène. Dans la deuxième (Phelps, 1972), la qualité éducationnelle ne varie pas en fonction des caractéristiques socio-économiques et sociodémographiques. En revanche, les signaux sur la qualité éducationnelle, par exemple, des élèves allogènes sont supposés moins précis que ceux des élèves indigènes⁷. Ce qui se traduit par une différence dans la variance du signal, et donc une dispersion plus forte pour les élèves allogènes : le signal de qualité envoyé par un élève allogène est dès lors considéré par les enseignants comme moins fiable que celui provenant d'un élève indigène.

Les inefficiences générées par la discrimination statistique associée à l'interaction enseignant-élève peuvent bien entendu perdurer. À aptitudes innées identiques, les élèves ayant des caractéristiques discriminées et/ou victimes de stéréotypes négatifs (que ce soit au niveau sociétal ou uniquement à l'interne du système scolaire) sont moins incités à fournir un effort, ce qui a pour conséquence de confirmer, voire d'aggraver, les différences générées par la discrimination statistique ; le coût d'amélioration du signal envoyé à l'enseignant est en effet le même que pour les élèves non discriminés, mais pas le bénéfice attendu. D'une certaine manière, les prophéties des enseignants (fondées ou non) deviennent « autoréalisatrices », conduisant dès lors au renforcement des stéréotypes initiaux, comme l'ont notamment démontré Steele et Aronson (1995).

Une fois encore, une politique de discrimination positive et une amélioration de l'information sur la qualité éducationnelle des enfants (notamment ceux intégrant le système en cours de scolarité) peuvent diminuer les effets négatifs générés par la discrimination statistique en milieu scolaire. Mais les évaluations externes en soi peuvent également freiner ces effets pervers⁸. En effet, dans le cas d'une totale externalisation de l'évaluation, le bénéfice escompté de l'effort est, à qualité éducationnelle donnée, indépendant des caractéristiques socio-économiques et sociodémographiques des élèves.

7. Une hypothèse peu restrictive si l'on considère les enfants immigrés de première génération, qui plus est lorsqu'ils intègrent le système scolaire en cours de scolarité obligatoire, l'information accumulée par le système étant moindre que pour les enfants natis et les enfants immigrés de deuxième génération.

8. Ajoutons toutefois que dans le cas où la menace du stéréotype est plus forte pour les évaluations externes que pour les internes (par une communication vers l'externe du système qui stigmatise certains groupes sociodémographiques), les évaluations externes auront a priori plutôt tendance à accroître les inégalités de performance scolaire.

LES ÉVALUATIONS COMMUNES DE MATHÉMATIQUES AU SECONDAIRE INFÉRIEUR DU CANTON DE GENÈVE

Dans la présente section, nous discutons les principales hypothèses soulevées précédemment à partir des données micro de la base de données scolaires (BDS)⁹ et de la base de données des évaluations communes (EVACOM) du secondaire inférieur public dans le canton de Genève pour l'année 2005-2006. Cette division scolaire concerne trois degrés (7^e-9^e), et contrairement à la division primaire jusqu'à récemment¹⁰, ses établissements jouissent depuis les années 1960 d'un fonctionnement décentralisé. Depuis le début des années 2000, l'organisation repose sur une sélection/orientation des élèves en deux filières, le regroupement A à exigences scolaires étendues, et le regroupement B (ou C) à exigences scolaires moindres¹¹. L'effectif des classes dans le regroupement B est réduit et le choix d'options est moins large que dans le regroupement A.

En termes de transition vers le secondaire supérieur, et contrairement à la tendance observée dans le reste de la Suisse (et plus particulièrement dans les cantons alémaniques), le canton de Genève se caractérise par un faible taux de transitions vers les formations professionnelles duales (un élève sur vingt environ aujourd'hui contre près d'un sur six en 1990) et une forte proportion d'élèves qui s'oriente vers une formation gymnasiale (près d'un sur deux aujourd'hui contre 15% en 1990), donnant directement accès aux études universitaires après obtention du titre de maturité gymnasiale fédérale. Les autres transitions sont la formation certificative (École de culture générale et

9. La base de données scolaires (BDS) du canton de Genève existe depuis la fin des années 1960 et recense toutes les personnes scolarisées dans le canton depuis la petite enfance jusqu'au niveau tertiaire (excluant toutefois les étudiants de l'Université de Genève). Outre les informations purement scolaires (degré, établissement, etc.) elle contient des informations sur les caractéristiques de la personne (genre, âge civil, âge théorique, nationalité) et de sa famille (adresse, nationalité-s, langue-s parlée-s à la maison, catégories socioprofessionnelles [CSP]). Un seul et unique numéro d'identification de la personne dans la BDS permet d'ajouter, par exemple, les informations contenues dans la base EVACOM (résultats scolaires, numéro d'identification classe, numéro d'identification enseignant).

10. Les directions d'établissement se sont mises en place durant l'année scolaire 2008-2009 (voir la contribution de Solaux, Huguenin, Payet & Ramirez dans le présent ouvrage).

11. La transition entre la dernière année du primaire (6^e) et la 7^e se fait essentiellement sur la base des résultats en mathématiques et en français en fin de 6^e. Dans trois établissements (sur les 19 existants en 2005-2006), il n'y a pas de distinction de filière au 7^e degré. Dès le 8^e degré, les élèves sont toutefois regroupés selon trois niveaux d'exigences dans le cadre des enseignements liés aux principales matières prises en considération lors de la transition vers le postobligatoire, à savoir les mathématiques et l'allemand. Une table de correspondance permet aux institutions scolaires et professionnelles du secondaire supérieur d'identifier le regroupement A ou B dans lequel se trouverait un élève sortant d'un établissement pratiquant l'hétérogénéité dans la formation des classes. À des fins de simplification du propos, nous n'avons pas considéré les résultats des élèves fréquentant les établissements à classes hétérogènes, ni les élèves en regroupement C qui concerne une petite minorité.

écoles professionnelles à plein temps¹²) et les structures d'accueil et d'insertion qui représentent 15 % des transitions aujourd'hui contre 5 % en 2000. La transition vers le secondaire supérieur varie grandement selon le regroupement en secondaire inférieur, puisque plus des deux tiers des élèves sortant de A se dirigent vers la formation gymnasiale, pendant que 54% des élèves de B s'orientent vers une formation certifiante et le reste vers les structures d'accueil et d'insertion (SRED, 2009).

Comme exposé auparavant, les élèves du secondaire inférieur à Genève passent une évaluation cantonale (EC) en mathématiques dans les trois degrés¹³. Ces EC sont externalisées dans leur élaboration et totalement standardisées dans leur passation et leur correction, quand bien même ces deux tâches sont effectuées par l'enseignant de la matière. Leur poids est avant tout symbolique puisque leur résultat ne compte que pour 20 à 25 % de la moyenne du dernier trimestre. Ce n'est toutefois pas tant le score à l'EC en soi qui nous intéresse ici, mais la différence entre celui-ci et la moyenne annuelle afin de discuter les diverses hypothèses tirées des deux sections précédentes. L'échantillon considéré compte 7604 élèves en regroupement A, dont 2282 au dernier degré, et 2445 élèves en regroupement B, dont 963 au dernier degré. La proportion de filles dans le regroupement A (51,6 %) est légèrement supérieure à celle des garçons et la part des élèves de nationalité suisse est de 71,1 %. Dans le regroupement B, les proportions de filles et d'élèves de nationalité suisse sont respectivement de 42,8 % et 48 %. La structure en termes de catégories socioprofessionnelles¹⁴ diffère également selon le regroupement puisque la proportion d'enfants provenant de milieux familiaux plutôt défavorisés est de 52,2 % en B et 30,1 % en A.

Les graphiques 1 et 2 représentent pour différents percentiles de la distribution des moyennes annuelles¹⁵, l'écart (moyen du percentile) entre le résultat à l'EC et la moyenne annuelle¹⁶. Dans le graphique 1, les valeurs sur l'ordonnée correspondent à l'écart pour chaque percentile de la distribution

12. Les écoles secondaires professionnelles à plein temps délivrent des maturités professionnelles qui donnent accès aux nouvelles structures de formation professionnelle de niveau tertiaire, les Hautes écoles spécialisées.

13. En plus des maths, les EC concernent également le français et l'allemand aux trois degrés. Pour l'EC en maths, les élèves du regroupement B passent la même que ceux de A mais le taux de réussite est fixé à 50 % pour les B (voire moins selon les résultats finalement obtenus pour ce regroupement [Soussi *et al.*, 2009]) et à 60 % pour les A.

14. Les cinq catégories socioprofessionnelles (CSP) considérées sont: 1) Cadres supérieurs et dirigeants; 2) Employés, cadres inférieurs et intermédiaires; 3) Petits indépendants; 4) Ouvriers et 5) Divers autres. Cette dernière catégorie inclut les parents chômeurs, rentiers de l'assurance invalidité ou vieillesse et, au même titre que la CSP ouvriers, est considérée par l'administration comme une population socio-économiquement défavorisée.

15. Les points ont été reliés pour souligner les tendances; ces « courbes » ne correspondent pas strictement à la forme exacte de la fonction de distribution des moyennes annuelles.

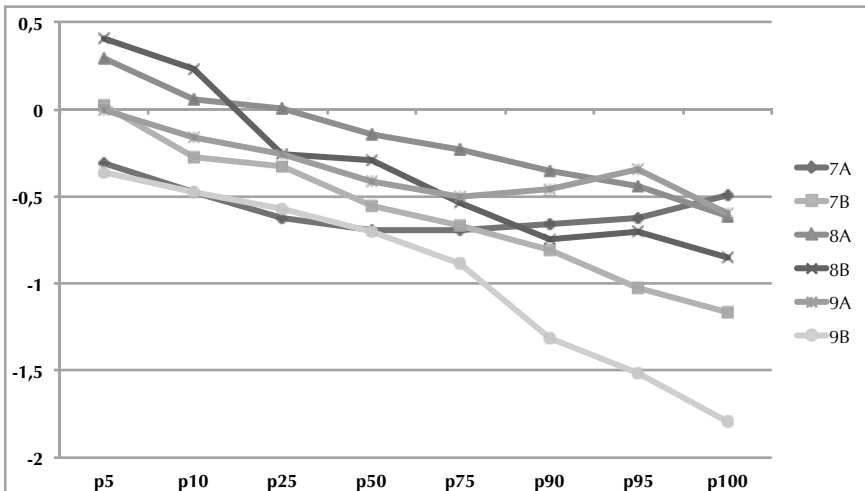
16. La note maximale dans le système scolaire genevois est de 6, le minimum étant généralement fixé à 1, pour la présence.

des moyennes annuelles observées sur l'ensemble des élèves d'un degré et d'un regroupement, alors que dans le graphique 2 c'est à partir de la distribution des moyennes annuelles observées à l'intérieur des classes que les valeurs en ordonnée sont calculées. Autrement dit, dans le premier on compare le score relatif de l'élève à l'EC par rapport à son degré et son regroupement, et par rapport à sa classe dans le deuxième.

Comme on peut aisément le voir à partir du graphique 1 notamment, les élèves relativement moins performants sur l'année s'en sortent relativement mieux à l'EC. Ceci est plus particulièrement visible pour les élèves du regroupement B, où les élèves du 9^e degré se situant dans les 5% les moins performants en moyenne annuelle (p0-p5) ont un écart moyen entre l'EC et leur moyenne annuelle de -0,36 contre -1,80 pour les élèves se situant parmi les 5% les plus performants en moyenne annuelle (p95-p100)¹⁷. Si pour ces mêmes élèves, on prend comme référence de comparaison la classe (cf. graphique 2), ces écarts se réduisent à -0,46 et -1,08. Que ce soit au niveau de l'ensemble des élèves de mêmes degré et regroupement ou à l'intérieur de chaque classe, c'est dans le regroupement B que l'on observe les écarts les plus importants.

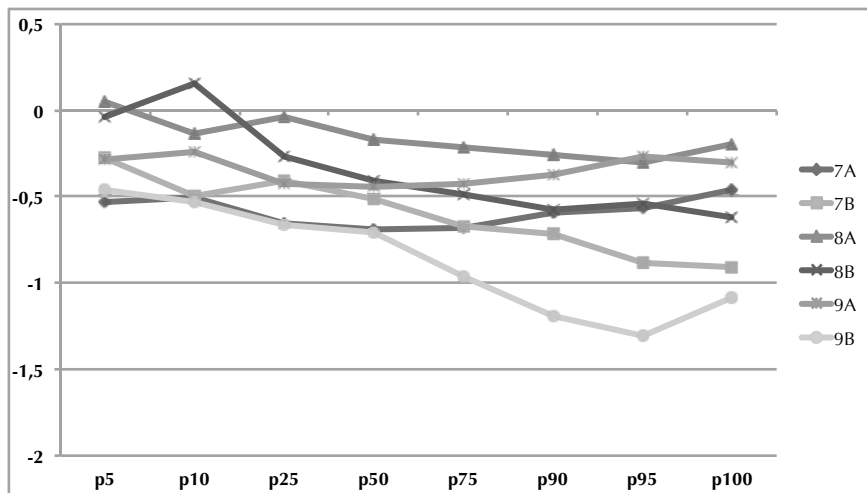
**Graphiques 1 et 2: Écart entre le résultat à l'EC de maths
et la moyenne annuelle (de maths)
pour différents percentiles de la moyenne annuelle (de maths)**

1. Percentiles déterminés sur l'ensemble des élèves de mêmes degré et regroupement



17. Il est important de souligner que les moyennes annuelles n'ont pas été « nettoyées » du résultat obtenu à l'EC, ce qui introduit un léger biais dans nos calculs mais qui n'altère en rien les tendances observées. Il se pourrait même, a priori, que la correction de ce biais ne fasse qu'accroître ces tendances.

2. Percentiles déterminés à l'intérieur des classes



Sources: BDS et EVACOM 2005-2006, Genève; propres estimations.

Sur la base de cette simple illustration statistique, il semble que l'évaluation (totalement) standardisée et (partiellement) externalisée réduise les inégalités en termes de moyenne annuelle de mathématiques dans le secondaire inférieur du canton de Genève. Les explications de cette tendance peuvent toutefois être multiples, comme discuté auparavant, et donc impliquer des politiques et mesures différenciées dans le cas où l'on voudrait utiliser ces évaluations comme outil pour accroître l'efficacité sociale. Si l'on raisonne strictement en termes d'incitations de l'élève par rapport à l'EC (i.e. comme dans les modèles principal-agent), cet effet hétérogène négatif (dans le sens où il freine les inégalités de résultats) est peut-être simplement dû au faible poids accordé à l'EC dans la moyenne annuelle. Un faible poids réel qui diminue forcément le rendement de l'effort. Toutefois, le fait que la tendance soit plus marquée dans le regroupement B, et encore plus nettement au dernier degré (9B), peut au contraire suggérer que le « vrai » poids de l'EC n'est pas dans la moyenne, mais dans sa valeur symbolique. En effet, pour les élèves du regroupement B (dont la probabilité de transiter en fin de 9^e année directement vers le marché du travail est sensiblement plus élevée que pour ceux du regroupement A), un relativement bon résultat à l'EC leur permet d'améliorer le signal envoyé aux employeurs potentiels, qui plus est si l'élève est conscient que les résultats qu'il a obtenus sur la base des évaluations internes ne reflètent pas sa « vraie » qualité éducative.

Empiriquement, la différence entre le résultat à l'EC et la moyenne annuelle de l'élève est une variable dépendante fort pratique, car elle permet

d'éliminer les effets inobservables exclusivement liés à la personne¹⁸ lorsque l'on exploite des données individuelles (Rivkin, Hanushek & Kain, 2005). On peut donc plus aisément discuter les effets du contexte et de l'interaction enseignant-élève sur la variable dépendante. Des deux graphiques présentés, nous pourrions déduire que les différences entre élèves d'un même degré et regroupement sont en bonne partie dues aux différences entre classes. En termes d'identification toutefois, il n'est pas possible à ce stade de savoir si ces différences entre classes sont dues à la composition des classes, à l'enseignant ou encore à des différences entre établissements. Pour ce faire, nous proposons un premier pas, une simple question : quelles sont les variables qui comptent pour expliquer les différences observées entre élèves de l'écart entre leur résultat à l'EC et leur moyenne annuelle ? Le tableau 1 présente deux simples modèles d'analyse de la variance (ANOVA) différenciés selon les regroupements A et B. L'ensemble des variables indépendantes sont qualitatives, représentant un numéro d'identification (établissement, classe, enseignant), une catégorie socioprofessionnelle, voire simplement binaires pour le genre et la nationalité suisse.

Tableau 1 : Modèles ANOVA sur les écarts individuels entre le résultat à l'EC de maths et la moyenne annuelle (de maths) selon regroupement

Facteur	Regroupement A		9A	Regroupement B		9B
Établissement	0,0128**	0,0029**	0,0102**	0,0077*	0,0068*	0,0127
Classe	0,3228**	0,0885**	0,0490**	0,2405**	0,1011**	0,0508*
Enseignant	0,0012	0,0016	0,0022	0,0492**	0,0335**	0,0905**
Genre	0,0019**	0,0012**	0,0064**	0,0024**	0,0009	0,0026*
Suisse	0,0018**	0,0008**	0,0004	0,0010*	0,0001	0,0001
CSP	0,0053**	0,0026**	0,0046**	0,0005	0,0004	0,0012
Enseignant*sexe	–	0,0216*	0,0275	–	0,0511**	0,0747*
Enseignant*suisse	–	0,0232**	0,0253	–	0,0472	0,0638**
Enseignant*CSP	–	0,0725**	0,0868	–	0,1243**	0,1534**
R2 ajusté	0,3780	0,3977	0,3530	0,3637	0,4025	0,4162
N	7604	7604	2282	2445	2445	963

*prob. $F \leq .05$; **prob. $F \leq .01$. Les coefficients correspondent simplement à la part de la variance expliquée par chaque facteur ou interaction de facteurs (les autres facteurs étant constants). Les nombres d'établissements, de classes et d'enseignants sont respectivement de 16, 368 et 235 dans le regroupement A, et de 16, 208 et 163 dans le regroupement B. Sources : BDS et EVACOM 2005-2006, Genève; propres estimations.

18. Caractéristiques inobservables qui permettent d'analyser dans ce cas le comportement de l'élève, son attitude générale envers l'école, etc., au-delà de ses caractéristiques visibles telles que genre, ethnie/nationalité, catégorie socioprofessionnelle/formation des parents, etc.

Considérons pour l'instant uniquement les deuxième et cinquième colonnes de ce tableau qui présentent les résultats du premier modèle. Tant l'établissement que la classe influent sur la variance observée à partir des écarts de performance individuelle. L'enseignant ne semble toutefois pas influencer sur les écarts observés dans le regroupement A, contrairement à ce que nous pouvons voir dans la cinquième colonne correspondant au regroupement B. À l'inverse, les caractéristiques individuelles semblent davantage influencer les écarts observés dans le regroupement A, ce qui n'est pas surprenant étant donné la plus grande taille et la plus forte hétérogénéité caractérisant ce dernier regroupement.

Les troisième et quatrième colonnes présentent les résultats du deuxième modèle pour le regroupement A, alors que les sixième et septième montrent ceux obtenus pour le regroupement B. Comme on peut le voir, l'interaction des variables socio-économiques et sociodémographiques avec la variable enseignant permet dans tous les cas d'améliorer le pouvoir explicatif, notamment pour le 9^e degré du regroupement B. Le poids de l'interaction enseignant-catégorie socioprofessionnelle semble clairement supérieur à ceux des interactions enseignant-sexe et enseignant-nationalité suisse, notamment dans le regroupement B. Si l'on se concentre exclusivement sur les élèves du dernier degré, la différence entre les deux regroupements apparaît nettement. Dans le B, où la probabilité de transition vers le marché du travail est élevée, et l'hétérogénéité de la population d'élèves moins forte que dans le regroupement A, les effets associés à l'interaction enseignant-élève semblent être également plus importants.

CONCLUSION

Les évaluations externes (ou partiellement externes) peuvent-elles renforcer l'équité sociale, du fait qu'elles sont moins affectées par les inefficiences sociales liées à l'interaction enseignant-élève? La réponse est définitivement empirique. Les simples observations que nous avons présentées à partir de la variance mesurée sur des écarts individuels entre le score à une évaluation totalement standardisée et la moyenne annuelle en mathématiques, soulignent l'importance de l'interaction enseignant-élève sur les inégalités observées en termes de performance scolaire, au même titre que les effets de contexte. Dans la filière à exigences scolaires moindres du secondaire inférieur dans le canton de Genève, caractérisée par une forte probabilité de transition finale vers le marché du travail et une moindre hétérogénéité socio-économique, les effets induits par l'interaction enseignant-élève sont plus marqués et sensiblement plus importants que les effets de contexte.

Cet article a tenté d'apporter un éclairage sur les avantages et désavantages des évaluations centralisées, avant tout en termes d'équité sociale,

les résultats des études empiriques réalisées sur le sujet – que ce soit sur la base de données locales, nationales ou internationales – ne permettant pas à ce jour de dégager un consensus sur les bienfaits de ce type d'évaluation (Mons, 2009 ; OCDE, 2007). Dans notre discussion théorique, nous avons abordé la question délicate de la discrimination dont l'interaction enseignant-élève peut être empreinte. De manière simplifiée, cette discrimination est le produit de stéréotypes négatifs, basés sur les caractéristiques sociodémographiques et socio-économiques, et sur les préjugés et attentes des enseignants quant à la qualité éducative de leurs élèves. Dans le cadre des évaluations externes et standardisées, les inefficiences sociales liées à l'interaction enseignant-élève peuvent être réduites, surtout lorsque la correction est totalement externalisée.

En termes méthodologiques, cet article suggère de mieux exploiter les différents degrés d'externalisation et de standardisation des évaluations pour mieux comprendre les inégalités de performance scolaire. En termes politiques, le message est très clair : tout ce qui peut améliorer la qualité de la mesure de la qualité éducative des élèves ne peut qu'augmenter la qualité des systèmes scolaires. L'interaction enseignant-élève est génératrice d'effets puissants tant sur les acquis des enfants que sur leur évaluation, comme l'a encore démontré une étude fort complète basée sur des données administratives aux États-Unis (Rivkin *et al.*, 2005). Les recherches empiriques sur le rôle et l'importance de la qualité (observable et non observable) de l'enseignant constituent sans aucun doute un champ de recherche prometteur en économie de l'éducation.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Arrow, K. (1973). The Theory of Discrimination. In O. Aschenfelder & A. Rees (Ed.), *Discrimination in the Labor Market* (pp. 3-33). Princeton, NJ : Princeton University Press.
- Bishop, J. H. & Wössmann, L. (2004). Institutional effects in a simple model of educational production. *Education Economics*, 12, 17-38.
- Bourdieu, P. & Passeron, J.-C. (1970). *La reproduction*. Paris : Minuit.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC : U.S. Government Printing Office.
- Conférence suisse des directeurs de l'Instruction publique. (2009). *HarmoS*. Berne. Consulté le 8 avril 2011 dans <http://www.edk.ch/dyn/11737.php>
- Costrell, R. M. (1994). A Simple Model of Educational Standards. *American Economic Review*, 84, 956-971.
- Dee, T. S. (2004). Teachers, Race and Student Achievement in a Randomized Experiment. *Review of Economics and Statistics*, 86(1), 195-210.
- Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review*, 95(2), 158-165.

- Institut de recherche et de documentation pédagogique. (2010). *Pratiques cantonales concernant l'organisation d'épreuves, d'examens, de tests ainsi que l'obtention de certificats ou diplômes. Enseignement primaire, secondaire premier cycle. Suisse romande*. Neuchâtel.
- Mons, N. (2009). *Les effets théoriques et réels de l'évaluation standardisée. Complément à l'étude: Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats*. Bruxelles: EACEA & Eurydice.
- Organisation de coopération et de développement économiques. (2007). *PISA 2006: Science Competencies for Tomorrow's World* (Vol. 1). Paris.
- Organisation de coopération et de développement économiques. (2011). *Raising Education Outcomes in Switzerland* (OECD Economics Department Working Papers n° 838). Paris.
- Phelps, E. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, 62(4), 569-661.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417-458.
- Rosenthal, A. R. & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Service de la recherche en éducation. (2009). *L'enseignement à Genève. Ensemble d'indicateurs du système genevois d'enseignement et de formation*. Genève.
- Soussi, A., Guilley, E., Guignard, N. & Nidegger, C. (2009). *Évaluation des acquis des élèves à l'école obligatoire. L'évaluation cantonale et commune: état de situation du dispositif existant, points forts et points faibles, perspectives d'avenir*. Genève: SRED.
- Steele, C. M. & Aronson, J. (1995). Stereotype Threat and The Intellectual Test-Performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.
- Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64, 416-424.
- Wirthner, M. & Ntamakiliro, L. (2008). Des épreuves de référence au service de l'évaluation des enseignants. In J. Weiss (Éd.), *Quelle évaluation des enseignants au service de l'école? Actes du séminaire AIDEP 2007, Leysin 6-7 décembre* (pp. 91-99). Neuchâtel: IRDP.
- Wössmann, L. (2005). The Effect of Heterogeneity of Central Examinations: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, 13(2), 143-169.

Quelle est l'importance des compétences évaluées dans les avantages liés à la formation sur le marché du travail?

Jean-Marc Falter
Université de Genève,
Leading House en économie de l'éducation

INTRODUCTION

À l'instar d'autres disciplines s'intéressant à la formation, l'évaluation a pris une place centrale en économie de l'éducation. Les débats sur l'efficacité ou l'efficience économique des politiques éducatives se basent sur les scores obtenus par les élèves lors de tests standardisés portant sur des matières comme la lecture, les mathématiques ou les sciences. Il va de soi que les objectifs d'un système scolaire ne se limitent pas à l'obtention d'un bon score dans ces matières. Des travaux récents (Heckman & Rubinstein, 2001 ; Heckman, Stixrud & Urzua, 2006) ont mis en évidence l'importance de compétences dites non cognitives comme le comportement ou l'attitude dans le succès sur le marché du travail. L'évolution de la demande de compétences sur le marché du travail, caractérisée par une polarisation croissante de l'emploi entre postes de travail nécessitant des compétences accrues et postes à très faibles qualifications (voir Autor & Dorn, 2009 ; Oesch & Rodriguez Menés, sous presse) est également susceptible de remettre en cause le modèle d'évaluation basé sur des compétences cognitives de base. Néanmoins, à un niveau macro-économique, les modes actuels d'évaluation trouvent une certaine légitimité. Plusieurs résultats indiquent que la croissance économique est positivement corrélée aux variations de

score moyen en compétences de base comme le calcul ou l'écriture (voir par exemple Hanushek & Wössmann, 2007).

Il ne fait guère de doute que de meilleurs scores en calcul, écriture ou lecture sont associés à des effets positifs, ceci aussi bien pour les personnes que pour les sociétés dans lesquelles elles vivent. Ce qui ne signifie pas pour autant qu'une discussion sur leur pertinence soit malvenue. En effet, la performance scolaire englobe plusieurs dimensions, au premier rang desquelles figure l'insertion professionnelle. À ce titre, le système scolaire finlandais, qui est souvent présenté comme « un champion » de la performance scolaire si l'on se base sur les résultats des diverses études PISA de l'OCDE, semble faire pâle figure lorsque l'on se penche sur le taux de chômage des jeunes. En Finlande, celui-ci s'élevait à 14,8% en 2008, contre une moyenne de 12,4% dans les pays de l'OCDE. On ne peut certes pas rejeter toute la responsabilité des déséquilibres du marché du travail sur le système scolaire, mais cet exemple souligne que l'évaluation scolaire doit reposer sur plusieurs instruments de mesure. Ceci paraît d'autant plus évident si l'on appréhende la diversité des compétences nécessaires sur le marché du travail. Dans cette optique, l'accent mis sur des matières comme les mathématiques ou la lecture est très certainement lié à un a priori favorable par rapport aux professions intellectuelles, qui connaissent une très forte expansion dans les pays développés. Cette approche est légitime, mais elle ne saurait négliger le fait qu'une majorité de jeunes ne se destinent pas à une formation tertiaire. Bertschy, Cattaneo et Wolter (2009) ont montré, dans le cas de la Suisse, que les scores PISA avaient un impact positif sur la probabilité de réussir un diplôme d'apprentissage de niveau secondaire, mais que cet effet était relativement modeste. Les compétences de base mesurées dans des tests comme ceux des études PISA sont probablement indispensables dans le monde actuel. Cependant, leur importance pour la réussite professionnelle dépend vraisemblablement du type de formation suivie et de la profession choisie.

Cette contribution propose une synthèse de plusieurs travaux empiriques qui mesurent l'effet des compétences de base sur les parcours éducatifs et la réussite sur le marché du travail en Suisse. D'abord, nous vérifions si les compétences mesurées dans les tests PISA déterminent les transitions au sein du système éducatif suisse. Comme le définit l'OCDE (2010) « PISA teste l'aptitude des élèves à appliquer les connaissances acquises à l'école aux situations de la vie réelle, plutôt que leur maîtrise d'un programme scolaire précis. » Ceci signifie que les scores PISA ne devraient qu'imparfaitement expliquer les transitions au sein d'un système éducatif. Dans quelle mesure ces tests sont un bon indicateur de la réussite scolaire est donc une question d'importance. Grâce aux données de TREE (Transition de l'école

à l'emploi)¹, nous pouvons suivre la cohorte ayant participé à l'enquête PISA 2000 jusqu'à l'entrée sur le marché du travail.

Ensuite, nous discutons de l'effet conjoint des compétences de base et de la formation sur le marché du travail. Les avantages liés à la formation sont-ils dus à ce type de compétences ou, alternativement, à des compétences plus spécifiques comme celles acquises suite à un apprentissage? Pour répondre à cette question, nous commenterons les résultats obtenus à partir de l'enquête ALL pour la Suisse (*Adult Literacy and Lifeskills Survey*). ALL est une enquête internationale initiée par l'OCDE qui mesure les compétences de base (calcul, écriture et lecture) de la population adulte. Ces aptitudes sont relativement similaires à celles mesurées dans les tests PISA, étant donné qu'elles se distinguent de connaissances spécifiques apprises lors de la formation ou sur le lieu de travail. Notre approche consiste à comparer les effets des compétences de base sur le marché du travail en distinguant la formation générale de la formation professionnelle.

COMPÉTENCES ÉVALUÉES ET TRANSITIONS SCOLAIRES

Les résultats des enquêtes PISA ont eu un fort retentissement au sein du grand public et des médias. Des jugements, souvent hâtifs, ont été émis sur les performances relatives des différents systèmes scolaires suisses. Quelle que soit la valeur que l'on attribue à ces controverses, il ne fait aucun doute que ces enquêtes ont un impact sur la politique éducative. Ceci est d'autant plus vrai qu'à l'heure actuelle, il n'existe pas d'autres instruments en Suisse permettant de comparer les performances scolaires entre les cantons. À notre sens, il convient dans un premier temps d'analyser comment ces mesures d'aptitudes des élèves, établies à l'âge de 15 ans, affectent la réussite scolaire et les transitions au sein du système éducatif. À cette fin, nous utilisons les données de TREE. Il s'agit d'un suivi longitudinal des élèves ayant participé à l'enquête PISA 2000 entre 2001 et 2007, c'est-à-dire jusqu'à l'entrée sur le marché du travail.

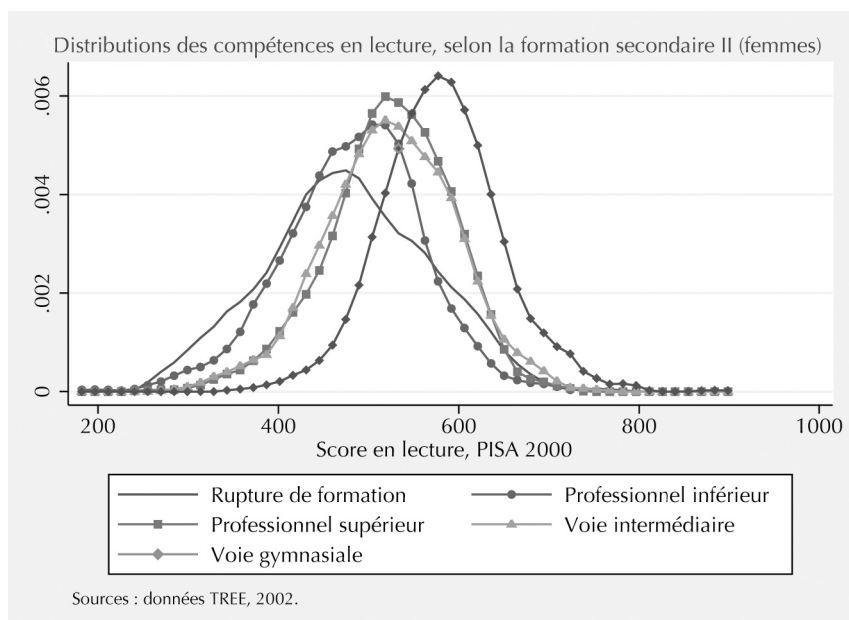
La première transition qui a fait l'objet de nos analyses est celle qui conduit aux études secondaires supérieures, au terme de l'école obligatoire. Il s'agit d'une transition cruciale, particulièrement dans le cas de la Suisse, car les jeunes effectuent un choix entre une formation générale et une formation professionnelle. Suite à des regroupements, nous avons distingué cinq types de formation : éducation de niveau inférieur, formation professionnelle à basses qualifications, formation professionnelle à hautes qualifications, formations intermédiaires et formation gymnasiale. La première catégorie regroupe les jeunes en rupture momentanée ou définitive de formation, ceux qui suivent une formation n'aboutissant pas à une certification ainsi

1. Voir <http://tree.unibas.ch/>

que d'autres formations offrant de faibles débouchés sur le marché du travail. La formation professionnelle a été subdivisée en deux catégories selon l'exigence des professions à l'aide d'une échelle développée par Stalder (2005), elle-même basée sur des informations qualitatives recueillies auprès de professionnels actifs dans le domaine de l'orientation scolaire. Les formations intermédiaires regroupent des formations non professionnelles ne permettant pas d'accéder directement à une formation universitaire. Enfin, la voie gymnasiale mène à un diplôme secondaire permettant l'accès aux hautes écoles.

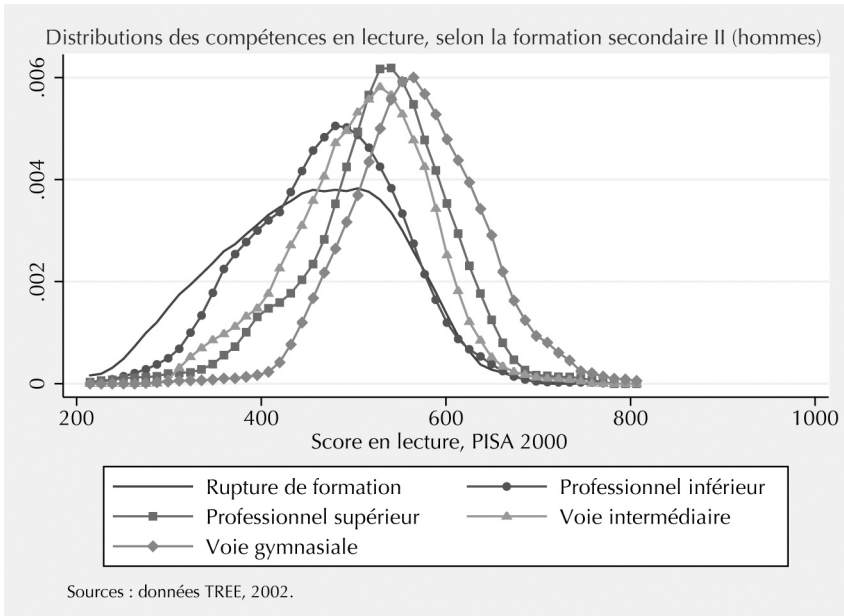
Dans les graphiques 1 et 2 sont reportées les distributions des scores en lecture pour les femmes et les hommes respectivement selon le type de formation suivie en 2002, soit deux ans après le test². Comme on pouvait s'y attendre, on constate des différences d'aptitude selon la formation choisie. Néanmoins, ces différences ne sont pas toujours très marquées, les courbes de distribution se chevauchant. Il en ressort notamment que la formation professionnelle de haut niveau attire des individus relativement similaires à ceux choisissant la formation gymnasiale. À l'inverse, la formation professionnelle de niveau inférieur est clairement distincte.

Graphique 1 : Distributions des compétences selon la formation secondaire II (femmes)



2. PISA 2000 s'est concentré sur la lecture; il s'agit donc de la seule compétence pour laquelle nous disposons d'informations pour l'ensemble de l'échantillon.

Graphique 2: Distributions des compétences selon la formation secondaire II (hommes)



Pourquoi des individus aux aptitudes similaires suivent-ils des voies différentes? Deux explications peuvent être avancées. La première a trait aux variations régionales, qui peuvent se manifester par des différences culturelles ou institutionnelles. Concernant ces dernières, on pense plus particulièrement à l'âge de sélection des élèves. En effet, il a été démontré que les cantons qui sélectionnent précocement ont de plus fortes inégalités dues à l'origine sociale des parents que ceux où la sélection s'opère tardivement (Bauer & Riphahn, 2006). On peut également évoquer la structure économique de la région de domicile qui va déterminer l'offre de places d'apprentissage. La deuxième explication porte sur l'origine sociale et migratoire des élèves. Les contraintes économiques peuvent forcer des personnes venant de milieux défavorisés à renoncer aux études supérieures. Ces effets peuvent se faire sentir à un âge relativement bas, si les individus anticipent des coûts de formation futur trop élevés pour les ressources du ménage. On peut également avancer des arguments de nature plus sociologique, la décision de poursuivre des études étant souvent conditionnée par des valeurs communes aux groupes sociaux. Finalement, les aspirations individuelles dépendent vraisemblablement de ce qui a été accompli par les parents.

Dans le cadre d'une analyse multivariée³, il ressort que les compétences influencent fortement les transitions vers la formation secondaire supérieure (voir tableau 1). Comme on pouvait s'y attendre, les scores PISA sont positivement corrélés avec une formation académique et négativement avec une formation professionnelle de niveau inférieur ou une rupture de formation. Concernant la formation professionnelle de niveau supérieur, les effets des scores PISA varient selon les sexes. Les scores influencent positivement ce choix pour les hommes et négativement pour les femmes. En parallèle, l'impact de l'origine sociale demeure élevé, même lorsque l'on prend en compte les scores PISA. Par exemple, avoir des parents qui ont achevé une formation supérieure (maturité ou plus) augmente la probabilité de poursuivre des études gymnasiales de 28 % et 27 % pour les femmes et les hommes respectivement, par rapport à une personne dont les parents n'ont pas accompli d'études au-delà de l'école obligatoire.

Tableau 1 : Effets marginaux, transitions vers le secondaire supérieur, TREE 2002

	Formation professionnelle faible exigence		Formation professionnelle haute exigence		Études gymnasiales	
	Femmes	Hommes	Femmes	Hommes	Femmes	Hommes
Score en lecture/100	-0,177*** (0,012)	-0,301*** (0,023)	-0,064*** (0,016)	0,066*** (0,020)	0,335*** (0,019)	0,267*** (0,018)
<i>Éducation des parents :</i>						
Niveau intermédiaire	-0,043** (0,018)	-0,080*** (0,032)	0,006 (0,025)	-0,024 (0,033)	0,063* (0,033)	0,085** (0,037)
Niveau supérieur	-0,122*** (0,019)	-0,232*** (0,0325)	-0,154*** (0,025)	-0,035 (0,035)	0,277*** (0,036)	0,266*** (0,039)

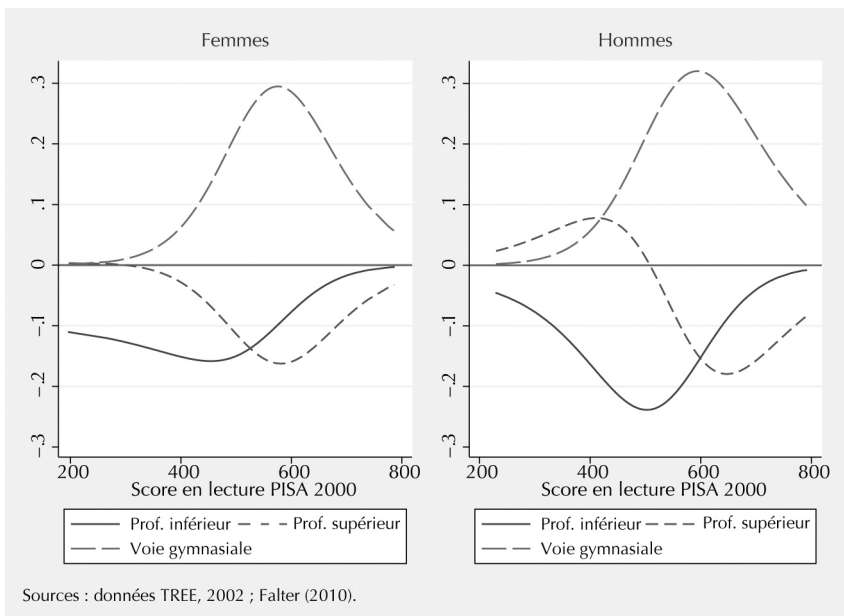
Notes: méthode d'estimation: logit multinomial; entre parenthèses: écarts types robustes. *p < .1; **p < .05; ***p < .001. Voies de formation non reportées: rupture de formation, formations intermédiaires (voir texte). Variables de contrôle: né(e) en Suisse, francophone, italophone, nombre de livres au sein du ménage, langue parlée à la maison différente de celle du test, taille de la commune, cantons à sélection précoce (11 ans), nombre de voies au secondaire I, taille de la fratrie, position au sein de la fratrie, âge (en mois). Nombres d'observations: 2500 (échantillon féminin) et 1941 (échantillon masculin). Sources: TREE, 2002; Falter (2010).

Les effets marginaux commentés ci-dessus sont estimés à la moyenne de l'échantillon. Ces effets sont donc susceptibles de varier selon les caractéristiques des individus. Pour cette raison, nous avons analysé les interactions entre les scores en lecture et l'origine sociale des élèves, mesurée

3. Les résultats détaillés sont décrits dans Falter (2010).

grossièrement par le niveau de formation des parents. Si les effets découlant de l'originale sociale étaient indépendants du niveau de compétence des élèves, cela mettrait en évidence une allocation inefficace des élèves. Nous avons estimé les effets marginaux associés à des parents ayant achevé une formation supérieure pour différents niveaux de compétences. Les autres variables du modèle sont constantes et fixées à la moyenne de l'échantillon. L'évolution des effets parentaux selon le score PISA est reportée dans le graphique 3 pour les femmes et les hommes. Plusieurs constatations s'imposent. Tout d'abord, les effets marginaux dépendent du niveau des compétences évaluées aussi bien pour les hommes que pour les femmes. Dans le cas de la transition vers la formation professionnelle supérieure, on observe même que le signe des effets marginaux s'inverse, passant de négatif à positif, pour l'échantillon masculin. En deuxième lieu, aussi bien pour la formation gymnasiale que pour la formation professionnelle supérieure, les effets marginaux sont à leur maximum pour un niveau de compétence approximativement égal à 600 (soit environ un écart type au-dessus de la moyenne de l'échantillon). Les signes opposés de ces effets marginaux suggèrent que ces deux formations sont des substituts à ce niveau de compétence. Enfin, les variables parentales n'ont guère d'effet sur le choix des études aux extrêmes de la distribution des compétences. Pour ces personnes, la sélection se fait essentiellement sur la base des compétences et non pas des variables parentales. En d'autres termes, on peut affirmer qu'aux extrêmes de la distribution, les scores PISA

Graphique 3 : Effets marginaux liés au critère « au moins un parent avec une formation supérieure »



Sources : données TREE, 2002 ; Falter (2010).

sont un bon indicateur des futurs diplômés, et par là même de la position future sur le marché du travail. Si ceci peut paraître trivial, il n'en est rien. Checchi & Flabbi (2007), appliquent une méthodologie similaire à l'Italie et à l'Allemagne, pays où les élèves sont également sélectionnés précocement dans des filières différenciées. Les résultats pour l'Italie montrent que les variables parentales produisent leur effet indépendamment du score PISA de l'élève alors que l'Allemagne présente un profil similaire à celui de la Suisse.

Il est également intéressant de savoir si les compétences mesurées dans PISA continuent d'avoir un impact pour les transitions vers les formations tertiaires. De ce fait, nous avons estimé un modèle de choix séquentiels. Nous faisons l'hypothèse que les individus sont confrontés à une série de trois choix. Le premier porte sur le type de formation, professionnelle ou générale. Le deuxième concerne le degré d'exigence de la formation choisie et, finalement, le troisième choix survient à la fin des études secondaires avec la possibilité de poursuivre des études tertiaires⁴. Nos résultats montrent que les variables de compétences perdent de leur importance à mesure que l'on progresse dans le cursus des études, particulièrement pour les personnes qui poursuivent la voie gymnasiale (tableaux 2 et 3). Il est à noter que ce phénomène est observé pour d'autres variables de notre modèle comme l'origine sociale. Nous constatons néanmoins que le score PISA conserve un impact positif sur la probabilité de poursuivre une formation tertiaire pour

Tableau 2 : Odds ratios, transitions vers la formation tertiaire, voie professionnelle

	Professionnel ou général	Haute ou faible exigence	Éducation secon- daire ou tertiaire
Score en lecture/100	0,986*** (0,000)	1,008*** (0,000)	1,006** (0,027)
<i>Éducation des parents :</i>			
Niveau intermédiaire	0,787 (0,159)	1,300 (0,203)	0,903 (0,797)
Niveau supérieur	0,275*** (0,000)	1,837** (0,043)	2,074 (0,103)
Observations		2741	

Notes: méthode d'estimation: logit séquentiel; entre parenthèses: valeur « p ». *p < .1; **p < .05; ***p < .001. Variables de contrôle: né(e) en Suisse, francophone, italophone, nombre de livres au sein du ménage, langue parlée à la maison différente de celle du test, cantons à sélection précoce (11 ans), âge (en mois), hommes.
Sources: TREE, 2002; Falter (2010).

4. Le premier et le deuxième choix peuvent être inversés. Qualitativement, les résultats sont inchangés.

les personnes ayant accompli une formation professionnelle supérieure. Cela pourrait signifier que le développement d'une formation professionnelle de type tertiaire permettrait aux individus à fortes capacités d'accéder à une formation supérieure, indépendamment du choix de formation à l'âge de 15 ans.

Tableau 3: Odds ratios, transitions vers la formation tertiaire, voie générale

	Professionnel ou général	Intermédiaire	Éducation secon- daire ou tertiaire
Score en lecture/100	1,014*** (0,000)	1,015*** (0,000)	1,003 (0,244)
<i>Éducation des parents :</i>			
Niveau intermédiaire	1,271 (0,159)	1,546 (0,124)	0,658 (0,329)
Niveau supérieur	3,634*** (0,000)	2,528** (0,001)	0,733 (0,485)
Observations		2741	

Notes: méthode d'estimation: logit séquentiel; entre parenthèses: valeur « p ». *p < .1; **p < .05; ***p < .001. Variables de contrôle: né(e) en Suisse, francophone, italophone, nombre de livres au sein du ménage, langue parlée à la maison différente de celle du test, cantons à sélection précoce (11 ans), âge (en mois), hommes.

Sources: TREE, 2002; Falter (2010).

COMPÉTENCES COGNITIVES DE BASE ET MARCHÉ DU TRAVAIL

Compétences cognitives et primes liées à la formation

Un niveau de formation plus élevé conduit à une meilleure situation sur le marché du travail, par exemple sous la forme de salaires plus élevés, d'un moindre risque de chômage ou d'un emploi plus stable. Par contre, on connaît moins bien ce qui génère cette prime liée à la formation. Dans quelle mesure découle-t-elle des compétences enseignées à l'école? Les diplômes agissent-ils comme un signal de compétences innées mais difficilement observables par les employeurs potentiels (voir le modèle de signal présenté par Spence, 1973)? Prises au pied de la lettre, ces questions sont susceptibles de remettre en cause la légitimité des investissements éducatifs. De manière moins radicale, elles soulèvent néanmoins des interrogations sur le contenu des programmes d'études, et par là même, des modalités de l'évaluation.

Les données de l'enquête ALL (*Adult Literacy and Lifeskills Survey*) pour la Suisse nous procurent des informations sur les compétences de base des adultes. Cette enquête, réalisée sous l'égide de l'OCDE, qui mesure les capacités de lecture, de calcul et de résolution de problème, a souvent été considérée comme le PISA des adultes. Les tests sont élaborés de manière à ce qu'il n'y ait pas de références à un environnement professionnel particulier (pour plus de détails, voir Organisation de coopération et de développement économique [OCDE], 2005). Ces mesures de compétence sont fortement corrélées avec le niveau de formation. Par exemple, l'écart entre les personnes actives ayant achevé une formation universitaire et celles dotées d'un diplôme professionnel secondaire s'élève à environ un écart type de la moyenne pour les trois compétences mesurées (voir Falter & Pasche, 2007, 2009). Les femmes ont un niveau de compétence plus faible que les hommes, ce qui s'explique par leur moindre insertion sur le marché du travail. De ce fait, elles tendent à perdre des compétences non mises à profit dans un cadre professionnel. Finalement, on observe également des effets d'âge, les personnes relativement jeunes obtenant des meilleurs scores. Ceci pourrait s'expliquer par le type de compétences mesurées qui induit un effet de cohorte ou une détérioration des capacités cognitives liée au vieillissement.

Plusieurs indicateurs du marché du travail peuvent être utilisés afin de déterminer l'impact de l'éducation et des compétences sur la situation professionnelle des individus. L'indicateur le plus pertinent est le salaire car il inclut plusieurs aspects comme le risque de chômage, les conditions de travail et, surtout, la productivité. À partir des données ALL, on observe qu'une année de formation supplémentaire génère une prime salariale approximativement égale à 8%. Lorsqu'on inclut les compétences dans l'équation, la prime liée à l'éducation diminue d'environ 10% à 20% (voir tableau 4, colonnes I et II). Ce résultat est conforme à ceux trouvés dans la littérature internationale (voir Bowles, Gintins & Osborne 2001). Si on fait l'hypothèse que toutes les compétences de base ont été acquises dans le cadre scolaire et qu'elles ont été mesurées et reportées sans erreurs statistiques, on pourrait affirmer que les compétences de base ne représentent que 10% à 20% des bénéfices de l'éducation (Bowles *et al.*, 2001). Le solde serait attribuable à d'autres compétences comme, à titre d'exemple, la confiance en soi, la persévérance ou des compétences professionnelles spécifiques. Ce type d'interprétation ne conférerait que peu d'importance aux compétences de base généralement évaluées dans les études internationales.

De manière convaincante, Pasche (2009) soutient que l'interprétation faite ci-dessus repose sur des hypothèses trop restrictives. La plus contraignante porte sur le lieu d'acquisition des compétences, celles-ci n'étant pas toutes acquises dans le cadre scolaire mais aussi dans l'environnement familial. Cette distinction devient importante au cas où les compétences

acquises dans le cadre scolaire devaient être mieux rémunérées sur le marché du travail que celles acquises en dehors de l'école. Selon Pasche (2009), les compétences de base acquises dans le cadre scolaire sont rémunérées à un taux trois fois plus élevé que celles acquises en dehors du cadre scolaire⁵. En tenant compte de cette différence, la part des rendements de l'éducation attribuable aux compétences s'élève fortement sans qu'on puisse pour autant la déterminer précisément.

Tableau 4 : Compétences de base et marché du travail : synthèse des résultats

	I	II	III	IV	V	VI
Éducation (années)	0,0779 (0,0055)	0,0671 (0,0057)	–	–	0,0587 (0,007)	0,0670 (0,006)
Compétences de base (ALL)		0,1082 (0,0142)	–0,0561 (0,0512)	0,0052 (0,0624)	0,1985 (0,022)	0,0906 (0,026)
ALL * expérience	–	–	0,0508 (0,0189)	0,0300 (0,0253)	–	–
Insuffisance de compétence (IC)	–	–	–	–	0,1345 (0,040)	0,0088 (0,041)
Excès de compétence (EC)	–	–	–	–	–0,1847 (0,055)	–0,1674 (0,052)
Formation professionnelle * ALL	–	–	–	–	–0,0987 (0,028)	–0,0135 (0,036)
Formation professionnelle * IC	–	–	–	–	–0,0062 (0,055)	–0,1152 (0,062)
Formation professionnelle * EC	–	–	–	–	–0,1771 (0,079)	–0,0646 (0,089)
Observations	1146	1146	475	339	1291	1095
R ²	0,3794	0,4054	0,3124	0,1393	0,4397	0,3146

Notes : Dans toutes les colonnes, la variable dépendante est le taux de salaire mensuel standardisé (en log). Coefficients tirés des sources suivantes : I-II : Pasche (2009) ; III-IV : Falter et Pasche (2009). V-VI : Falter (2010). Données : ALL (*Adult Literacy and Lifeskills Survey*). Les modèles III et IV ont été estimés uniquement sur des personnes ayant comme plus haut niveau de formation respectivement une formation professionnelle secondaire et une formation universitaire. À l'exception de VI, les estimations se réfèrent uniquement à la population masculine.

5. Pasche (2009) effectue la distinction entre compétences acquises à l'école et compétences acquises hors de l'école en régressant, dans le cadre d'une estimation, les scores de compétences sur les diplômes obtenus et un ensemble de variables de contrôle. Les valeurs prédites de l'estimation sont considérées comme les compétences acquises à l'école, le résidu étant les compétences acquises en dehors de l'école.

Compétences selon le type de formation : le rôle de l'asymétrie d'information

L'importance des compétences de base varie vraisemblablement selon le type de formation. Malheureusement, il est difficile de mesurer la contribution des compétences de base par niveau d'éducation. Comme alternative, on peut analyser les mécanismes par lesquels ces compétences sont rémunérées sur le marché du travail et vérifier s'il existe des différences selon le type de formation. Falter et Pasche (2009) analysent l'apprentissage des employeurs, c'est-à-dire comment les employeurs prennent en compte des informations sur la productivité de leurs employés. Ces analyses reposent sur l'hypothèse d'asymétrie d'information sur le marché du travail. Au moment de l'embauche, les employeurs ne connaissent pas parfaitement la productivité de la personne qu'ils engagent. Ils se réfèrent donc à un signal de cette productivité, généralement le diplôme détenu par le travailleur. Concernant les compétences de base comme la lecture ou le calcul, ce signal n'aura pas forcément la même fiabilité selon la formation suivie. En effet, la variance des scores est plus élevée pour les formations de niveau secondaire que pour la formation universitaire. Ainsi, les personnes au bénéfice d'une formation professionnelle secondaire pourraient avoir des difficultés à valoriser leurs compétences sur le marché du travail. Cependant, à mesure que les employés acquièrent de l'expérience professionnelle, l'importance du signal dans la rémunération devrait décroître alors que celle des compétences de base devrait augmenter.

L'estimation du modèle d'apprentissage des employeurs s'effectue en analysant l'effet conjoint de l'éducation, des compétences de base et de l'expérience professionnelle sur les salaires. L'effet positif lié aux compétences devrait augmenter avec l'expérience professionnelle alors que, symétriquement, l'effet du diplôme devrait s'atténuer. Sur l'ensemble de l'échantillon, on observe que les compétences de base accroissent les salaires de manière significative. Un écart type supplémentaire du score moyen des compétences de base procure une prime salariale d'environ 10%. Cependant, ce rendement ne semble pas augmenter avec l'expérience professionnelle, ce qui contredit les hypothèses du modèle. Lorsqu'on estime le modèle par type d'éducation, on observe des différences remarquables (voir tableau 4, colonnes III et IV). Dans le cas de la formation professionnelle secondaire, les compétences de base n'offrent aucun bénéfice salarial à l'entrée sur le marché du travail. Par contre, les gains liés à ces compétences augmentent fortement avec l'expérience professionnelle. Pour les autres types de formation étudiés, les résultats ne mettent pas clairement en évidence cet effet « d'apprentissage » de la part des employeurs. Ces observations pour la Suisse sont conformes à celles obtenues aux États-Unis ou en Allemagne (Arcidiacono, Bayer & Hizmo, 2008; Bauer & Haisken-deNew, 2001).

Les données ALL contiennent également des informations sur les tâches effectuées par les travailleurs sur la place de travail, synthétisées sous la forme d'un indice d'utilisation des compétences au travail. Falter et Pasche (2009) estiment le modèle d'apprentissage des employeurs à l'aide de cette variable d'utilisation des compétences au lieu des salaires comme variable dépendante. Les résultats confirment largement ceux basés sur les salaires. Pour les personnes au bénéfice d'une formation professionnelle secondaire, on observe une relation négative entre le niveau de compétence et l'utilisation de ces compétences sur la place de travail à l'embauche. Ce résultat est surprenant mais souligne probablement l'importance d'acquiescer d'autres compétences spécifiques à une profession, probablement plus facile à démontrer lors d'un apprentissage. Ceci débouche également sur une inadéquation des compétences. Cependant, cette situation n'est que temporaire vu que l'on observe une forte augmentation de l'utilisation des compétences à mesure que les individus acquièrent de l'expérience professionnelle. À nouveau, cette relation n'est vérifiée que pour la formation professionnelle.

Adéquation des compétences sur la place de travail

ALL contient des questions sur l'adéquation entre les capacités des personnes et les postes qu'elles occupent. Les personnes interrogées répondent généralement que leurs compétences sont en adéquation avec le poste occupé, sans qu'il soit possible de déterminer si les travailleurs ne surévaluent pas leurs compétences de base. Alternativement, on peut définir l'adéquation des compétences sur la base du niveau de compétence mesuré dans la base de données et des réponses sur l'utilisation des compétences sur la place de travail. On considérera, par exemple, qu'une personne faiblement compétente mais qui occupe un poste de travail où il est fréquemment fait usage de compétences de base se trouve en inadéquation de compétence. Il en sera de même pour une personne très compétente qui utilise peu ses capacités sur la place de travail. Avec cette approche, l'OCDE (2005) estime qu'environ 25 % de la population active se trouve dans une situation d'excès de compétence dans les données internationales de ALL⁶. Une proportion similaire se trouve en insuffisance de compétence. Ces inadéquations soulèvent de nombreuses questions même si, comme nous l'avons vu précédemment, elles proviennent en partie d'une situation temporaire.

À l'aide d'une définition similaire à celle utilisée par l'OCDE (2005), Falter (2010) analyse la probabilité d'être soit en insuffisance de compétence, soit en excès de compétence pour la Suisse. Puis, il étudie les conséquences

6. Les données internationales ALL couvrent les pays suivants : Bermudes, Canada, États-Unis, Italie, Norvège, l'État du Nuevo León au Mexique et la Suisse.

en termes de salaire. Il en ressort, toutes choses étant égales par ailleurs, que le phénomène d'excès de compétence touche plus particulièrement les femmes, les personnes jeunes et les personnes travaillant à temps partiel⁷. À l'inverse, l'insuffisance de compétence touche plus fréquemment les hommes que les femmes. En ce qui concerne les niveaux de formation, il est à noter que la formation professionnelle ne se distingue pas par une plus forte inadéquation des compétences.

L'analyse des salaires procure des enseignements intéressants. Les personnes en excès de compétence subissent une pénalité de salaire approximativement égale à 10% toutes choses étant égales par ailleurs, alors que l'insuffisance de compétence est associée à une prime salariale comprise entre 6% et 10%⁸. Les analyses ont été par la suite affinées, afin de mesurer d'éventuelles différences selon le type de formation achevée par les travailleurs. Un excès de compétence produit-il les mêmes conséquences pour une personne ayant suivi une formation générale que pour une personne ayant suivi une formation professionnelle? A priori, on pourrait penser que d'autres compétences, plus spécifiques au métier exercé, devraient diminuer l'effet d'un excès de compétence pour les personnes ayant obtenu un diplôme de formation professionnelle. Les résultats empiriques semblent confirmer cette prédiction empirique car les pénalités salariales liées à l'excès de compétence sont moindres pour les personnes au bénéfice d'une formation professionnelle. Ce résultat suggère l'existence de compétences annexes liées aux formations de type non académique, limitant la pertinence des compétences de base.

CONCLUSION: COMPÉTENCES DE BASE ET NOUVELLES ÉVALUATIONS

L'évaluation des compétences des élèves a souvent recours à des tests standardisés dans des matières de base comme la lecture ou le calcul. La pertinence de ces tests ne fait guère de doute, vu que les scores obtenus dans ces matières ont un fort pouvoir prédictif, aussi bien sur l'orientation scolaire que sur la réussite professionnelle. Néanmoins, force est de constater que la maîtrise des matières testées n'est pas l'unique déterminant du succès professionnel. Les différences que l'on observe entre formation

7. Les variables de contrôle comprennent la formation, les métiers occupés, l'industrie, la présence d'enfant au sein du ménage, le niveau de compétence cognitive, la région de résidence.

8. À première vue, une prime salariale pour une insuffisance de compétence peut paraître contre-intuitive. Il faut noter que l'équation de salaire contrôle les compétences cognitives de base. Cela signifie que pour un niveau de compétence donné, les personnes ayant plus d'activités liées à ces compétences sur la place de travail ont un salaire plus élevé.

professionnelle et formation générale confirment ce constat. Quelles en sont les implications concrètes pour les systèmes d'évaluation ?

Quelles peuvent être ces « autres » compétences ? Des travaux récents dans le domaine de l'économie de l'éducation ont mis en exergue le rôle important des compétences dites non cognitives. De manière plus spécifique, ce groupe de compétences se réfère à des indicateurs psychologiques comme l'indice d'estime de soi de Rosenberg (1965) ou l'indice de contrôle personnel de Rotter (1966). Ces variables comportementales sont corrélées avec le niveau de formation, sans que la causalité entre les deux variables soit bien établie. Pasche (2009) démontre, à l'aide de données américaines, que les employeurs utilisent le niveau de formation des individus à l'embauche afin de pouvoir discriminer les personnes relativement bien dotées en termes de compétences non cognitives et celles qui ne le sont pas. Cela signifie que des compétences non cognitives plus élevées font effectivement partie des rendements de la formation scolaire. La prise en compte de ce type de compétences dans un système d'évaluation demeure un défi difficile. En effet, l'évaluation scolaire doit être liée aux objectifs implicites ou explicites assignés au système scolaire. Bien que les compétences non cognitives soient partiellement acquises dans un cadre scolaire, on ne peut pas dire qu'elles soient associées à des objectifs de performance.

La nécessité d'évaluer d'autres compétences, quelle que soit leur nature, dépend également de leur complémentarité avec les compétences de base. Une série de travaux récents sur la construction du capital humain montrent que la complémentarité entre les différentes compétences est très forte (voir Heckman & Rubinstein, 2001 ; Heckman, Stixrud & Urzua, 2006). Dans la perspective de l'évaluation, ce constat est à double tranchant. Si les compétences de base ne sont qu'un élément parmi d'autres, elles n'en demeurent pas moins indispensables.

L'évaluation scolaire porte principalement sur des compétences de base (PISA) ou des compétences génériques (IALS, ALL) qui sont acquises par le biais d'une formation générale. À notre sens, il s'agit de la principale limite du mode d'évaluation actuel. Nous avons discuté des différences notables quant à l'effet des compétences générales sur le marché du travail selon le type de formation. Il existe donc une nécessité de fournir un instrument en adéquation avec la formation professionnelle, d'où les travaux menés en Allemagne sur un projet intitulé familièrement *VET-PISA* (le projet est décrit par Baethge, Arends & Winther, 2009). Ce type de projet est doublement important. Premièrement, il comblerait un vide pour ce qui demeure la voie de formation modale dans de nombreux pays, à commencer par la Suisse et l'Allemagne. Deuxièmement, couplées avec les indicateurs traditionnels de compétences, des informations sur les compétences spécifiquement professionnelles permettraient à terme de mieux comprendre quelles sont les compétences réellement nécessaires sur le marché du travail à court et

à long terme. Enfin, lorsqu'on élargit le spectre de notre discussion, des outils d'évaluation de la formation professionnelle sont nécessaires si l'on veut encourager la formation tout au long de la vie. En effet, aussi bien du point de vue de la demande émanant des travailleurs que des décideurs publics, une évaluation fiable de ce type de formation est nécessaire afin d'éviter de coûteux gaspillages.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Arcidiacono, P., Bayer, P. & Hizmo, A. (2008). *Beyond signaling and human capital: education and the revelation of ability* (NBER Working Paper n° 13951). Cambridge: National Bureau of Economic Research.
- Autor, D. & Dorn, D. (2009). This Job is « Getting Old » : Measuring Changes in Job Opportunities using Occupational Age Structure. *American Economic Review Papers and Proceedings*, 99(2), 45-51.
- Baethge, M., Arends, L. & Winther, E. (2009). International Large-Scale Assessment on Vocational and Occupational Education and Training. In F. Oser, U. Renold, E. G. John, E. Winther & S. Weber (Ed.), *VET-Boost, Towards a Theory of Professional Competencies* (pp. 3-24). Rotterdam: Sense-Publishers.
- Bauer, P. & Riphahn, R. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters*, 91(1), 90-97.
- Bauer, T. K. & Haisken-DeNew, J. P. (2001). Employer learning and the returns to schooling. *Labour Economics*, 8(2), 161-180.
- Bertschy, K., Cattaneo, A. & Wolter, S. (2009). PISA and the Transition into the Labour Market. *LABOUR: Review of Labour Economics and Industrial Relations*, 23, 111-137.
- Bowles, S., Gintis, H. & Osborne, M. (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature*, 39, 1137-1176.
- Checchi, D. & Flabbi, L. (2007). *Intergenerational mobility and schooling decisions in Italy and Germany: the impact of secondary school track* (IZA Discussion Paper N° 2876). Bonn: Institute for the Study of Labor (IZA).
- Falter, J.-M. (2010). *Parental Background, Track Choices and Schooling Inequality in Switzerland*. Manuscrit non publié, Université de Genève, Leading House en économie de l'éducation.
- Falter, J.-M. & Pasche, C. (2007). *Compétences, formation et marché du travail en Suisse*. Neuchâtel: Office fédéral de la statistique.
- Falter, J.-M. & Pasche, C. (2009). Does Vocational Education Affect Employer Learning? Evidence from Switzerland. *Empirische Pädagogik*, 23(4), 459-478.
- Hanushek, E. A. & Wössmann, L. (2007). *Education Quality and Economic Growth*. Washington, DC: The World Bank.

- Heckman, J. & Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2), 145-149.
- Heckman, J., Stixrud, J. & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour. *Journal of Labor Economics*, 24(3), 411-482.
- Oesch, D. & Rodriguez Menés, J. (sous presse). Upgrading or Polarization? Occupational Changes in Britain, Germany, Spain and Switzerland, 1990-2008. *Socio-Economic Review*.
- Organisation de coopération et de développement économiques. (2005). *Learning a living: first results of the adults literacy and lifeskills survey*. Ottawa et Paris.
- Organisation de coopération et de développement économiques. (2010). *Programme international pour le suivi des acquis des élèves (PISA)*. Consulté le 15 avril 2011 dans http://www.oecd.org/about/0,3347,fr_2649_35845621_1_1_1_1_1,00.html
- Pasche, C. (2009). *Schooling, Ability and Wages*. Thèse de doctorat en Sciences économiques, Université de Genève.
- Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.
- Rotter, J. B. (1966). *Generalized Expectancies for Internal Versus External Control of Reinforcement*. Washington, DC: American Psychological Association.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355-374.
- Stalder, B. E. (2005). *Das intellektuelle Anforderungsniveau von 105 Berufslehren*. Bern: TREE.

Notices biographiques

Anne Barrère est agrégée de lettres modernes et docteure en sociologie. Elle est actuellement professeure de sciences de l'éducation à l'Université Paris Descartes. Son principal chantier de recherche concerne le travail à l'école dans l'enseignement secondaire, analysé du point de vue de différents acteurs (*Travailler à l'école*, PUR, 2003), notamment les chefs d'établissement (*Sociologie des chefs d'établissement, Les managers de la République*, PUF, 2006). C'est à partir de cette dernière enquête qu'elle a mené une réflexion sur l'évaluation.

Adresse : Université Paris Descartes
Faculté de Sciences Humaines et Sociales
Laboratoire CERLIS
45, rue des Saints-Pères
F-75270 Paris Cedex 06
Courriel : anne.barrere@parisdescartes.fr

Pascal Bressoux est docteur en sciences de l'éducation, professeur à l'Université Pierre-Mendès-France de Grenoble où il dirige le Laboratoire des Sciences de l'Éducation (EA n° 602). Ses travaux portent sur l'efficacité de l'enseignement : il s'intéresse aux effets du contexte et des pratiques scolaires sur les acquis des élèves ainsi qu'aux effets du jugement des enseignants sur les perceptions de soi des élèves. Il a publié notamment *Modélisation statistique appliquée aux sciences sociales* (De Boeck, 2008) et, en collaboration avec Pascal Pansu, *Quand les enseignants jugent leurs élèves* (PUF, 2003).

Adresse : Université Pierre-Mendès-France
Laboratoire des Sciences de l'Éducation
1251, avenue Centrale
B.P. 47
F-38040 Grenoble Cedex 9
Courriel : Pascal.Bressoux@upmf-grenoble.fr

Branka Cattonar, docteure en sociologie, est chercheuse au CIRTES (Centre interdisciplinaire de recherche travail, État et société) et au GIRSEF (Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation), et chargée de cours invitée à l'EDEF (École d'éducation et de formation) à l'Université catholique de Louvain. Ses travaux portent sur l'évolution des métiers de l'éducation (en particulier des identités professionnelles, des modèles de professionnalité et de formation) en lien avec les politiques éducatives et les nouveaux modes de régulation du système scolaire. Plus récemment, ses travaux ont été consacrés aux rapports entre les connaissances et les politiques éducatives.

Adresse: Groupe interdisciplinaire de recherche
sur la socialisation, l'éducation et la formation (GIRSEF)
Place Montesquieu, 1/14
B-1348 Louvain-la-Neuve
Courriel : branka.cattonar@uclouvain.be

Marcel Crahay, dont la thèse de doctorat portait sur les implications pédagogiques de la théorie de Piaget à l'éducation préscolaire, a été professeur de pédagogie théorique et expérimentale à l'Université de Liège, jusqu'en 2004. Tout en conservant une implication dans son université d'origine, il est actuellement professeur ordinaire à l'Université de Genève et occupe un poste intitulé « Développement, apprentissage et intervention en situations scolaires ». Il s'intéresse aux apports de la psychologie aux problématiques éducatives. Il porte également un intérêt à l'organisation des systèmes éducatifs et, plus généralement, aux dispositifs qui peuvent rendre l'école plus efficace et moins inégalitaire. Il est l'auteur ou le coauteur de nombreux articles et de plusieurs livres, dont les derniers ont pour ambition de mettre à disposition des enseignants les apports de la psychologie cognitive et sociale dans la perspective de favoriser les apprentissages scolaires de tous.

Adresse: Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : Marcel.Crahay@unige.ch

Jean-Marc Falter est maître d'enseignement et de recherche en économie à l'Université de Genève. Il a obtenu dans cette même université son doctorat en économie sur le thème du travail indépendant. Depuis 2006, il occupe le poste de Directeur chargé de la recherche au sein de la Leading House en économie de l'éducation, un pôle d'excellence basé à l'Université de Genève et financé par la Confédération helvétique. Ses recherches portent sur les domaines de l'économie de l'éducation, du marché du travail et des inégalités.

Adresse : Université de Genève
Département d'Économie Politique
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : Jean-Marc.Falter@unige.ch

Georges Felouzis est professeur de sociologie des politiques éducatives à l'Université de Genève. Il a créé et anime le Groupe Genevois d'Analyse des Politiques Éducatives (GGAPE). Ses travaux portent sur la sociologie de l'école et de l'enseignement supérieur. Il est l'auteur de plusieurs articles et ouvrages sur l'enseignement secondaire, notamment *L'apartheid scolaire*, coécrit avec Françoise Liot et Joëlle Perroton (Le Seuil, 2005), ainsi que sur l'enseignement supérieur. Ses travaux portent aujourd'hui sur l'évaluation des politiques éducatives et des systèmes d'enseignement en Suisse et en France, et sur les comparaisons internationales en matière d'efficacité et d'équité à partir des enquêtes PISA.

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : Georges.Felouzis@unige.ch

Siegfried Hanhart est professeur en analyse économique et politique des systèmes éducatifs à l'Université de Genève. Il codirige l'Unité « politique, économie, gestion et éducation internationale » (PEGEI). Ses travaux portent sur la comparaison des coûts, de l'efficacité et du financement des systèmes éducatifs, tant scolaires que professionnels et universitaires. Il est le coauteur de plusieurs ouvrages et articles sur la formation professionnelle en entreprise.

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : Siegfried.Hanhart@unige.ch

Jean-Marc Huguenin est chargé de cours et de projets à l'Institut de hautes études en administration publique de Lausanne. Il est titulaire d'un *Master of Science in Finance*, Faculté des hautes études commerciales (HEC) de l'Université de Lausanne, et d'un brevet d'aptitude pédagogique pour l'enseignement secondaire supérieur du canton de Berne. Il a précédemment été économiste à la Banque nationale suisse (BNS) et dirigé les Économistes indépendants, un cabinet d'intelligence économique fondé en 1999. Ses domaines de recherche et d'expertise sont l'évaluation de la productivité et

de l'efficacité des organisations, l'économie, le management et l'évaluation des systèmes éducatifs.

Adresse : Institut de hautes études en administration publique

Quartier UNIL Mouline

CH-1015 Lausanne

Courriel : Jean-Marc.Huguenin@idheap.unil.ch

Claude Lessard est professeur titulaire au département d'Administration et Fondements de l'Éducation de la Faculté des sciences de l'éducation de l'Université de Montréal, dont il fut le doyen de 1991 à 1995. Il a été directeur/fondateur du Centre de recherche interuniversitaire sur la formation et la profession enseignante à Montréal (CRIFPE-Montréal). Il est titulaire de la chaire de recherche de l'Université de Montréal sur les métiers de l'éducation. Il s'intéresse à la sociologie de l'éducation, à la psychosociologie du corps enseignant, à l'analyse des politiques et à leurs effets sur le travail des enseignants et des directions d'établissement.

Adresse : Claude Lessard

205 Avalon

Rosemère, Québec J7A 3T5

Canada

Courriel : claudellessard@umontreal.ca

Laurent Lima, après un doctorat en sciences de l'éducation soutenu à Grenoble en 2001, a occupé un poste d'ingénieur d'étude au Laboratoire des Sciences de l'Éducation de Grenoble, puis un poste de maître de conférences dans ce même laboratoire. Ses travaux de recherche sur l'efficacité des systèmes d'enseignement se développent selon trois dimensions. La première étudie l'effet d'approches didactiques innovantes sur le développement de la compréhension de l'écrit à l'école primaire. La seconde envisage les liens entre pratiques enseignantes, organisations scolaires et performances des élèves. Enfin, la troisième porte sur l'étude des facteurs sociaux et institutionnels qui affectent les jugements émis par les étudiants de l'enseignement supérieur sur leurs études.

Adresse : Université Pierre-Mendès-France

Laboratoire des Sciences de l'Éducation

1251, avenue Centrale

B.P. 47

F-38040 Grenoble Cedex 9

Courriel : laurent.lima@upmf-grenoble.fr

Catherine Mangez est sociologue au GIRSEF (Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation) à l'Université catholique de Louvain. Elle termine cette année une thèse de doctorat sur les instruments des politiques éducatives en Belgique francophone, sous la

direction de Christian Maroy. Elle participe depuis quatre ans à la recherche européenne *Knowledge and Policy* qui traite du rôle des connaissances dans la construction et la régulation des politiques de santé et d'éducation en Europe. Ses principaux domaines de recherche sont la sociologie de l'éducation et l'analyse de l'action publique et de ses instruments.

Adresse: Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation (GIRSEF)

Place Montesquieu, 1/14

B-1348 Louvain-la-Neuve

Courriel: catherine.mangez@uclouvain.be

Christian Maroy est sociologue et professeur titulaire à l'Université de Montréal. Il est titulaire de la chaire de recherche du Canada en politiques éducatives, chercheur au CRIFPE (Centre de recherche interuniversitaire sur la formation et la profession enseignante) et chercheur associé au GIRSEF (Groupe interdisciplinaire de recherche sur la socialisation, l'éducation et la formation) à l'Université catholique de Louvain. Ses intérêts de recherche portent sur les politiques éducatives et les comparaisons internationales, la régulation et la gouvernance des systèmes d'enseignement, les établissements scolaires dans leur contexte local, la démocratisation de l'enseignement supérieur. Il a notamment publié *École, régulation et marché. Une comparaison de six espaces scolaires en Europe* (PUF, 2006).

Adresse: Université de Montréal

Département Administration et Fondements de l'Éducation

C.P. 6128, succursale Centre-Ville

Montréal, Qc, H3C 3J7

Canada

Courriel: christian.maroy@umontreal.ca

Nathalie Mons est maître de conférences en sociologie à l'Université Paris-Est Marne-la-Vallée et chercheure au Laboratoire techniques, territoires et sociétés (LATTS). Professeure invitée au *London Institute of Education* (2007-2009), elle consacre ses recherches à l'analyse des politiques publiques d'éducation (autonomie scolaire, décentralisation, évaluation, libre choix de l'école...) dans une perspective comparatiste internationale. L'ouvrage *Les nouvelles politiques éducatives* (PUF, 2007) ou le numéro spécial de la *Revue française de pédagogie* (2008) « Évaluation des politiques éducatives et comparaisons internationales » présentent ses axes de recherche.

Adresse: Université Paris-Est, LATTS et IFRIS

Bâtiment Bois de l'Étang

Cité Descartes

F-77420 Champs sur Marne

Courriel: nmons@wanadoo.fr

Catherine Paradeise, sociologue, docteure ès lettres, est actuellement professeure à l'Université Paris-Est Marne-la-Vallée, chercheure au LATTS (Laboratoire techniques, territoires et sociétés) et présidente de l'IFRIS (Institut francilien recherche innovation et société). Elle a notamment été directrice adjointe du département Sciences Humaines et Sociales (SHS) du CNRS et directrice adjointe de l'École normale supérieure (ENS) de Cachan. Elle a longtemps consacré ses travaux à l'analyse des marchés du travail, des professions, de la négociation collective et des relations professionnelles. Elle s'intéresse aujourd'hui aux organisations et aux professions scientifiques et universitaires.

Adresse : Université Paris-Est, LATTS et IFRIS
Bâtiment Bois de l'Étang
Cité Descartes
F-77420 Champs sur Marne
Courriel : Catherine.Paradeise@univ-mlv.fr

Jean-Paul Payet, sociologue, est professeur à l'Université de Genève. Ses recherches sur l'école en contexte désqualifié s'articulent à une réflexion sur les processus de reconnaissance des « acteurs faibles » dans les institutions contemporaines. Il anime l'équipe de recherche Sociologie de l'action-Transformations des institutions-Éducation (SATIE).

Adresse : Université de Genève
Faculté de Psychologie et des Sciences de l'Éducation
40, boulevard du Pont-d'Arve
CH-1211 Genève 4
Courriel : Jean-Paul.Payet@unige.ch

José V. Ramirez a obtenu un doctorat en sciences économiques et sociales, mention économie politique, de l'Université de Genève en 2000. Il a ensuite été successivement maître assistant et maître d'enseignement et de recherche à l'Université de Genève avant de devenir professeur en économie politique à la Haute école de gestion de Genève en 2004. Depuis 2006, il est également directeur adjoint de la Leading House en économie de l'éducation de Genève. Ses domaines de recherche et d'expertise sont directement en lien avec le marché du travail et l'éducation, notamment sur les questions d'inégalités, de ségrégation et de discrimination.

Adresse : Haute école de gestion de Genève
7, route de Drize
CH-1227 Carouge
Courriel : jose.ramirez@hesge.ch

Georges Solaux, enseignant du primaire, cadre de l'Éducation nationale puis professeur des universités a réalisé ses travaux de recherche à l'IREDU (Institut de recherche sur l'éducation) de l'Université de Bourgogne dont il fut le vice-président. Son expérience du système éducatif français ainsi que de nombreux travaux d'expertise effectués dans les pays en voie de développement l'ont conduit à s'intéresser à la genèse et à l'évaluation des politiques d'éducation et de formation.

Adresse : IREDU Pôle AAFE

Esplanade Érasme

B.P. 26513

F-21065 Dijon

Courriel : georges.solaux@u-bourgogne.fr

Jean-Claude Thoenig, directeur de recherche (ém.) au CNRS rejoint, après des études de sociologie à l'Université de Genève, le Centre de Sociologie des Organisations à Paris, devient professeur à l'École Polytechnique Fédérale de Lausanne, puis à l'INSEAD de Fontainebleau. Il a publié récemment avec Charles Waldman *De l'entreprise marchande à l'entreprise marquante* (Éditions d'Organisation, 2005), avec David Courpasson *Quand les cadres se rebellent* (Vuibert, 2008), et avec Claude Michaud *L'organisation et ses langages. Interpréter pour agir* (Presses de l'Université Laval, 2009). Il travaille actuellement avec Catherine Paradeise sur la production de la qualité académique.

Adresse : Dauphine Recherches en Management

Université Paris-Dauphine

Place du Maréchal de Lattre de Tassigny

F-75775 Paris Cedex 16

Courriel : jeanclaude.thoenig@free.fr

Table des matières

Sommaire	5
Politiques éducatives et évaluation :	
nouvelles tendances, nouveaux acteurs	7
<i>Georges Felouzis & Siegfried Hanhart</i>	
<i>Université de Genève</i>	
Quels rapports entretiennent les procédures évaluatives avec les sciences sociales en général ?	11
De la régulation par la qualification à la régulation par les <i>accounts</i> : émergence de nouveaux acteurs collectifs ?	13
Quelles sont les conditions de validité et de légitimité de l'évaluation ?	16
Quelles définitions de l'efficacité pour l'évaluation des politiques éducatives ?	19
<i>Efficacité interne et efficience des politiques éducatives</i>	20
<i>Efficacité externe des politiques éducatives :</i>	
<i>le problème des effets non mesurables</i>	21
<i>Efficacité interne versus efficacité externe ?</i>	22
Gouverner l'éducation par les nombres : usages et conséquences	23
Références bibliographiques	28
Réformes et ordres universitaires locaux	33
<i>Catherine Paradeise* & Jean-Claude Thoenig**</i>	
* <i>Université Paris-Est Marne-la-Vallée, LATTs et IFRIS</i>	
** <i>Université Paris-Dauphine, DRM</i>	
Les jugements de qualité percutés par « l'excellence »	35
Un statut théorique du local	36
Une typologie de la production locale de la qualité : quatre positionnements face au prestige et à l'excellence	39
<i>Le gratin</i>	40
<i>Les aspirants</i>	40
<i>Les patriciens</i>	42
<i>Les missionnaires</i>	43
Les appariements entre dimensions honorifiques et instrumentales de la qualité	43
<i>Le capital social</i>	44
<i>L'utilitarisme opportuniste</i>	46
<i>La collégialité</i>	47
	227

<i>L'égalitarisme</i>	49
Conclusion	50
Références bibliographiques	51
La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone : nouveau paradigme politique et médiation des experts	53
<i>Christian Maroy* & Catherine Mangez**</i>	
<i>*Universités de Montréal et de Louvain, CRIFPE et GIRSEF</i>	
<i>**Université de Louvain, GIRSEF</i>	
Un système scolaire fragmenté et décentralisé	55
De nouvelles politiques scolaires de pilotage et d'évaluation	56
<i>Les orientations générales des politiques scolaires</i>	57
Développement de règles centrales et transversales à tous les réseaux d'enseignement	57
Valorisation et encadrement de l'autonomie des établissements	57
Développement de l'évaluation et du pilotage	58
<i>La construction lente du pilotage du système scolaire</i>	58
Un changement de paradigme politique : de l'institution au système de production scolaire	62
<i>Les valeurs de référence</i>	64
<i>Les normes définissant des écarts entre le réel perçu et le réel souhaité</i>	64
<i>Les théories de l'action</i>	65
Conditions structurelles et acteurs du changement	66
<i>Conditions structurelles</i>	67
<i>Acteurs entrepreneurs</i>	68
<i>Ressources cognitives</i>	69
Conclusion	71
Références bibliographiques	72
L'évaluation des performances scolaires des élèves : un instrument d'évaluation des politiques éducatives ?	77
<i>Nathalie Mons* & Marcel Crahay**</i>	
<i>*Université Paris-Est Marne-la-Vallée, LATTS</i>	
<i>**Universités de Genève et de Liège</i>	
Introduction	77
Cadres théoriques de l'évaluation des politiques publiques	79
<i>La justification de l'évaluation dans les politiques publiques</i>	80
<i>Cadres théoriques en éducation :</i>	
<i>l'évaluation des acteurs opérationnels est privilégiée</i>	83
Politiques d'évaluation standardisée dans l'OCDE : une évaluation des acteurs politiques en retrait	84
<i>Des appareils statistiques nationaux pauvres qui permettent difficilement l'évaluation des politiques publiques</i>	85
<i>Quel usage est-il fait des évaluations standardisées ?</i>	88
Discussion	91
Références bibliographiques	95

La place de l'évaluation dans les politiques éducatives: le cas de la taille des classes à l'école primaire en France	99
<i>Pascal Bressoux & Laurent Lima</i>	
<i>Laboratoire des Sciences de l'Éducation</i>	
<i>Universités Pierre-Mendès-France et Joseph-Fourier, Grenoble</i>	
Le rapport Meuret, l'avis du HCEE et la <i>Note Évaluation</i> du MEN	102
Un réexamen des travaux antérieurs	104
Un réexamen des données de l'expérimentation CP à effectifs réduits	108
<i>Méthode</i>	108
Participants	108
Matériel	109
Procédure	109
<i>Résultats</i>	109
Une analyse des populations élèves et enseignants	109
Estimation des effets de l'expérimentation sur les acquis des élèves	111
<i>Discussion</i>	117
Conclusion	120
Références bibliographiques	122
Les contradictions organisationnelles de l'évaluation. L'expérience des chefs d'établissement français	125
<i>Anne Barrère</i>	
<i>Université Paris Descartes, CERLIS</i>	
Entre bureaucratie et postbureaucratie	127
<i>Une tension structurelle en France</i>	127
<i>Solutions quotidiennes</i>	128
Évaluation et pouvoir	131
<i>Du côté des familles</i>	132
<i>Du côté des enseignants</i>	133
L'évaluation : rationalité en moyens ou en valeur?	136
<i>Une solution à tout faire</i>	137
<i>Un projet unidimensionnel?</i>	137
Conclusion	138
Références bibliographiques	140
Développement de l'évaluation externe et restructuration du métier de direction d'établissement scolaire au Canada	143
<i>Branka Cattonar* & Claude Lessard**</i>	
<i>*Université de Louvain, CIRTES et GIRSEF</i>	
<i>**Université de Montréal, CRIFPE</i>	
Évolution du mode de régulation scolaire et redéfinition normative du rôle des directions d'établissement	143
<i>Vers une régulation par les « résultats »</i>	143
<i>L'établissement scolaire comme communauté éducative mobilisée et responsable</i>	144
<i>Le directeur d'établissement comme acteur central du changement</i>	145

L'enquête	148
Le pilotage par les résultats : des effets négatifs, mais une adhésion au principe	149
Le rapport au métier	152
<i>Une multitude de rôles et de responsabilités</i>	
<i>à assumer, une forte valorisation du travail pédagogique</i>	152
<i>Une satisfaction professionnelle globale</i>	157
Quelle restructuration du métier de direction d'établissement ?	159
Références bibliographiques	161
Évaluation, concertation, décision : quelle régulation pour le système éducatif ?	
Le cas de l'enseignement primaire genevois	165
<i>Georges Solaux*, Jean-Marc Huguenin**, Jean-Paul Payet*** & José V. Ramirez****</i>	
<i>*Université de Bourgogne, **IDHEAP, Lausanne, ***Université de Genève ****HEG et Leading House en économie de l'éducation, Genève</i>	
Le nouveau fonctionnement de l'enseignement primaire du canton de Genève	166
L'Observatoire du fonctionnement de l'enseignement primaire	170
Une équipe d'experts	172
Approche interdisciplinaire appliquée à la reconfiguration des établissements	176
Conclusion	181
Références bibliographiques	182
L'évaluation des élèves : externalisation et hétérogénéité	185
<i>José V. Ramirez</i>	
<i>HEG et Leading House en économie de l'éducation, Genève</i>	
L'externalisation de l'évaluation des élèves	187
Institutionnalisation de l'évaluation externe : signaux et incitations	188
Externalisation et interaction enseignant-élève	191
Les évaluations communes de mathématiques au secondaire inférieur du canton de Genève	193
Conclusion	198
Références bibliographiques	199
Quelle est l'importance des compétences évaluées dans les avantages liés à la formation sur le marché du travail ?	201
<i>Jean-Marc Falter</i>	
<i>Université de Genève, Leading House en économie de l'éducation</i>	
Introduction	201
Compétences évaluées et transitions scolaires	203
Compétences cognitives de base et marché du travail	209
<i>Compétences cognitives et primes liées à la formation</i>	209

<i>Compétences selon le type de formation :</i>	
<i>le rôle de l'asymétrie d'information</i>	212
<i>Adéquation des compétences sur la place de travail</i>	213
Conclusion : compétences de base et nouvelles évaluations	214
Références bibliographiques	216
Notices biographiques	219
Table des matières	227

Collection Raisons Éducatives

Concept éditorial

Raisons éducatives est une collection de la Section des sciences de l'éducation de l'Université de Genève publiée en partenariat avec les Éditions De Boeck Université.

Chaque édition traite un thème de façon approfondie, avec des regards multiples permettant la controverse scientifique, en regroupant des travaux effectués aussi bien dans la section que dans d'autres universités.

Raisons éducatives vise à...

- promouvoir un support scientifique de référence et de langue française, pouvant en principe couvrir l'ensemble des domaines de la recherche en éducation, et adressé en particulier à la communauté des chercheurs et des étudiants en sciences de l'éducation;
- privilégier les questions vives des théorisations éducatives, en développant une approche restituant la diversité disciplinaire des sciences de l'éducation, ainsi que leur spécificité et leur inscription plus large dans le domaine des sciences humaines et sociales;
- contribuer à donner forme, au fur et à mesure des publications et selon les cas:
 - à un repérage des objets et concepts émergents,
 - à un approfondissement des apories ou difficultés repérées comme majeures dans le champ,
 - à une ouverture de « chantiers » réflexifs nouveaux;
- développer les liens scientifiques entre universités sur des thèmes significatifs au plan des questions et des hypothèses de recherche.

Comité de rédaction

Le travail principal du comité de rédaction est de définir le thème retenu pour chaque publication de la collection, de repérer et susciter les contributions souhaitées ainsi que de veiller au respect de l'orientation éditoriale.

Laurent FILLIETTAZ & Lucie MOTTIER LOPEZ (responsables du comité de rédaction), Abdeljalil AKKARI, Kristine BALSLEV, Jean-Michel BAUDOUIN, Sylvie CÈBE, Joaquim DOLZ, Jean-Luc DORIER, Georges FELOUZIS, Gérard FIGARI, Siegfried HANHART, Francia LEUTENEGGER, Florence LIGOZAT, Valérie LUSSI, Olivier MAULINI, Françoise MERHAN, Jean-Paul PAYET, Christophe RONVEAUX, Bernard SCHNEUWLY, Sabine VANHULLE.

ADRESSE: RAISONS ÉDUCATIVES

Université de Genève – *Faculté de psychologie et des sciences de l'éducation* – Uni-Pignon, 40 Bd du Pont-d'Arve, CH-1211 Genève 4

Courrier électronique : publications-ssed-infos@unige.ch

Secrétariat: Leïla LOUCA

Comité scientifique

Le comité scientifique regroupe des chercheurs en éducation suisses et étrangers. La composition de ce comité est représentative des champs disciplinaires et des domaines de référence propres aux sciences de l'éducation. Chaque contribution de la série est soumise, pour expertise, à deux lecteurs du comité scientifique.

Abdeljalil AKKARI, Université de Genève (Suisse) – Cristina ALLEMANN-GHIONDA, Université de Cologne (Allemagne) – Jean-Marie BARBIER, Conservatoire National des Arts et Métiers, Paris (France) – Jean-Michel BAUDOUIN, Université de Genève (Suisse) – Élisabeth BAUTIER, Université de Paris VIII Saint-Denis (France) – Alex BLANCHET, Unité de recherche pour le pilotage des systèmes pédagogiques, Lausanne (Suisse) – Étienne BOURGEOIS, Université de Genève (Suisse) – Christian BRASSAC, Université de Nancy 2 (France) – Jean-Paul BRONCKART, Université de Genève (Suisse) – Pierre CASPARD, Institut national de recherche pédagogique, Paris (France) – Nadine CHARBONNEL, Université de Strasbourg (France) – Bernadette CHARLIER, Université de Fribourg (Suisse) – Anne-Marie CHARTIER, Institut national de recherche pédagogique, Paris (France) – Yves CHEVALLARD, IUFM, Aix-Marseille (France) – Marcel CRAHAY, Universités de Genève et de Liège (Suisse et Belgique) – Lucien CRIBLEZ, Université de Zurich (Suisse) – Isabelle DELCAMBRE, Université Charles de Gaulle, Lille 3 (France) – Joaquim DOLZ, Université de Genève (Suisse) – Christine FÉLIX, IUFM, Aix-Marseille (France) – Annick FLUCKIGER, Université de Genève (Suisse) – José-Luis GARCIA GARRIDO, Université nationale d'éducation à distance, Madrid (Espagne) – Christiane GILLIÉRON, Université de Genève (Suisse) – Jean-Luc GURTNER, Université de Fribourg (Suisse) – Jean HEBRARD, Ministère de l'éducation nationale, Paris (France) – France HENRI, Télé-université, Sherbrooke (Canada) – Rita HOFSTETTER, Université de Genève (Suisse) – Geneviève JACQUINOT, Institut national d'agronomie et Université de Paris VIII Saint-Denis (France) – Dominique LAFONTAINE, Université de Liège (Belgique) – Yves LENOIR, Université de Sherbrooke (Canada) – Jean-Philippe LERESCHE, Université de Lausanne (Suisse) – Louis LEVY-GARBOUA, Université Paris I (France) – Philippe LOSEGO, HEP Vaud Lausanne (Suisse) – Pierre MARC, Université de Neuchâtel (Suisse) – Eduardo MARTÍ, Université de Barcelone (Espagne) – Robert MARTY, Université de Perpignan (France) – Jean-Pierre MEUNIER, Université Catholique de Louvain (Belgique) – Cléopâtre MONTANDON, Université de Genève (Suisse) – Nicole MOSCONI, Université de Paris X Nanterre (France) – Lucie MOTTIER LOPEZ, Université de Genève (Suisse) – Christian NIDEGGER, Service de la recherche en éducation, Genève (Suisse) – Fernand OUELLET, Université de Sherbrooke (Canada) – Jean-Louis PAOUR, Université de Provence, Marseille (France) – Jean-Paul PAYET, Université de Genève (Suisse) – Anne-Nelly PERRET-CLERMONT, Université de Neuchâtel (Suisse) – Marie-Jeanne PERRIN-GLORIAN, Université Paris VII (France) – André PETITAT, Université de Lausanne (Suisse) – Gianreto PINI, Université de Genève (Suisse) – Yves REUTER, Université Charles de Gaulle, Lille 3 (France) – René RICKENMANN, Université de Genève (Suisse) – Laurence RIEBEN, Université de Genève (Suisse) – Christophe RONVEAUX, Université de Genève (Suisse) – Marie-Hélène SALIN, Université de Bordeaux 2 (France) – Frédéric SAUSSEZ, Université de Montréal (Canada) – Gérard SENSÉVY, IUFM de Bretagne, Université de Rennes (France) – Yves SCHWARZ, Université de Provence, Aix-en-Provence (France) – Jacques SILBER, Université de Tel-Aviv (Israël) – Claude SIMARD, Université de Laval, Québec (Canada) – François STOLL, Université de Zürich (Suisse) – Juan-Carlos TEDESCO, IIEP, Buenos Aires (Argentine) – Sabine VANHULLE, Université de Genève (Suisse).