Les transformations de l'organisation scolaire : retour vers la qualité de l'enseignement

Changes in school organization: back to teaching quality

*Las transformaciones de la organización escolar: retorno hacia la calidad de la
enseñanza*

Veränderungen der Schulorganisation: zurück zur Unterrichtsqualität

Matthis Behrens

Édition électronique

URL : <http://journals.openedition.org/rfp/2694>

DOI : 10.4000/rfp.2694

ISSN : 2105-2913

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 15 mars 2011

Pagination : 71-90

ISBN : 978-2-7342-1203-4

ISSN : 0556-7807

Référence électronique

Matthis Behrens, « Les transformations de l'organisation scolaire : retour vers la qualité de l'enseignement », *Revue française de pédagogie* [En ligne], 174 | janvier-mars 2011, mis en ligne le 15 mars 2015, consulté le 01 mai 2019. URL : <http://journals.openedition.org/rfp/2694> ; DOI : 10.4000/rfp.2694

Les transformations de l'organisation scolaire : retour vers la qualité de l'enseignement

Matthis Behrens

L'évaluation se trouve au cœur de l'activité scolaire. Alors qu'elle servait surtout à certifier des acquis ou à réguler les apprentissages, elle est devenue un outil de gestion dont les administrations scolaires se servent pour piloter l'organisation scolaire. En analysant les parties constitutives d'un système éducatif, et plus précisément dans les pays germanophones, l'article interroge les besoins d'information à l'intérieur d'un système de production scolaire, en tenant compte à la fois des besoins formatifs et des impératifs de l'accountability. Un travail avec les enseignants sous forme de recherche collaborative permet d'interpréter les résultats d'évaluation et de planifier des remédiations.

Descripteurs (TESE) : évaluation, diffusion de l'information, norme, formation des enseignants, politique en matière d'éducation, système éducatif.

Pourquoi les résultats en éducation sont-ils érigés en impératifs catégoriques ? Quels sont les mécanismes et la signification de cette obligation de résultats qui est faite aux systèmes scolaires ? Comment la comprendre ? Une manière de débiter est de se demander si le contraire est vrai : l'éducation aurait-elle fonctionné jusqu'ici sans produire des résultats ? Si nous commençons avec les résultats des évaluations, la réponse est clairement non car, depuis au moins le début du siècle dernier, l'évaluation fait partie de façon constitutive du travail scolaire et conditionne, sous différentes formes, l'activité de l'enseignant. Les travaux de docimologie franco-

phones (Piéron, 1963 ; Noizet & Caverni, 1978) témoignent d'un souci constant d'amélioration des pratiques d'évaluation et de nombreuses contributions conceptuelles ont permis, dans les années quatre-vingt et quatre-vingt-dix, à la fois de théoriser le champ de l'évaluation scolaire et d'opérer des distinctions théoriques importantes, notamment entre évaluation formative, formatrice, sommative et pronostique, pour ne mentionner que les plus importantes. Autrement dit, le système éducatif et ses acteurs ont été fortement préoccupés par l'évaluation, ses résultats, leur qualité et leur validité sous différentes formes. Malgré d'importantes réductions

budgétaires, surtout dans les pays occidentaux, ils ont obtenu et utilisé des résultats d'évaluation, ils se sont consolidés, ont accueilli une population croissante, ont élargi l'offre curriculaire et obtenu des taux de certification de fin d'études toujours plus importants durant ces dernières décennies. De ce point de vue, sans vouloir ouvrir le débat relatif à une éventuelle inflation des diplômes (Duru-Bellat, 2005), on peut affirmer que des résultats solaires existent et qu'ils sont plutôt positifs.

Pourquoi alors cette insistance si forte sur l'obligation de résultats en éducation alors que l'évaluation, standardisée ou non, est une pratique constitutive des systèmes éducatifs et a toujours produit des résultats ? Mon hypothèse est que les résultats d'évaluation ont pris une signification différente dès lors qu'on cherche à les utiliser pour qualifier la productivité du système éducatif. Cette nouvelle fonction a émergé ces dernières années sur fond de redéfinition du rôle institutionnel de l'école, confrontée à une logique de marché de plus en plus forte. Cette mutation est intervenue avec la mise en place généralisée des principes de la nouvelle gestion publique qui redéfinit le rôle et les critères d'efficience dans les rapports entre les pouvoirs publics, le marché scolaire et les utilisateurs. Alors que « l'évaluation mesurerait les apprentissages en s'intéressant principalement à l'élève, aujourd'hui son champ d'intervention est devenu beaucoup plus large et met en lien le pédagogique – sa terre d'élection traditionnelle – et le politique dont il est devenu un outil de pilotage » (Behrens, 2006). La réponse à cette question sera donc de chercher dans le rôle politique attribué à l'évaluation, dans les modifications méthodologiques qu'induit ce rôle et dans les effets qu'une évaluation politique produira, notamment dans l'articulation avec son utilisation pédagogique initiale.

Comme l'élargissement de la fonctionnalité évaluative est un processus en cours, je montrerai d'abord les changements intervenus en matière d'évaluation standardisée ces trente-cinq dernières années. Je procéderai ensuite à une analyse succincte des parties constitutives d'un système éducatif, dans la perspective de la nouvelle gestion publique en Europe, et plus précisément dans les pays germanophones, moins connus du lecteur francophone. J'utiliserai de façon ponctuelle les contributions anglo-saxonnes présentées dans ce numéro, à la fois pour étayer mon argumentation, mais aussi pour questionner les concepts et pratiques afférents. Sur la base de mon expérience personnelle tirée des échanges avec les ministères et administrations sco-

lares des cantons de Suisse romande dans leur nouveau *design* de l'architecture scolaire suisse en cours, je discuterai les besoins d'information à l'intérieur d'un tel système pour en modéliser ensuite les flux, en tenant compte à la fois des besoins formatifs et des impératifs de l'*accountability*.

L'ÉVOLUTION DES ÉVALUATIONS STANDARDISÉES EN EUROPE

L'inventaire des évaluations standardisées des élèves en Europe dans l'enseignement obligatoire, réalisé par l'Agence exécutive éducation, audiovisuel et culture (2009a), fait apparaître un tableau très varié de différents tests nationaux. Ils complètent les activités d'évaluation habituelles en classe, sommatives et formatives, qui sont prescrites par des textes législatifs ou des règlements et que les enseignants ou les établissements, dans la plupart des pays, organisent de manière autonome, avec des méthodologies choisies au gré des situations et des buts poursuivis, en fonction de l'avancement des processus d'enseignement et d'apprentissage.

Dans une première phase se situant avant les années quatre-vingt-dix, les tests nationaux et/ou régionaux externes et standardisés, c'est-à-dire basés sur des procédures fixées, permettent de placer les résultats des évaluations en classe dans un contexte plus large et de rendre comparables les performances des élèves à l'intérieur du périmètre couvert par le test. Les formes et les usages développés sont extrêmement variés. Les tests peuvent être obligatoires ou facultatifs et, dans certains cas, les résultats obtenus interviennent dans le processus de certification, en particulier lorsqu'ils sont pris en considération dans le calcul d'une note finale et garantissent ainsi une certaine possibilité de comparaison des diplômes. En règle générale, les tests sont construits à partir des programmes d'enseignement, peuvent couvrir plusieurs disciplines, mais plus particulièrement la langue 1, les mathématiques et les sciences naturelles, ou encore une deuxième langue. Ils s'appliquent à tous les élèves et interviennent surtout à l'interface entre l'enseignement primaire (CTIE 1, d'après la classification internationale type de l'éducation, voir OCDE, 1999) et l'enseignement secondaire inférieur (CTIE 2), ou à la fin de ce dernier. Quant à l'utilisation des résultats, les établissements scolaires concernés pourraient en profiter pour se situer par rapport aux performances nationales ou

pour faire des recommandations pour l'orientation future des élèves. À leur tour, les enseignants pourraient s'en servir pour adapter leur enseignement. Bref, on est en présence d'un traitement d'information à finalité variable qui, au niveau de l'enseignant, correspondrait à des routines de travail qu'Allal (2009) appelle triangulation des informations : l'enseignant pourrait ainsi mettre en perspective des informations tirées des évaluations en classe, des informations obtenues par des évaluations externes (tests standardisés) et d'autres informations provenant de ses observations personnelles. Cependant l'enseignant ne semble pas donner le même poids à ces trois sources d'information lorsqu'il doit se forger un avis global. En effet Gilliéron Giroud et Pulzer-Graf (2009) observent, lors d'une analyse de l'utilisation des résultats de tests externes indicatifs par les enseignants au début du cycle primaire (CTIE 1) en Suisse romande dans le canton de Vaud, que ceux-ci n'utilisent ces résultats que lorsqu'ils confirment les appréciations obtenues par les évaluations de l'enseignant. Ce qui voudrait dire que, selon le contexte, soit l'enseignant ne saisit pas la portée des résultats tirés des évaluations externes, soit les trois sources n'ont pas la même valeur d'information dans la constitution de son jugement.

À partir des années quatre-vingt-dix, on observe une expansion des tests nationaux et, dans certains cas, une réorientation des tests existants. Ce sont les systèmes qui, au-delà de l'identification des performances individuelles à des fins d'orientation ou de rééducation lors des apprentissages, cherchent à utiliser ces dispositifs également à des fins de pilotage, en évaluant les établissements et, à travers eux, le système éducatif dans son ensemble. En adoptant un point de vue macroscopique, les résultats obtenus semblent indiquer que les phénomènes éducatifs peuvent entrer en tension avec ce que ressentent les acteurs (enseignants, parents, élèves) à travers les évaluations *internes* de ces mêmes phénomènes. En effet « dans les tests nationaux la préoccupation pédagogique et didactique se trouve remplacée par un regard sociologique globalisant qui opère à un niveau agrégé, sur des moyennes, des dispersions, des corrélations, abstractions – données qui sont difficilement accessibles aux acteurs directement concernés et laissent souvent les enseignants incrédules ou, pour le moins, dubitatifs » (Crahay, 2009, p. 239 *sqq.*).

Comme le soulignent plusieurs auteurs (Behrens, 2010 ; Maroy, 2011 ; Mons, 2009), ce changement est en relation directe avec l'adoption des principes de la

nouvelle gestion publique par la plupart des administrations du monde de l'éducation. Cette idéologie de gestion publique a été décrite largement ailleurs et, dans cet article, nous n'y reviendrons pas au-delà d'une présentation très schématique nous permettant de désigner les acteurs concernés. En bref, elle part du principe que les unités administratives confrontées directement aux utilisateurs sont les agents les mieux à même de trouver des réponses adéquates et d'assurer l'efficacité de leur action. Dans le domaine scolaire, ces unités sont les établissements scolaires, composés d'enseignants et regroupés sous une direction. Cela rend nécessaire une autonomie d'action que les bureaucraties scolaires peinent à accorder. En règle générale, le cadre d'exercice de cette autonomie est donné par un contrat de prestation entre l'administration centrale et l'agent, qui définit les performances attendues, un budget calculé en conséquence, l'obligation d'un dispositif d'assurance de qualité garantissant le maintien des prestations et le devoir de rendre compte de manière scientifique des résultats obtenus.

La demande pour améliorer la qualité des systèmes est donc inhérente à la logique de la nouvelle gestion publique. Plus précisément, à l'instar des dispositifs d'assurance qualité dans l'économie ne servant qu'à attester l'existence de procédures gestionnaires permettant de maîtriser le processus de production et non pas la qualité du produit lui-même, l'usage politique qui est fait de la qualité en éducation sous-entend la « non-qualité » des systèmes éducatifs (Behrens, 2007). Par là-même, l'argument de la qualité est utilisé pour lancer le débat sur l'efficacité générale des systèmes éducatifs dans une période de restrictions budgétaires généralisée, ce qui conduit à une réflexion sur les missions essentielles de l'école et à charger les personnels de trouver eux-mêmes les solutions permettant de faire plus avec moins. Pendant la même période, le GATT (*General Agreement on Tariffs and Trade*) impose de considérer l'éducation comme une activité économique de service. Il exprime un courant de pensée puissant cherchant à abolir le monopole étatique en matière d'éducation et à ouvrir l'école publique au marché. Il aurait conduit, selon Berliner et Biddle (1995) et Bracey (2002), à l'instrumentalisation tactique de l'argument de la qualité, non pas pour améliorer l'école publique mais pour la déstabiliser. Bracey cité par Lind (2009, p. 18) aurait même parlé d'une guerre contre l'école publique, vision qui a récemment été confirmée par Neumann, membre du cabinet Bush qui, dans une interview avec le journal *Times* (2008), estimait que certains considéraient l'initiative *No*

Child Left Behind comme un cheval de Troie conduisant l'opinion publique à choisir une école de marché plutôt qu'une école publique : « *A way to expose the failure of public education and "blow it a bit". There was a number of people pushing hard for market forces and privatization.* »

Ce cadre influence à des degrés divers la plupart des administrations et conduit à travers les différents pays de l'Europe à un large mouvement de décentralisation et d'autonomie des établissements. Cette tendance est accompagnée d'un changement réglementaire ou législatif marqué par une liberté d'action accrue des établissements, mais aussi d'une intensification des évaluations (Agence exécutive éducation, audiovisuel et culture, 2009a, p. 20). S'adaptant au contexte politico-historique de chaque pays, les dispositifs prennent une multitude de formes qui combinent diverses fonctions : évaluations externes surtout à des fins de pilotage du système ; orientation des élèves lors des transitions (du secondaire inférieur vers le supérieur professionnel ou académique) ; comparaisons nationale ou régionale des performances des élèves obtenues lors d'évaluations internes réalisées par les enseignants ; régulation des processus d'enseignement et d'apprentissage. Ces évaluations internes ne sont généralement pas standardisées et permettent la régulation des processus d'apprentissage, raison pour laquelle Mottier Lopez et Laveault (2008) et Laveault (2009) suggèrent de les considérer comme ayant une finalité formative, peu importe s'il s'agit d'évaluations formatives ou sommatives. Ceci annulerait la distinction instrumentale entre formatif et sommatif décrite par Allal et Mottier Lopez (2005) en faveur d'une opposition entre une évaluation pour la surveillance du système et une évaluation pour le suivi des processus d'apprentissage.

Le débat sur la qualité des systèmes éducatifs est parallèlement renforcé par l'émergence, dans les années soixante, des premières enquêtes internationales sous l'égide de l'IEA (International Association for the Evaluation of Educational Achievement), dont Gaston Mialaret était un membre fondateur. C'est ainsi qu'un consortium de différents centres de recherche à l'échelle planétaire a étudié, dans l'étude pilote *Twelve Country Study*, la faisabilité d'études internationales comparant la réussite scolaire d'élèves de 13 ans dans douze pays, travaux qui se poursuivent jusqu'à nos jours dans le cadre des enquêtes internationales TIMSS (*Trends in International Mathematics and Science Study*) et PIRLS (*Progress in International Reading Literacy Study*). La démarche développée par l'IEA est reprise par l'OCDE dans le cadre des

enquêtes PISA (*Program for International Student Assessment*) à partir de 2000, qui établit de facto un *ranking*, voire une compétition, entre les pays, tout en se plaçant ainsi comme autorité économique transnationale dans la construction d'un champ éducatif global, qu'elle justifie par l'idéologie de l'économie de la connaissance (Lingard & Rawolle, 2010). Dans ce processus de mondialisation, l'école obligatoire se voit imposer la fonction d'assurer l'acquisition de compétences dont l'OCDE (OCDE, 2011), à travers PISA, estime que les jeunes de 15 ans auront besoin dans la vie quotidienne au terme de la scolarité obligatoire : « Plutôt que la maîtrise d'un programme scolaire précis, PISA teste l'aptitude des élèves à appliquer les connaissances acquises à l'école aux situations de la vie réelle. Les facteurs conditionnant leurs performances ainsi que leur potentiel pour l'apprentissage tout au long de la vie font également l'objet d'une analyse. » (OCDE, 2011)

La même logique est en œuvre au niveau de l'Union européenne. Dans un premier temps, l'intérêt pour l'éducation s'est développé lentement au sein de la Communauté européenne avec un premier discours sur les orientations de base, les collaborations entre États, la gouvernance, les réseaux et les projets de coopération économique et culturelle. Il change en 1971 lorsque les ministères mettent en œuvre une politique de coopération européenne en éducation. Les références à l'éducation sont faites plus particulièrement dans les domaines de la formation professionnelle et de l'enseignement supérieur (Lawn & Nóvoa, 2005). Vers la fin de cette décennie, des initiatives concernant la reconnaissance des diplômes, la mobilité des étudiants et surtout l'harmonisation des statistiques sont mises en œuvre. Mais le grand pas vers l'espace européen de l'éducation intervient en 1992 avec le traité de Maastricht, qui consacre l'éducation comme l'une des responsabilités majeures de l'Union européenne. Dès lors, il est question d'une contribution communautaire pour une éducation de qualité, ce qui ouvre la porte à une action de l'Union, tout en respectant encore la responsabilité des États membres selon les principes de la diversité et de la subsidiarité. Le *Livre blanc sur l'éducation et la formation* (Commission européenne, 1996) est l'un des premiers documents qui proposent la création d'un processus européen permettant de confronter et de diffuser largement les définitions des « compétences clés » et de trouver les meilleurs moyens de les acquérir, de les évaluer et de les certifier (Meunier, 2005). Ce sont donc, comme le soulignent Lawn et Nóvoa (2005), les questions sociétales, celles de

l'emploi et de la qualité qui préparent l'émergence d'un véritable espace européen de l'éducation. Avec l'objectif ambitieux de « l'économie de la connaissance, la plus compétitive et la plus dynamique du monde » formulé au sommet de Lisbonne en 2000 (Commission des communautés européennes, 2003), l'Union se donne implicitement le mandat pour développer une approche commune dans le domaine de l'éducation.

Depuis, l'Union européenne se positionne comme acteur supranational et fixe ses actions politiques d'abord dans le programme *Éducation et formation 2010*, et plus récemment dans l'édition de 2020 (Conseil de l'Union européenne, 2009), qui prévoit de nombreuses actions politiques qu'il serait trop long de détailler ici. En ce qui concerne la scolarité obligatoire, on peut retenir les objectifs suivants :

- faire en sorte que l'éducation, la formation tout au long de la vie et la mobilité deviennent des réalités ;
- améliorer la qualité et l'efficacité des systèmes d'éducation et de formation ;
- promouvoir l'égalité, la cohésion sociale et la citoyenneté active ;
- mettre en valeur la créativité et l'innovation, en particulier l'entrepreneuriat, et ce à tous les niveaux d'enseignement.

Pour mesurer si ces objectifs sont atteints, elle se fixe pour 2020 des cibles qui concernent directement l'enseignement obligatoire :

- au moins 95 % des enfants entre 4 ans et l'âge du début de l'enseignement primaire obligatoire devraient participer à l'enseignement préscolaire ;
- le taux moyen de jeunes de 15 ans ayant des lacunes en lecture, en mathématiques et en science devrait être inférieur à 15 % ;
- le taux moyen de décrochage scolaire devrait être inférieur à 10 % ;
- le taux moyen d'adultes entre 30 et 34 ans ayant un niveau de troisième cycle devrait atteindre au moins 40 % ;
- une moyenne d'environ 15 % des adultes (de 25 à 64 ans) devrait participer à l'éducation et la formation tout au long de la vie.

Ce sont des objectifs macroscopiques qui s'adressent au système. Ils comprennent des taux de scolarisation, de poursuite et d'interruption de formation et des indicateurs de performance des élèves, vraisemblablement calqués sur l'enquête PISA. Autrement dit, dans cette économie de la connaissance, l'institution scolaire est appelée à jouer un rôle important. Elle est exposée à des attentes sociétales larges dans

lesquelles il est moins question du développement de l'individu que de la capacité de la société à se maintenir dans une compétition globale. Apparaît donc clairement une obligation de résultats. Elle est destinée au système dans son ensemble et répercutée partiellement sur les apprenants.

LES RÉFÉRENTIELS ÉDUCATIFS

On observe alors dans les pays francophones européens l'émergence de référentiels avec l'indication des performances à atteindre. Dans la communauté française de Belgique d'abord, apparaissent en 1997 par décret les socles de compétences qui décrivent l'ensemble des compétences devant être maîtrisées. Ils sont très proches des plans d'études mais, au lieu de décrire comment structurer l'enseignement, ils définissent de façon exhaustive et sous forme de performances vérifiables ce qui doit être atteint à certains moments de la scolarité. Ils concernent les compétences pour les huit premières années de l'enseignement obligatoire (les six années du primaire et les deux premières années du secondaire) et les compétences terminales, un référentiel sous forme de compétences qui doivent être acquises à la fin du secondaire dans chaque discipline scolaire. Le but est de cadrer, voire de diminuer les grandes disparités disciplinaires provenant d'une situation de marché scolaire sur lequel prestataires publics et privés se livrent une concurrence, en vertu des dispositions constitutionnelles de la liberté d'enseignement, qui autorise toute personne à ouvrir une école et donne aux parents le libre choix de l'école qui sera fréquentée par leurs enfants (Rey & Carette, 2005). Cette grande disparité est tenue pour responsable des résultats PISA, peu satisfaisants.

En France, ce sont des réflexions similaires qui démarrent quelques années plus tard. Mais, au lieu de procéder par décret, les autorités lancent un débat national sur l'avenir de l'école en 2003. La préoccupation est de définir et représenter ce qui est indispensable pour réussir sa vie dans une société du *xxi^e* siècle. Selon Thélot, le message principal qui se dégage est que l'école fasse vraiment réussir tous les élèves. Pour ce faire, l'école a une mission éducative, elle assure la maîtrise d'un « socle commun des indispensables et elle organise la diversité des parcours » (Thélot, 2004, p. 31), tout en s'efforçant d'être juste. Ces indispensables forment un référentiel appelé « socle commun ». Il est choisi en fonction de valeurs

et se décline en termes de connaissances, de compétences et de règles de comportement. Il ne s'identifie pas aux programmes mais il définit les compétences sociales qui doivent être maîtrisées pour être intégrées à la nation. C'est une formulation vague qui, d'après Meunier (2005), n'a pas pour objectif de préparer à l'emploi. La loi d'orientation et de programme pour l'avenir de l'école du 24 avril 2005 en émane. Officiellement cette loi « assigne au système éducatif des missions renouvelées autour d'un objectif central : assurer la réussite de tous les élèves » (Meunier, 2005, p. 80).

Les deux exemples francophones ont été abondamment débattus (voir par exemple Gauthier & Le Gouvello, 2010 ; Maroy & Mangez, 2011), nous ne les explorerons donc pas davantage. L'intérêt pour notre propos est qu'ils témoignent d'une préoccupation forte de redéfinir la mission de l'école : après l'effort de massification de l'enseignement secondaire dans des contextes scolaires marqués par une situation de marché, l'école est appelée à se préoccuper de l'efficacité et de l'efficience de son action, mesurées à l'aune de compétences que tous les élèves doivent acquérir. Les deux exemples proposent donc des référentiels de performance des élèves qui sont déclinés en compétences. Dans le cas de la Belgique francophone, ces compétences sont considérées comme une base pour les plans d'études. En France, elles se déclinent de manière plus abstraite en fonction d'un débat sociétal et se placent ainsi volontairement en dehors des programmes scolaires. Mais, dans les deux cas, le référentiel précède la construction des dispositifs d'évaluation ce qui, dans le cas de la communauté francophone belge, posera ultérieurement le problème de la vérification et de la traçabilité évaluative des compétences retenues (Rey, 1996 ; Rey & Carette, 2005 ; Malaise, 2011). Une autre particularité importante de ces référentiels de performance est que les exigences ainsi fixées doivent être atteintes par tous les élèves.

LES STANDARDS ET LES DISPOSITIFS D'ASSURANCE QUALITÉ DANS LES SYSTÈMES ÉDUCATIFS GERMANOPHONES

Les pays germanophones étudiés (nous y comptons la Suisse pour l'influence germanophone et la parenté structurale) s'inspirent du concept anglophone des standards dans la construction de leurs référentiels. Ils ont tous les trois une structure fédérale dans laquelle

le rapport entre centre et instances décisionnelles régionales (*Länder*, cantons, régions) peut varier fortement.

L'Allemagne

L'Allemagne (82,2 millions d'habitants) est un pays fédéraliste avec un gouvernement central (*Bundesregierung*) et un gouvernement pour chacun des 16 pays (*Bundesländer* ou *Länder*) faisant partie de la confédération. La compétence décisionnelle en matière d'éducation est l'affaire des *Länder*. Le ministère du gouvernement fédéral n'intervient pas dans le domaine de la scolarité obligatoire. Les ministres d'éducation des différents *Länder* concertent leurs actions au sein de la conférence des ministres d'éducation (KMK ou *Kultusministerkonferenz*). Celle-ci décide en 1997 de soumettre les systèmes scolaires à une analyse comparative internationale dans le cadre des enquêtes internationales TIMSS, PISA et PIRLS, afin de pouvoir identifier les forces et faiblesses des élèves allemands dans quelques domaines de compétence jugés centraux. En 1999, le gouvernement central et les *Länder* lancent conjointement un forum de discussion sur l'éducation (*Forum Bildung*). Cinq thématiques se trouvent au centre des débats : les finalités éducatives de demain, l'égalité des chances, l'assurance qualité dans la compétition internationale, l'apprentissage tout au long de la vie et une nouvelle culture d'enseignement et d'apprentissage. Le forum recommande une plus grande autonomie des établissements, la nécessité d'informer régulièrement sur l'état de l'éducation et de prendre des mesures pour un développement systématique de la qualité de l'enseignement (Oelkers & Reusser, 2008). En 2002, la KMK utilise les résultats obtenus dans les comparaisons internationales pour justifier un passage vers un suivi régulier des résultats appelé « monitoring », la création de standards et une intensification des évaluations qui, selon les expériences faites aux États-Unis, suffirait pour obtenir de meilleurs résultats – conviction du « *try harder* », effectivement défendue dans la première génération des systèmes d'*accountability* sur la base de standards (voir Robinson, dans ce numéro).

Les données obtenues par des tests vérifiant les standards sont intégrées dans un rapport sur l'éducation (*Bildungsbericht*), inspiré des rapports *Regards sur l'éducation 2009 : les indicateurs de l'OCDE* (OCDE, 2009) et *Chiffres clés de l'éducation 2009 : une vue d'ensemble sur les systèmes éducatifs européens* de l'Agence exécutive éducation, audiovisuel et culture (2009b) de l'Union Européenne, ou des

publications de la DEPP en France sur *L'état de l'école* (MENJV, 2010). Le but est d'informer le grand public à intervalle régulier sur la capacité de l'institution scolaire à remplir sa mission et à répondre aux exigences que la société lui fixe (Rürup, 2003), et de permettre de cette manière de mettre en place un dispositif de monitoring du système scolaire. Le rapport allemand paraît pour la première fois en 2003 et confirme la nécessité de fixer des objectifs et des standards communs exigeants et de vérifier régulièrement qu'ils sont atteints (KMK, 2004). Par ailleurs, « la KMK est appelée à mettre à disposition des ressources suffisantes et à les utiliser de manière optimale. Le but est de préserver les chances éducatives [*Bildungschancen*] de tous et de toutes, en marquant l'enseignement par un enseignement basé sur l'individualisation » (KMK, 2004, p. 5, traduit par l'auteur). Autrement dit, la KMK est invitée non seulement à contrôler la bonne marche du système, mais en plus à le réformer par une approche pédagogique innovante pour la réussite de laquelle elle doit mettre des ressources suffisantes à disposition.

Dans la même année, le ministère fédéral de l'éducation publie la fameuse expertise Klieme, Avenarius, Blum *et alii* (2004) sur le développement des standards nationaux de formation, qui structurera fortement le débat sur les standards, en Allemagne comme en Suisse. Le rapport plaide pour des standards basés sur des modèles de compétences conceptualisés à partir de considérations théoriques d'ordre didactique et psychopédagogique. Ces modèles doivent décrire les dimensions de compétences jugées centrales, ce qui peut être attendu, les niveaux de performance et comment les évaluer. Ce ne sera que sur la base de tels travaux validés empiriquement qu'il sera possible de formuler des standards. L'expertise de Klieme, Avenarius, Blum *et alii* (2004) recommande que soient élaborés et mis en œuvre des standards de base (*Mindeststandards*) qui définissent le seuil inférieur des performances à atteindre. Mais en l'absence de tels travaux, la KMK opte pour des standards moyens, autrement dit sans précision plus détaillée des niveaux de performance (Oelkers & Reusser, 2008). Proches des plans d'études et compris comme permettant de préciser ces derniers, elle les adopte pour la fin de la scolarité obligatoire en langue 1, mathématiques et langue 2. Elle les complète en 2004 par des standards pour la 4^e année primaire et les élargit aux sciences naturelles (Eurydice, 2011a). Toujours en 2004, elle crée l'institut de recherche IQB (Institut für Qualität der Bildung), rattaché à l'Université Humboldt de Berlin, qui est mandaté pour développer les modèles de compé-

tences, les valider empiriquement et procéder ainsi à la vérification des standards au niveau national.

Depuis sa fondation, l'IQB développe les instruments de test en produisant des items sous forme de tâches. Dans une première étude appelée *Pilotierungsstudie*, elle étudie la validité de contenu des items par rapport aux compétences retenues comme standards. Une deuxième étude permet de classer les items retenus sur une courbe de difficulté et de fixer des niveaux de compétence. La méthodologie appliquée ressemble à celle de PISA. Les vérifications des standards ont commencé en 2008. Ils remplacent les échantillons complémentaires de PISA pour chaque *Land*. En parallèle, l'IQB étudie comment les standards sont intégrés dans l'enseignement. Les problématiques abordées sont l'analyse de l'intégration des standards dans l'enseignement, ainsi que la gestion des innovations dans la pratique enseignante. Autrement dit, la conception des standards en termes de compétences n'est pas considérée comme un artefact de mesure emprunté à la méthodologie PISA, mais comme une innovation pédagogique majeure. De ce fait, la plupart des *Länder* sont en train de prendre des mesures importantes pour adapter les plans d'études à une pédagogie par compétences ou établir un « tronc commun » (*Kerncurriculum*) intégrant les compétences formulées dans les standards. D'autres mesures sont prévues pour la formation des enseignants et le développement de procédures de qualité dans les établissements (Oelkers & Reusser, 2008, p. 453).

Mentionnons également que, dans le cadre du passage vers un monitoring par les résultats, les établissements scolaires gagnent en autonomie financière et administrative, autonomie qui est délimitée par un mandat de prestation fixé avec l'autorité de tutelle. Ils jouissent ainsi d'une certaine autonomie dans le domaine pédagogique, leur permettant d'apporter des réponses appropriées aux exigences du contexte local dans lequel ils évoluent. Dans ce sens, les établissements sont responsables d'assurer la qualité de l'enseignement et le suivi du programme (Eurydice, 2011a, p. 292 *sqq.*).

L'Autriche

L'Autriche (8,2 millions d'habitants) est également un pays fédéraliste, mais la compétence décisionnelle pour la scolarité obligatoire est répartie entre le gouvernement fédéral et les neuf *Länder*. Le gouvernement central joue un rôle bien plus important que les *Länder*, qui souvent sont uniquement en charge

des règlements d'application. Il édicte des lois cadres, définit les programmes cadres qu'il négocie avec les *Länder* et homologue des moyens d'enseignement. Les *Länder* engagent les enseignants, choisissent parmi les moyens d'enseignement homologués et soutiennent les communes dans le financement de l'infrastructure scolaire. Les établissements jouissent d'une certaine autonomie budgétaire et ont le droit d'adapter les curriculums aux particularités locales.

En Autriche, les résultats de PISA et de PIRLS en littérature interrogent comme ailleurs. Cependant ils se situent en 2000 clairement au-dessus de la moyenne de l'OCDE et sont meilleurs que ceux de l'Allemagne et de la Suisse. En revanche, d'autres indicateurs comme le niveau de compétence des jeunes à la fin de la scolarité obligatoire et l'iniquité du système d'évaluation dans la transition vers le secondaire soulèvent des questions. Ces éléments conduisent en 2003 à la création d'une commission (*Zukunftskommission* ou commission d'avenir) qui est appelée à développer un concept global de réformes, en particulier des propositions pour la mise en place d'un système national de management de la qualité, des propositions pour l'amélioration de l'enseignement et des propositions pour le renforcement du mandat pédagogique des enseignants. La commission propose toute une série de mesures, dont la mise en œuvre de standards de base. Des travaux sont entrepris et conduisent en 2008 à la révision de la loi scolaire, qui donne au ministère fédéral la possibilité de décréter des standards si cela est jugé nécessaire. Ce sera fait en 2009. L'argumentation ministérielle justifiant leur introduction évoque la nécessité d'une politique basée sur les résultats (*Evidenzorientierte Bildungspolitik*) et l'alignement sur un nouveau paradigme du monitoring par les résultats (*Outputorientierung*), qui serait en voie de généralisation dans les pays germanophones (Breit, Friedl-Lucyshyn, Furlan *et al.*, 2010). Le but premier des standards est d'assurer l'acquisition des compétences fondamentales (*Grundkompetenzen*) que chaque élève doit atteindre. Ce cadre donne un statut légal à des travaux pilotes qui, depuis 2001, analysent le curriculum afin d'identifier des domaines de compétences obligatoires (*Pflichtgegenstände*) et de définir les performances qui doivent être attendues à certains moments précis du parcours scolaire. Les équipes de projet comprennent dans une première phase des enseignants et des inspecteurs, mais quasiment pas de chercheurs. Leurs propositions ne sont que peu utilisables, notamment parce qu'elles ne parviennent pas à montrer l'utilisation des standards dans l'enseignement. À l'instar de l'expertise de Klieme, des cher-

cheurs sont ensuite associés aux travaux, ce qui permet l'adoption des premières propositions testées dans 320 établissements (Oelkers & Reusser, 2008). Les standards sont développés en langue 1 et en mathématiques pour la 4^e et la 8^e année scolaire et en langue 2 pour la 8^e année. Une première vérification est prévue dans l'année 2011-2012 (Breit, Friedl-Lucyshyn, Furlan *et al.*, 2010).

En 2008 est également créé l'institut fédéral BIFIE (Bildungsforschung Innovation und Entwicklung für das österreichische Schulwesen, l'Institut national pour la recherche en éducation, l'innovation et le développement du système scolaire autrichien). Son rôle est de participer au développement des modèles de compétences et des descripteurs permettant de décrire et de mesurer les standards par des tâches dont on estime que la réussite permet d'inférer la maîtrise des standards ; de développer des exemples de tâches pour l'enseignement ; de vérifier que les standards sont bien atteints ; enfin d'analyser les résultats, d'informer les élèves, enseignants, directions d'établissements et administrations scolaires, ainsi que de conseiller les décideurs.

La vérification que les standards sont bien atteints sera effectuée auprès de tous les élèves de l'enseignement public et de l'enseignement privé, mais les résultats ne pourront être utilisés pour l'évaluation individuelle. Ils sont destinés en premier lieu aux enseignants dans la perspective d'un *feedback* sur leur enseignement qui, dans la même lignée, devra désormais appliquer les principes d'une pédagogie par compétences (Eurydice, 2010, p. 14). De ce fait, la vérification ne porte pas seulement sur les résultats d'élèves, mais aussi sur les conditions cadres de l'enseignement qui sont relevées par un questionnaire contextuel. De même, les directions d'établissement recevront les résultats afin de pouvoir analyser les forces et faiblesses de leur établissement et de pouvoir prendre les mesures de soutien appropriées. Le corps d'inspection recevra quant à lui des données agrégées des établissements dont il a la surveillance. Enfin les résultats provenant de la vérification des standards seront analysés dans une macroperspective et seront intégrés dans un rapport sur l'éducation. Ce rapport a servi à piloter le système pour la première fois en 2009 (Eurydice, 2011b).

La Suisse

Avec 7,6 millions d'habitants, la Suisse est le plus petit et le plus décentralisé des trois pays germanophones étudiés. Également de structure fédérale, elle

est composée de 26 cantons qui sont souverains en matière d'éducation. Le gouvernement fédéral n'a pas de politique d'éducation en matière de scolarité obligatoire. Par contre, les cantons coordonnent aujourd'hui leur travail au sein d'une conférence des ministres appelée CDIP (Conférence suisse des directeurs cantonaux d'instruction publique). La CDIP se compose de deux conférences régionales qui regroupent les cantons selon leur appartenance linguistique (une conférence alémanique et une conférence latine représentant la Suisse francophone et italophone).

En 2001, la CDIP se donne des lignes directrices qui circonscrivent les objectifs de la coordination et les modalités de coopération dans le domaine de la scolarité obligatoire. Deux objectifs importants sont retenus pour l'assurance qualité du système éducatif : l'harmonisation de l'école obligatoire par la fixation de standards de performance et la mise en place d'un monitoring de l'éducation en collaboration avec la Confédération. Mais ce sera la première enquête PISA, dont les résultats sont en dessous des attentes, qui créera le consensus nécessaire à une réforme majeure du système. En 2002, la CDIP propose un plan d'action visant l'harmonisation structurelle des systèmes cantonaux par le biais de standards de formation. Cherchant une uniformisation des curriculums cantonaux par l'intermédiaire de la définition de performances minimales que tous les élèves devront atteindre, elle passe à un monitoring par les résultats et utilise son initiative pour contrer des velléités de transférer à la Confédération la responsabilité des compétences en matière d'éducation (Behrens, 2008).

En effet, quelques années plus tard, le Parlement fédéral, également appelé à se prononcer sur la situation de la formation en Suisse, soumet au vote deux articles constitutionnels sur ce sujet. Le 21 mai 2006, ils sont acceptés par le peuple suisse, à une majorité très nette de 86 % des votants, de même que par tous les cantons, ce qui initiera un bouleversement profond du système suisse de la formation. Désormais les autorités publiques responsables de la formation sont tenues par la Constitution fédérale de réglementer de manière uniforme certains paramètres fondamentaux du système éducatif. Si les cantons ne devaient pas parvenir à trouver une solution commune pour la scolarité obligatoire, la Confédération leur en imposerait une. Mise sous pression pendant les débats parlementaires préparatoires, la CDIP se devait donc de présenter un projet capable de modifier les structures cantonales, faute de quoi elle

risquait de perdre sa crédibilité dans ce processus d'ajustement.

Les mesures d'harmonisation, appelées HarmoS, mettent donc en œuvre une vaste réforme des structures éducatives cantonales. Le projet est politique puisqu'il propose un nouveau cadre légal sous forme d'un concordat entre cantons, dans lequel ces derniers renoncent à une partie de leur souveraineté en matière éducative en faveur d'un niveau national. Ce texte de loi est adopté par 15 des 26 cantons (totalisant 76,8 % de la population) et entre en vigueur le 1^{er} août 2009. Il comprend plusieurs dimensions concernant les principaux paramètres du système, à savoir les finalités de la formation de base, les domaines disciplinaires, l'enseignement des langues étrangères, le début de la scolarisation, la durée des degrés (cycles) scolaires, les instruments d'assurance qualité et l'aménagement de la journée scolaire. Sous le label d'instruments d'assurance qualité, on trouve l'idée des standards de performance qui, en adoptant les recommandations de l'expertise Klieme, seront inférés de modèles de compétences à développer et à valider scientifiquement pour chaque discipline. Le texte de loi affirme clairement la primauté des standards sur les curriculums. C'est donc d'abord une réponse pragmatique et inédite à un paysage éducatif qui n'a pas réussi à harmoniser au niveau national les contenus et les objectifs de formation. L'autre élément fort du concordat est la création de plans d'études au niveau régional, alignés sur les standards. Les performances de l'école obligatoire seront régulièrement évaluées dans le cadre d'un monitoring national du système éducatif. La vérification de l'atteinte des standards nationaux de formation par des tests fera partie intégrante de la récolte des données à des fins de monitoring (CDIP, 2007, articles 7, 8 et 10 du concordat HarmoS).

Les modèles de compétences sont élaborés par des consortiums de recherche spécialement créés à cet effet. Ils s'inspirent du cadre européen commun de référence pour les langues (Division des politiques linguistiques, 2000) et l'appliquent de façon analogue pour la langue 1, les mathématiques, la langue 2 et les sciences naturelles. Le mandat initial met clairement en avant la dimension de la mesure (Behrens, 2008), bien que celle-ci soit destinée à s'effacer partiellement en faveur de l'harmonisation curriculaire qui est privilégiée dans une deuxième phase (Behrens, 2010). De ce fait, les modèles de compétences ne sont pas clairement construits dans une visée didactique ou psychopédagogique. Ils sont davantage une fixation normative de quelques balises

curriculaires permettant d'harmoniser les contenus ou les objectifs d'enseignement dans les plans d'études à développer.

De même, le projet scientifique d'élaboration des standards HarmoS n'a pas fourni un cadre suffisant pour valider empiriquement les modèles de compétences. Si les équipes de recherche ont pu détailler chaque discipline en sous-domaines et déterminer des aspects cognitifs, elles n'ont pas réussi à se mettre d'accord sur une approche cognitivo-didactique commune. De ce fait, il leur a été difficile de choisir des tâches dont la réussite permettrait d'inférer la maîtrise d'une compétence contenue dans le modèle. En particulier en langue 1, certaines parties du modèle restent à réviser. Le test national dit de validation, portant sur 12 000 élèves de deux âges différents, n'a permis que de sélectionner les items satisfaisants d'un point de vue psychométrique et de les hiérarchiser sur une courbe de difficulté progressive. Le nombre d'items étant partiellement insuffisants, notamment pour obtenir une invariance linguistique due au caractère multilingue du pays, ni le lien entre items et compétences, ni la fixation de niveaux de performance n'ont pu être solidement établis. Ces faiblesses seront à prendre en compte lors de la première vérification des standards prévue pour 2014. Pour le moment, la manière dont les cantons entendent procéder n'est pas encore claire. Étant donné le coût d'une telle opération, les cantons hésitent à créer une institution chargée de ce travail. Différentes options sont étudiées. Il est prévu d'effectuer la vérification des standards à l'aide de différents échantillons, ce qui permettra de comparer les performances des cantons. Par ailleurs, la CDIP a l'intention de faire remplacer les échantillons complémentaires de PISA de chaque canton par les épreuves basées sur les standards. En prenant l'option de travailler par échantillon, la CDIP limite l'utilisation qui peut être faite des standards à l'évaluation du système, ce qui lui permet de déclarer que les standards ne seront pas utilisés ni pour l'évaluation des établissements, ni pour celle des enseignants. En outre, bien que les standards doivent être formulés de façon claire et accessible, ils ne sont pas directement destinés aux enseignants et à leur travail en classe.

Il est également important de mentionner que le concordat HarmoS redéfinit les principes de subsidiarité entre le niveau national, les régions et les cantons. Ainsi il impose une architecture scolaire dans laquelle différents éléments de monitoring scolaire doivent être alignés sur les standards : les cantons gèrent de façon indépendante les disposi-

tifs d'évaluation des établissements scolaires et déterminent le degré d'autonomie qu'ils entendent leur accorder. La région est responsable de l'harmonisation des plans d'études et de la coordination des moyens d'enseignement. En plus, les cantons francophones de la Suisse romande choisissent d'introduire des épreuves de référence pour tous les élèves. Cette initiative est l'aboutissement d'une longue réflexion sur l'évaluation scolaire (Behrens, 2010) et elle est reprise dans le cadre des mesures accompagnant l'introduction du plan d'études romand et intégré dans la convention scolaire romande, texte de loi intercantonal dans lequel les cantons romands précisent l'application qu'ils entendent faire du concordat HarmoS, tout en le complétant par quelques mesures supplémentaires. Contrairement aux tests de vérification des standards, les épreuves romandes prennent le plan d'études comme référentiel. L'objectif est de vérifier que chaque élève atteigne bien les objectifs prévus.

En parallèle, la Confédération et la CDIP mettent en place un dispositif de monitoring tel que demandé par le nouvel article constitutionnel. Le pivot central est une compilation de données statistiques, administratives et de recherche, organisées selon les cycles d'enseignement et qu'elles traitent sous l'angle de l'efficacité, de l'efficience et de l'équité. Le rapport sur l'éducation contiendra les données relatives à la vérification des standards. Il est renouvelé tous les 4 ans. Jusqu'à ce jour, deux rapports ont paru, en 2006 (dans une version pilote) et en 2010 (Centre suisse de coordination pour la recherche en éducation, 2010).

FONCTIONNALITÉS, CARACTÉRISTIQUES ET USAGES DES DISPOSITIFS

L'exemple des trois pays germanophones présentés est intéressant puisqu'on assiste à trois tentatives de création d'un référentiel central critérié, permettant d'ancrer les performances et de les comparer. Les standards sont donc l'élément central dans un dispositif piloté par les résultats. Les expériences faites aux États-Unis avec les premières générations de standards, en particulier l'écart entre standards, curriculums et tests (Resnick & Zurawsky, 2005), ont été intégrées, et on se trouve ici devant une tentative de construire un système cohérent globalement. On observe en effet, du moins au niveau du discours ministériel, l'exigence que les standards soient ali-

gnés avec le curriculum et même dans certains cas avec les moyens d'enseignement et d'autres dispositifs d'évaluation individuelle. La genèse des standards dans les trois pays suit deux modèles : ils sont développés à partir d'un curriculum existant (comme en Autriche) ou bien ils se situent en dehors de ceux des *Länder* ou des cantons (comme en Allemagne et en Suisse) et précèdent des ajustements curriculaires ou un nouveau curriculum à créer. Les standards se déclinent en compétences, soit pour des raisons de mesure, soit parce qu'il y a une ambition de réformer l'enseignement vers une pédagogie par compétences. Dans les deux cas, ils fixent des seuils ou niveaux de performance pour un nombre limité de compétences par discipline : soit un seuil inférieur que tous les élèves doivent atteindre (standards minimaux ou de base) et qui a l'avantage d'être potentiellement une cible à atteindre (Baslé, 2008, p. 94), soit un seuil moyen (standards réguliers), soit un niveau d'excellence (standards maximaux). Ils deviennent ainsi prescription mesurable pour tous les acteurs du système.

Les trois pays s'inspirent à un moment ou un autre de l'expertise Klieme, Avenarius, Blum *et alii* (2004) et en suivent, à des degrés divers, les recommandations. Celle-ci apporte d'abord une clarification conceptuelle en distinguant les standards de contenu (*content standards*), proches des programmes ; les standards de performance (*performance standards*), qui précisent le niveau de performance de l'élève ; et les standards définissant le cadre nécessaire pour que les apprentissages puissent avoir lieu (*opportunity to learn standards*). Dans les trois pays, les types de standards retenus sont les standards de performance. Ensuite, cette expertise demande que les standards soient inférés de modèles de compétences construits selon des principes didactiques et scientifiquement validés. Cette exigence traduit bien la préoccupation d'obtenir des standards à la fois valides pour l'analyse de la performance des systèmes scolaires et significatifs pour les pratiques en classe. Autrement dit, Klieme tient compte des critiques formulées outre-Atlantique à l'égard des standards, en particulier de la part de ceux qui regrettent le manque d'alignement entre standards, curriculum, moyen d'enseignement et tests (Resnick & Zurawsky, 2005 ; Elmore, 2002 ; Müller & Silver, 2005). En résumé, Klieme, Avenarius, Blum *et alii* (2004, p. 22) demandent que les standards :

- reflètent aussi clairement que possible les objectifs fixés et les compétences exigées ;

- gardent un lien étroit avec les disciplines d'enseignement et tiennent compte des principes qui les régissent ;
- soient applicables, c'est-à-dire s'appuient sur des ressources disponibles jugées réalistes et suffisantes.

D'après Klieme, les standards ne définissent qu'une partie d'un domaine d'étude et peuvent être considérés comme un noyau de celui-ci. Les standards renvoient à des compétences qui s'acquièrent par un processus d'apprentissage cumulatif. La formulation des standards doit être claire, concise et compréhensible pour tous les acteurs concernés. Non seulement ils fixent un seuil, mais ils établissent également des distinctions entre différents niveaux de compétences sur un même continuum d'apprentissage.

Le rôle que remplissent les standards à l'intérieur de chaque système varie cependant d'un pays à l'autre. Dans les documents officiels, les trois pays utilisent l'argument des faibles résultats à l'enquête PISA pour justifier le passage vers un monitoring du système par les résultats. L'enjeu est de pouvoir évaluer le système dans son ensemble, ce qui accentue la fonctionnalité évaluative des standards. Cependant les pays hésitent à se fixer des cibles claires (Baslé, 2008, p. 49), même si les standards basés sur des modèles de compétences empiriquement validés, tels que recommandés par Klieme, Avenarius, Blum *et alii* (2004), devraient le permettre. Surtout dans les deux pays avec un fort pouvoir décentralisé, l'utilisation qui est faite des résultats est un objet de négociation essentiellement en phase de construction des dispositifs d'évaluation. Comme c'était le cas en Suisse romande, cette utilisation tient compte de l'équilibre des pouvoirs en place, des débats sociétaux, de la tradition évaluative, des structures existantes, tout en cherchant à donner suite aux injonctions faites au niveau national. Il s'en suivra une multitude de dispositifs et des usages très variés des résultats. Une enquête interne faite auprès des sept ministères d'éducation des cantons romands sur l'utilisation des résultats tirés des épreuves communes mentionnées plus haut montre une certaine propension à vouloir utiliser à des fins d'*accountability* administrative des résultats initialement destinés aux enseignants pour faciliter le travail avec leurs élèves en classe. Ainsi les résultats prennent comme fonction de rassurer l'administration scolaire (Maroy, 2011). Pourtant, si l'on tient compte à la fois de l'instabilité des résultats d'une année à l'autre et du rythme très lent auquel s'opèrent les changements dans un système, on peut s'interroger quant à la

pertinence de destiner les résultats en premier lieu aux administrations scolaires. Par ailleurs, si ces dernières sont les destinataires des informations générées, il se pourrait que les résultats ne présentent pas ou guère de plus-value pour les enseignants et soient perçus comme un dispositif se rajoutant à ceux existants pour contrôler leur travail. Dès lors, leur acceptation par les enseignants sera bien plus difficile.

Toutefois, les administrations scolaires semblent devenir conscientes d'un risque de surévaluation, pour des raisons de coût et de temps pris sur le temps d'enseignement. L'Allemagne comme la Suisse ont déjà remplacé les échantillons complémentaires de l'enquête PISA par *Land* ou canton par des tests de vérification des standards jugés plus précis et plus pertinents, ou prévoient de le faire. L'enquête PISA serait donc réduite au strict minimum et des items d'ancrage tirés des standards nationaux développés afin de pouvoir assurer une comparaison internationale.

Un autre point commun réside dans la vérification des standards par les trois pays, dans une logique plus large de monitoring du système qu'ils conçoivent comme un processus systématique et cyclique de suivi des résultats du système. Élément central de ce processus, un rapport pluriannuel sous forme de compilation de résultats statistiques, administratifs et de recherche comprendra des mesures de performance et intégrera les vérifications des standards à venir. De ce fait, on assiste en Suisse à l'émergence d'un nouveau niveau décisionnel qui se situe au niveau national, sans pour autant relever des compétences de la Confédération. En Allemagne semble également s'opérer un renforcement du pouvoir central de la KMK.

En Suisse, et dans une moindre mesure en Allemagne, apparaît une deuxième fonction des standards, celle d'harmonisation curriculaire. Le projet HarmoS suisse, en fixant les attentes de performance s'appuyant sur un modèle de compétences, a réussi à fixer un cadre technique servant de base au développement du plan d'études régional alémanique, tout en évitant d'ouvrir un débat politique épineux pour trouver un consensus sur les finalités et les contenus de l'école obligatoire. La Suisse romande qui ne regroupe que 7 cantons est également concernée par cette volonté d'harmonisation à travers les standards. Cependant il y existe une longue tradition de coordination en matière de plan d'études. Des travaux pour un nouveau plan d'études ont été engagés en 1990, bien avant le lancement du projet HarmoS, pour aboutir 20 ans plus tard. Comme le concordat

HarmoS impose la primauté des standards sur les plans d'études, il faudrait vérifier la compatibilité des standards avec le plan d'études romand et, le cas échéant, l'adopter. En revanche, les 18 cantons alémaniques ou bilingues, sans expérience notable de coordination curriculaire, se trouvent au début d'un processus délicat, rendu plus difficile encore par la reprise politique très polémique du sujet. En Allemagne, la refonte curriculaire existe également mais elle est plus indirecte et moins forte. D'après Oelkers et Reusser (2008, p. 440), les standards sont considérés comme une précision, voire une concrétisation des plans d'études. Selon une enquête de la KMK, la plupart des *Bundesländer* réajustent leurs plans d'études afin de pouvoir tenir compte des standards nationaux (Oelkers & Reusser, 2008, p. 453).

En Allemagne et en Autriche, les standards sont simultanément appelés à être utilisés pour réformer l'école obligatoire, ce qui constitue une troisième fonction que j'appellerai formative. Dans ce cas, les standards sont utilisés pour introduire une pédagogie des compétences dont la base théorique est la définition de Weinert (2001, p. 27 *sqq.*). Cette définition est très large puisqu'une compétence est comprise comme « les capacités et aptitudes cognitives dont l'individu dispose ou qu'il peut acquérir pour résoudre des problèmes précis, ainsi que les dispositions motivationnelles, volitives et sociales qui s'y rattachent pour utiliser avec succès et responsabilité les résolutions de problèmes dans des situations variables ». La compétence décrit donc le potentiel de l'individu à faire quelque chose. Elle s'oppose à la performance. Son évaluation nécessiterait des « dispositions devant être mobilisées contextuellement dans des tâches qui [...] doivent être imprévisibles, ce qui limite la part de l'apprentissage par cœur et de la mémorisation de solutions décontextualisées » (Elmiger, 2007, p. 37). Autrement dit, les résultats d'apprentissage issus de telles situations (par nature ouvertes) sont difficiles à tester. Cardinet, grand spécialiste de la mesure, avait déjà pointé les limites de cette ambition. Se référant à la pédagogie par objectifs, précurseur des standards, il estime que « trouver une façon de définir des objectifs pédagogiques [ou standards, pour changer la lecture] qui soient à la fois généraux et mesurables paraît [...] la quadrature du cercle : ou bien on contrôle de façon précise, mais les objectifs [ou standards] sont multiples et arbitraires, ou bien on se contente de finalités [ou standards] générales, mais non opérationnalisées » (Cardinet, 1988, p. 158). Autrement dit, « la signification réelle de la performance observée peut être très différente de ce que l'enseignant souhaiterait contrôler... Inver-

sement, un échec ne révèle pas nécessairement l'absence de la compétence contrôlée » (Cardinet, 1988, p. 160). À mon avis, la situation n'a pas changé. Cette difficulté a pu être observée dans l'élaboration des modèles de compétences du projet HarmoS. En effet les items retenus ne sont pas validés par rapport aux compétences du modèle et peinent à présenter des situations complexes de résolution de problèmes (Behrens, 2010). On peut se demander si l'effet de réduction curriculaire observé ailleurs (Müller & Silver, 2005) ne réduit pas à néant les ambitions d'un modèle à la fois capable de mesurer et pertinent pour une réforme vers une pédagogie par compétences.

Ces différentes fonctionnalités se superposent donc et peuvent même se modifier en cours d'élaboration des standards. Elles ont une influence sur la manière de concevoir les modèles de compétences et d'en inférer les standards (Behrens, 2010). En Suisse, par exemple, les mandats donnés aux consortiums chargés d'élaborer les modèles de compétences étaient d'abord formulés dans une finalité évaluative. Face à l'obligation de soumettre les propositions de standards à une large consultation, les responsables ont préféré privilégier la fonction curriculaire. Ainsi les standards finalement retenus se sont éloignés des propositions faites initialement par les consortiums et seront plus difficiles à vérifier à l'avenir.

Dans les trois pays, la vérification que les standards sont bien atteints sera effectuée en testant les élèves, mais les standards ne seront pas utilisés pour les certifier. Néanmoins l'utilisation des standards pour piloter des réformes de l'enseignement en Allemagne et en Autriche veut toutefois que les résultats soient communiqués aux élèves, aux enseignants et aux établissements afin de les informer de leur positionnement par rapport à une moyenne nationale et dans l'espoir de pouvoir ainsi initier des processus de remédiation. Mais, étant donné que les enseignants dans les trois pays ne sont pas habitués à des évaluations externes standardisées, le sujet est très sensible. En Autriche, les résultats des tests de vérification des standards auront donc une « fonction miroir » pour reprendre le concept de Thélot (2002). En Suisse, une comparaison entre établissements avec palmarès public n'est pas à l'ordre du jour non plus. En Allemagne, la situation n'est pas encore stabilisée (Oelkers & Reusser, 2008). Dans le rapport allemand de 2003, tous les *Bundesländer* confirment leur intention de développer des dispositifs d'évaluation externe. L'IQB dirige l'élaboration de tests comparatifs basés sur les standards en utilisant un test appelé VERA (*Vergleichsarbeiten*) et développé par l'Univer-

sité de Landau. Il est repris dans la plupart des *Bundesländer*. On observe également de nombreux tests et évaluations qui utilisent plutôt le curriculum comme référentiel. Dans les exemples allemands, les informations produites sont utilisées pour comparer des classes afin de permettre aux établissements de rendre compte de leurs activités. En Suisse, les tests nationaux sur la base d'échantillons d'élèves servent uniquement à l'évaluation du système. Au niveau des cantons, la situation est extrêmement variée, certains cantons connaissant des épreuves standardisées. Dans les cantons alémaniques, des dispositifs d'évaluation offerts par des entreprises privées sont surtout utilisés pour faciliter la transition à la fin de la scolarité obligatoire (Vermot, Behrens & Marc, 2011). Pour le moment la Suisse n'a pas de véritable institution responsable des tests, mais des réflexions sont en cours pour capitaliser le développement d'items en utilisant les forces en place dans les cantons. En Suisse romande, le tableau est similaire. La CIIP (Conférence intercantonale d'instruction publique) envisage des épreuves romandes communes. Ces dernières sont également comprises comme un outil d'assurance qualité puisqu'elles ont pour but de vérifier l'atteinte des objectifs du plan d'études romand. Comme la majorité des cantons disposent d'ores et déjà d'évaluations dont les fonctions varient fortement (Institut de recherche et de documentation pédagogique, 2010), chaque canton les conçoit différemment selon son histoire. Il est intéressant de relever que le cadre législatif ne se prononce pas de façon univoque sur ce qui pourra être fait des résultats. Il est donc difficile d'anticiper clairement les fonctions, la forme et les modalités de passation des épreuves à venir : à qui sont-elles destinées ? Veut-on les utiliser pour la certification des élèves en fin de scolarité obligatoire ? Doivent-elles permettre d'orienter la future carrière de l'élève, en particulier dans la transition vers la formation professionnelle ? À quel moment faut-il administrer les épreuves ? S'agit-il d'un dispositif de régulation des apprentissages ? Qu'en est-il de l'évaluation des enseignants et des établissements ?

APPORTS DES CONTRIBUTIONS ANGLOPHONES

Il apparaît donc dans la situation des pays germanophones que les standards n'amènent pas une uniformisation des structures scolaires mais fixent un cadre dans lequel seront construits les dispositifs en fonction de la structure politique, de l'histoire et de la culture évaluative locale. Néanmoins, selon la fonc-

tionnalité dominante, les questions relatives aux standards peuvent différer. Les expériences en Suisse romande sembleraient indiquer que la demande de résultats serait justifiée par le besoin de rassurer les administrations quant au bon fonctionnement de leurs classes. Dès lors, on peut se demander si l'évaluation se réalise au détriment de la fonction formative. Ainsi Stobart, dans ce numéro, se demande-t-il à quelle fin est utilisée l'évaluation dans sa fonctionnalité formative. Il propose deux possibilités, celle de l'élève et celle de l'enseignant. Si l'attention est portée sur les enseignants qui, rendus responsables des résultats obtenus par leurs élèves, deviennent l'objet des mesures politiques, on risque d'oublier la perspective de l'élève, les processus d'acquisition qu'il met en route, son autonomie, y compris également l'appropriation de compétences métacognitives lui permettant de développer ses propres stratégies d'apprentissage, sa motivation, etc.

Cette analyse rejoint le discours d'Elmore (2002), qui dit que la structure organisationnelle et la culture des établissements aux États-Unis ne favorisent pas les apprentissages mais préparent à la passation de tests. Si l'on veut modifier cette situation, il faut agir sur les enseignants et les amener à modifier leurs pratiques. Le constat d'une formation professionnelle insuffisante des enseignants pourrait expliquer les efforts vers le développement d'une nouvelle culture évaluative intégrée dans les pratiques d'enseignement. D'après Tochon (2006), il existerait aux États-Unis une tendance croissante à engager des enseignants autodidactes ou des enseignants spécialistes de leur discipline sans formation pédagogique. Cette conception serait également le fruit de tentatives de dérégulation du système scolaire et s'accentuerait en période de pénurie d'enseignants.

Ce constat d'une nécessité accrue de formation des enseignants apparaît comme un fil rouge à travers les contributions faites dans ce numéro. Une exigence banale de prime abord, mais bien délicate, surtout lorsqu'on la pense en termes de réussite des élèves faibles scolairement. À cet égard, il faut rappeler l'influence importante de facteurs contextuels et socioculturels dans l'explication des résultats de performance des élèves. Les contributions de ce numéro n'abordent pas cette question et partent toutes du présupposé qu'il est possible d'améliorer la performance des établissements et des classes sans tenir compte de leur influence, attitude qui a été critiquée dans le débat *school effectiveness versus school improvement* (Thrupp, 2001) ou dans l'analyse de l'influence des contraintes du marché scolaire sur les

stratégies de réussite des établissements (Felouzis & Perroton, 2007). En analysant les recherches étudiant les effets de la formation des enseignants sur la performance des élèves, Timperley (dans ce numéro) montre que certaines modalités donneraient de meilleurs résultats. Il s'agirait de perfectionnement sous forme de recherches ou d'enquêtes collaboratives portant sur les pratiques des enseignants. Huberman (1986, 1995) place les enseignants dans une posture de chercheur qui, soutenu par des experts externes, interroge sa pratique et propose des remédiations. C'est ce type de dispositif qui est présenté et discuté par Annan et Robinson ici, dans des contextes fort différents.

L'exemple néo-zélandais, mais aussi les travaux new-yorkais présentés dans ce numéro (respectivement par Annan et Robinson), montrent d'abord à quel point il est difficile pour les enseignants d'interpréter les résultats d'évaluation afin d'en tirer des conclusions pour ajuster leurs pratiques, surtout avec les élèves faibles. Ce constat n'est pas étonnant car des difficultés similaires ont pu être observées dans l'interprétation des résultats de PISA et dans l'utilisation des résultats d'évaluations standardisées (Dierendonck, 2008) et il est à craindre que les résultats provenant des vérifications des standards se heurtent aux mêmes difficultés. Les raisons peuvent être multiples. Une première difficulté réside dans le fait que les items des tests ne sont pas toujours bien alignés sur les standards, que les standards se veulent détachés des curriculums, ou encore que les standards formulés en termes de compétences ne correspondent pas à la préoccupation de l'enseignant. Mais lorsque l'évaluation est alignée avec le curriculum, il est possible d'observer des formes d'interprétation variées. L'observation de Gilliéron Giroud et Pulzer-Graf (2009), que nous avons évoquée en début d'article, est confirmée par les résultats de Nabors Oláh, qui observe également (dans ce numéro) une tendance à la réinterprétation des résultats d'évaluation en fonction des jugements préexistants des enseignants. C'est seulement dans certains cas que les enseignants s'engageraient à raisonner réellement en termes de remédiation vis-à-vis des difficultés constatées et il faut s'interroger sur les conditions qui doivent être remplies pour initier un tel travail.

Il n'est pas étonnant de constater que les trois études de cas font état de recours à des experts en éducation qui interviendraient dans les établissements pour aider les enseignants dans l'interprétation des résultats d'évaluation sous forme de tests standardisés, pour assurer l'accès à des ressources sup-

plémentaires et pour apporter un accompagnement *ad hoc*. Ces interventions sont conçues dans le cas néo-zélandais comme des mesures de formation continue et peuvent prendre des formes très différentes. Annan (dans ce numéro) rapporte la mise en réseau des établissements par des enquêtes collaboratives. Ainsi on trouve des enseignants, des chefs d'établissement, des représentants du ministère et des prestataires professionnels dans un groupe qui analyse les performances d'un établissement. On peut se demander ici quelle dynamique de collaboration se développe entre personnes de niveau hiérarchique très différent. À en croire Robinson (dans ce numéro), les prestataires professionnels ou les experts peuvent provenir d'un marché privé pour l'accompagnement scolaire, auquel cas il s'agit de définir les modalités financières de ces opérations.

Toutes ces contributions développent l'idée d'une boucle cybernétique comprenant la mesure, l'interprétation des résultats et des actions correctrices. Certains auteurs la rattachent aux processus rétroactif ou proactif d'évaluation formative mais ce qui est frappant, c'est que le modèle de référence dans le contexte de la production scolaire est la roue de qualité de Deming (cf. Nabors Oláh, dans ce numéro), un modèle cybernétique issu de la gestion des processus industriels (Behrens, 2007). Dans certains cas, un effort est fait pour tenir compte d'autres informations que les résultats de tests : des indicateurs qualitatifs tels que la culture de l'établissement et le leadership, par exemple (cf. Robinson, dans ce numéro).

CONCLUSION

Le monitoring par les résultats et, par conséquent, l'obligation de résultats suscitent un engouement considérable dans le monde politico-administratif responsable de la bonne gestion de l'école, probablement parce qu'ils alimentent l'idée de pouvoir mieux contrôler son fonctionnement. Ceci malgré le fait que les résultats de recherche sur des effets de tels modes de gestion restent mitigés quant à l'amélioration des performances à long terme (Lind, 2009 ; Mons, 2009) et aux gains d'efficacité et d'efficience que produirait une autonomie accrue des établissements (Felouzis & Perroton, 2007 ; Mons, 2009). En effet les facteurs influençant une amélioration des résultats semblent plus complexes et ne peuvent être considérés indépendamment d'éléments contextuels (Thrupp, 2001), ce qui rend une instrumentalisation

difficile. Néanmoins l'idée centrale de pouvoir maîtriser à des fins de monitoring ce mécanisme compliqué reste très attrayante et l'orientation graduelle des systèmes vers un monitoring par les résultats se poursuivra.

Les trois exemples germanophones montrent qu'une même base théorique (cf. Klieme, Avenarius, Blum *et al.*, 2004) se trouve à l'origine de trois dispositifs, qui diffèrent cependant en fonction des contextes locaux, des traditions scolaires et des structures politiques existantes. Au niveau conceptuel, Klieme a pris en considération les expériences anglophones en matière de standards, mais il n'est pas certain que ce cadre théorique se maintienne pendant l'opérationnalisation définitive des dispositifs. Dans les trois pays germanophones, la transformation des systèmes scolaires est en cours. Comme nous le dit Baslé (2008), c'est une négociation constante et délicate entre les différents acteurs dans le but de définir les droits et devoirs de chacun. Au centre de cette architecture, on trouve un référentiel de performances sous forme de standards et basé sur un modèle de compétences. Les programmes et des dispositifs d'évaluation permettant de vérifier si les standards sont atteints sont en train d'être alignés. Dans ce processus, le niveau politico-administratif semble se renforcer pour le moment (Rönnberg, 2008 ; Mons & Pons, 2006) en imposant des prises d'information dont la fonction principale est d'assurer l'*accountability* administrative à l'intérieur du système. On peut se demander si cette tendance, observée également en Suisse romande, ne risque pas d'alourdir considérablement l'efficacité du système scolaire et si elle contribue réellement à la qualité de l'enseignement. Mais ces questions restent ouvertes pour le moment, même si du côté des enseignants certaines réserves commencent à se faire sentir.

L'architecture scolaire émergeant de ces processus peut être analysée en strates fonctionnelles, chacune confrontée à ses propres besoins d'informations, enjeux, défis et risques à gérer. Dans son analyse d'un canton suisse alémanique, Bähr (2006) distingue essentiellement un niveau politico-administratif, comprenant les décideurs politiques et une administration centrale, et un niveau formatif comprenant les directions d'établissement et les enseignants dans leurs classes. La forte présence d'acteurs internationaux, nationaux et régionaux mise en évidence par Lingard et Rawolle (2010) et reproduite dans les États fédéraux entre pouvoir central et pouvoirs régionaux plaiderait pour une strate supplémentaire. Elle peut interférer dans la prise de décision au niveau politico-

administratif (il suffit de penser à l'effet des résultats de l'enquête PISA ou des tests nationaux de vérification des standards à venir).

Bähr (2006, p. 131), s'inspirant de la théorie des systèmes sociaux de Luhmann (1984), estime que les acteurs de chaque niveau ont besoin d'informations pour exercer leur fonction. Ainsi les représentants politiques raisonnaient en fonction des enjeux de pouvoir, du choix des programmes et des personnes et de l'*accountability*, alors que les acteurs administratifs sont surtout concernés par des questions d'efficacité et de réglementation. Au niveau formatif, les demandes sont différentes à cause des fonctions professionnelles exercées. Les responsables d'établissement sont concernés par le respect du cadre réglementaire, le traitement équitable des utilisateurs du système et le leadership par rapport à l'équipe pédagogique. Les enseignants, enfin, sont concernés par des données leur permettant de répondre aux questions pédagogiques et/ou didactiques, en rapport avec les disciplines de référence. Autrement dit, ces acteurs ont des préoccupations et des visions de qualité divergentes. Par conséquent, leurs besoins d'information en termes de fréquence et de contenu varient également. Cependant il n'est pas certain qu'ils ne puissent être satisfaits par une même prise d'information.

Du point de vue de la recherche, un travail important reste à faire pour comprendre la perception qu'ont les acteurs de leur système et des facteurs qui permettent d'optimiser son efficacité. Resnick et Besterfield-Sacre (2007) présentent des pistes de travail intéressantes et proposent un cadre d'analyse original pour comprendre la gestion de l'efficacité d'un système de performance scolaire. Une manière d'y répondre pourrait également être de rappeler que la classe est un espace protégé dans lequel les enseignants agissent, à l'instar des médecins, en vertu de leurs compétences et de leur statut professionnel. Les informations relatives à la régulation des apprentissages resteraient confinées à cet espace et ne montreraient que sous certaines conditions vers la direction de l'établissement, mais non pas vers le niveau politico-administratif. Des procédures, la nature des données et les flux vers le niveau supé-

rieur resteraient à définir et permettraient de satisfaire aux besoins d'*accountability* de la hiérarchie scolaire. Ils ne devraient pas se limiter aux mauvaises performances des élèves mais inclure d'autres informations. De même, l'établissement scolaire jouirait plus fortement encore de son autonomie et pourrait, en fonction des problèmes rencontrés, solliciter des appuis externes. Si cette piste devait être suivie, de nombreux travaux restent à faire. Il serait particulièrement important d'analyser les flux d'information entre les composants du système en fonction de leurs intérêts et de questionner le ratio entre l'efficacité dans l'acquisition et la plus-value que représente cette donnée par rapport à d'autres, en tenant compte de la nature des processus en jeu.

Dans cette perspective, les contributions anglophones montrent une ouverture possible. À l'opposé des préoccupations structurelles observées dans les exemples germanophones, elles reviennent à la question de la formation. En effet une première génération de standards n'a pas permis d'obtenir de meilleurs résultats scolaires (Robinson, dans ce numéro, citant Fuhrman & Elmore, 2004), en particulier avec des élèves en difficulté. De façon unanime, la lecture, la compréhension et l'interprétation des résultats provenant de tests de vérification de standards ne sont pas chose simple, au point que les équipes enseignantes ont dû être accompagnées par des spécialistes de l'évaluation. Autrement dit, le passage à un monitoring du système par les résultats ne sera possible que si les responsables politico-administratifs prennent conscience qu'une réforme ne se décrète pas et qu'il faut un travail minutieux de suivi du terrain. Ainsi il ne suffit pas non plus de prescrire des standards pour les voir intégrés directement ou indirectement dans les pratiques. Il faut impérativement tenir compte du contexte, mettre à disposition des ressources et rendre possibles des dispositifs de suivi qui, comme nous l'ont montré Huberman (1986, 1995) et Timperley (dans ce numéro), donnent les meilleurs résultats s'ils se construisent sur la trame de recherches collaboratives.

Matthis Behrens
Matthis.Behrens@irdp.ch
Institut de recherche et de documentation
pédagogique, Suisse

BIBLIOGRAPHIE

- AGENCE EXÉCUTIVE ÉDUCATION, AUDIOVISUEL ET CULTURE (2009a). *Les évaluations standardisées pour les élèves en Europe : objectifs, organisation et utilisation des résultats*. Bruxelles : Agence exécutive éducation, audiovisuel et culture.
- AGENCE EXÉCUTIVE ÉDUCATION, AUDIOVISUEL ET CULTURE (2009b). *Chiffres clés de l'éducation 2009 : une vue d'ensemble sur les systèmes éducatifs européens*. Bruxelles : Agence exécutive éducation, audiovisuel et culture.
- ALLAL L. (2009). « Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire ». In L. Mottier Lopez & M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles : De Boeck, p. 29-45.
- ALLAL L. & MOTTIER LOPEZ L. (2005). « L'évaluation formative de l'apprentissage : revue de publications en langue française ». In OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires*. Paris : OCDE, p. 265-290.
- BÄHR K. (2006). « Erwartungen von Bildungsadministrationen an Schulleistungstests ». In H. Kuper & J. Schneewind (dir.), *Rückmeldung und Rezeption von Forschungsergebnissen: Zur Verwendung wissenschaftlichen Wissens im Bildungssystem*. Münster : Waxmann, p. 125-142.
- BASLÉ M. (2008). *Économie, conseil et gestion publique. Suivi et évaluation des politiques publiques et des programmes*. Paris : Economica.
- BEHRENS M. (2006). « Standards : quand le politique s'empare de l'évaluation ». In G. Figari & L. Mottier Lopez (dir.), *Recherche sur l'évaluation en éducation : problématiques, méthodologies et épistémologie (20 ans de travaux autour de l'ADMEE-Europe)*. Paris : L'Harmattan, p. 93-100.
- BEHRENS M. (2007). « Introduction ». In M. Behrens (dir.), *La qualité en éducation. Pour réfléchir à la formation de demain*. Québec : Presses de l'Université du Québec, p. 3-18.
- BEHRENS M. (2008). « Les standards de formation, l'anneau de Möbius de l'école suisse ? » *Revue internationale d'éducation de Sèvres*, n° 48, p. 127-136.
- BEHRENS M. (2010). « Standards ou les escaliers mouvants de Poudlard ». In P. Gilliéron Giroud & L. Ntamakiliro (dir.), *Réformer l'évaluation scolaire : mission impossible ?* Berne : Peter Lang, p. 113-138.
- BERLINER D. & BIDDLE B. (1995). *The manufactured crises. Myths, fraud, and the attack on America's public school*. New York : Addison-Wesley et Longman.
- BRACEY G. (2002). *The war against America's public schools. Privatizing schools, commercializing education*. Boston : Allyn & Bacon.
- BREIT S., FRIEDL-LUCYSHYN G., FURLAN N. et al. (2010). *Bildungsstandards in Oesterreich, Überprüfung und Rückmeldung*. Salzburg : Bildungsforschung, Innovation und Entwicklung für das oesterreichische Schulwesen.
- CARDINET J. (1988). « La maîtrise, communication réussie ». In M. Huberman (dir.), *Assurer la réussite des apprentissages scolaires*. Neuchâtel : Delachaux & Niestlé, p. 127-154.
- CENTRE SUISSE DE COORDINATION POUR LA RECHERCHE EN ÉDUCATION (2010). *L'éducation en Suisse. Rapport 2010*. Aarau : Centre suisse de coordination pour la recherche en éducation.
- COMMISSION DES COMMUNAUTÉS EUROPÉENNES (2003). « Éducation et formation 2010 ». *L'urgence des réformes pour réussir la stratégie de Lisbonne*. Bruxelles : Commission des communautés européennes. Disponible sur Internet à l'adresse : <http://europa.eu/legislation_summaries/education_training_youth/general_framework/c11071_fr.htm> (consulté le 26 janvier 2011).
- COMMISSION EUROPÉENNE (1996). *Livre blanc sur l'éducation et la formation. Enseigner et apprendre. Vers la société cognitive*. Bruxelles : Commission européenne. Disponible sur Internet à l'adresse : <http://europa.eu/documents/comm/white_papers/pdf/com95_590_fr.pdf> (consulté le 26 janvier 2011).
- CONFÉRENCE SUISSE DES DIRECTEURS CANTONAUX DE L'INSTRUCTION PUBLIQUE (2007). *Accord intercantonal sur l'harmonisation de la scolarité obligatoire (concordat HarmoS)*. Berne : Conférence suisse des directeurs cantonaux de l'instruction publique. Disponible sur Internet à l'adresse : <http://edudoc.ch/record/24710/files/HarmoS_f.pdf> (consulté le 26 janvier 2011).
- CONSEIL DE L'UNION EUROPÉENNE (2009). « Conclusions du Conseil du 12 mai 2009 concernant un cadre stratégique pour la coopération européenne dans le domaine de l'éducation et de la formation ("Éducation et formation 2020") ». *Journal officiel de l'Union européenne*, 28 mai 2009, C 119. Disponible sur Internet à l'adresse : <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52009XG0528%2801%29:FR:NOT>> (consulté le 26 janvier 2011).
- CRAHAY M. (2009). « Articuler l'évaluation en classe et le pilotage des systèmes, est-ce possible ? » In L. Mottier Lopez & M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles : De Boeck, p. 233-251.
- DIERENDONCK C. (2008). *Comment les évaluations externes des acquis des élèves sont-elles perçues par les enseignants du primaire dans les cantons de Neuchâtel, Vaud et Fribourg ?* Document de travail, n° 08.1002. Neuchâtel : Institut de recherche et de documentation pédagogique. Disponible sur Internet à l'adresse : <<http://publications.irdp.relation.ch/ftp/1217508758081002.pdf>> (consulté le 26 janvier 2011).
- DIVISION DES POLITIQUES LINGUISTIQUES (2000). *Un cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Strasbourg : Conseil de l'Europe. Disponible sur Internet à l'adresse : <http://www.coe.int/t/dg4/linguistic/cadre_fr.asp> (consulté le 26 janvier 2011).
- DURU-BELLAT M. (2005). *L'inflation scolaire. Les désillusions de la méritocratie*. Paris : Éd. du Seuil.

- ELMIGER D. (2007). « Un modèle de compétences pour les langues enseignées à l'école ». *Babylonia*, n° 4, p. 35-39.
- ELMORE R. (2002). « The price of accountability. Want to improve schools? Invest in the people who work in them ». *Results*, November 2002. Disponible sur Internet à l'adresse : <<http://www.nsd.org/members/results/results11-02.pdf>> (consulté le 26 janvier 2011).
- EURYDICE (2010). *Eurybase. National system overviews on education systems in Europe and ongoing reforms. 2010 Edition*. Bruxelles : Agence exécutive éducation, audiovisuel et culture. Disponible sur Internet à l'adresse : <http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/national_summary_sheets/047_AT_EN.pdf> (consulté le 26 janvier 2011).
- EURYDICE (2011a). *Eurybase. Organisation des Bildungssysteme in der Bundesrepublik Deutschland 2009/2010*. Bruxelles : Agence exécutive éducation, audiovisuel et culture. Disponible sur Internet à l'adresse : <http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/eurybase_full_reports/DE_DE.pdf> (consulté le 26 janvier 2011).
- EURYDICE (2011b). *Eurybase. Organisation des Bildungssysteme in Österreich 2008/2009*. Bruxelles : Agence exécutive éducation, audiovisuel et culture. Disponible sur Internet à l'adresse : <http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/eurybase_full_reports/AT_DE.pdf> (consulté le 26 janvier 2011).
- FELOUZIS G. & PERROTON J. (2007). « Repenser les effets d'établissement : marchés scolaires et mobilisation ». *Revue française de pédagogie*, n° 159, p. 103-118.
- FUHRMAN S. & ELMORE R. (2004). *Redesigning accountability systems for education*. New York : Teachers College Press.
- GAUTHIER R.-F. & LE GOUVELLO M. (2010). « The French curricular exception and the troubles of education and internationalisation: Will it be enough to "rearrange the deckchairs"? » *European Journal of Education*, vol. 45, n° 1, p. 74-88.
- GILLIÉRON GIROUD P. & PULZER-GRAF P. (2009). *Apprentissage de la lecture et promotion de fin de 2^e année primaire : points de vue d'enseignantes sur les nouvelles dispositions institutionnelles*. Lausanne : Unité de recherche pour le pilotage des systèmes pédagogiques.
- HUBERMAN M. (1986). « Un nouveau modèle pour le développement professionnel des enseignants ». *Revue française de pédagogie*, n° 75, p. 5-15.
- HUBERMAN M. (1995). « Networks that alter teaching: Conceptualizations, exchanges and experiments ». *Teachers and Teaching: Theory and Practice*, vol. 1, n° 2, p. 193-211.
- INSTITUT DE RECHERCHE ET DE DOCUMENTATION PÉDAGOGIQUE (2010). *Pratiques cantonales concernant l'organisation d'épreuves, d'examen, de tests ainsi que l'obtention de certificats ou diplômes. Enseignement primaire, secondaire premier cycle, Suisse romande*. Neuchâtel : Institut de recherche et de documentation pédagogique. Disponible sur Internet à l'adresse : <http://www.irdp.ch/publicat/irdp_dossiers_comparatifs/epreuves/2010_2011/epreuvesdocumentcomplet1011.pdf> (consulté le 26 janvier 2011).
- KLIEME E., AVENARIUS H., BLUM W. et al. (2004). *Le développement de standards nationaux de formation : une expertise*. Bonn : Ministère fédéral de l'Éducation et de la Recherche. Disponible sur Internet à l'adresse : <http://www.bmbf.de/pub/le_developpement_de_standards_nationaux_d_formation.pdf> (consulté le 26 janvier 2011).
- KMK (2004). *Veröffentlichungen der Kultusministerkonferenz. Bildungsstandards der Kultusministerkonferenz. Erläuterung zur Konzeption und Entwicklung*. München, Neuwied : Luchterhand. Disponible sur Internet à l'adresse : <http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf> (consulté le 26 janvier 2011).
- LAVEAULT D. (2009). « L'amélioration de l'efficacité du système éducatif : sur quels indicateurs s'appuyer ? » In X. Dumay & V. Dupriez (dir.), *L'efficacité dans l'enseignement. Promesses et zones d'ombre*. Bruxelles : De Boeck, p. 177-194.
- LAWN M. & NÓVOA A. (2005). « L'espace, l'éducation et l'Europe ». In M. Lawn & A. Nóvoa (dir.), *L'Europe réinventée, regards critiques sur l'espace européen de l'éducation*. Paris : L'Harmattan.
- LIND G. (2009). « Ameerika als Vorbild? Erwünschte und unerwünschte Folgen aus Evaluationen ». In T. Bohl & H. Kiper (dir.), *Lernen aus Evaluationsergebnissen*. Bad Heilbrunn : Klinkhardt.
- LINGARD B. & RAWOLLE S. (2010). « Globalization and the rescaling of education politics and policy ». In M. Larsen (dir.), *New thinking of comparative education: Honouring the work of Robert Cowen*. Rotterdam : Sense Publishers.
- LUHMANN N. (1984). *Soziale Systeme. Grundrisse einer allgemeinen Theorie*. Frankfurt am Main : Suhrkamp.
- MALAISE S. (2011). « Classification hiérarchique de compétences : comparaison d'une méthode basée sur l'analyse des fréquences de réussite et d'une méthode d'analyse statistique implicative ». *Bulletin de l'ADMEE-Europe*, n° 2011/1.
- MAROY C. (2011). « Accountability et confiance dans l'institution scolaire ». *Éducation comparée*, n° 5 [à paraître].
- MAROY C. & MANGEZ C. (2011). « La construction des politiques d'évaluation et de pilotage du système scolaire en Belgique francophone : nouveau paradigme politique et médiation des experts ». In G. Felouzis & S. Hanhart (dir.), *Politiques éducatives et évaluation. Nouvelles tendances, nouveaux acteurs* [à paraître]. Bruxelles : De Boeck.
- MEUNIER O. (2005). « Standards, compétences de base et socle commun ». In M. Behrens (dir.), *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif, réalisé dans le cadre du projet HarmoS*. Neuchâtel : Institut de recherche et de documentation pédagogique, p. 69-100.

- MINISTÈRE DE L'ÉDUCATION NATIONALE, DE LA JEUNESSE ET DE LA VIE ASSOCIATIVE (2010). *L'état de l'école : 29 indicateurs sur le système éducatif français*. Paris : Ministère de l'Éducation nationale, de la jeunesse et de la vie associative. Disponible sur Internet à l'adresse : <<http://www.education.gouv.fr/cid53863/l-etat-ecole-indicateurs-sur-systeme-educatif-francais.html>> (consulté le 26 janvier 2011).
- MONS N. (2009). *Les effets théoriques et réels de l'évaluation standardisée*. Bruxelles : Agence exécutive éducation, audiovisuel et culture. Disponible sur Internet à l'adresse : <http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111FR.pdf> (consulté le 26 janvier 2011).
- MONS N. & PONS X. (2006). *Les standards en éducation dans le monde francophone, une analyse comparative*. Neuchâtel : Institut de recherche et de documentation pédagogique.
- MOTTIER LOPEZ L. (2009). « L'évaluation en éducation : des tensions aux controverses ». In L. Mottier Lopez & M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles : De Boeck, p. 7-25.
- MOTTIER LOPEZ L. & LAVEAULT D. (2008). « L'évaluation des apprentissages en contexte scolaire : 30 ans de travaux. Développements, enjeux et controverses ». *Mesure et évaluation en éducation*, vol. 31, n° 3, p. 5-34.
- MÜLLER K. & SILVER R. (2005). « Standards in education: Review of US literature ». In M. Behrens (dir.), *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif, réalisé dans le cadre du projet HarMoS*. Neuchâtel : Institut de recherche et de documentation pédagogique, p. 9-68.
- NOIZET G. & CAVERNI J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : PUF.
- OCDE (1999). *Nomenclature des systèmes d'éducation, guide d'utilisation de la CITE-97 dans les pays de l'OCDE*. Paris : OCDE.
- OCDE (2009). *Regards sur l'éducation 2009 : les indicateurs de l'OCDE*. Paris : OCDE.
- OCDE (2011). *PISA en français*. Paris : OCDE. Disponible sur Internet à l'adresse : <http://www.oecd.org/document/24/0,3746,en_32252351_32235731_38378840_1_1_1_1,00.html> (consulté le 26 janvier 2011).
- OELKERS J. & REUSSER K. (2008). *Qualität entwickeln-Standards sichern- mit Differenz umgehen. Bildungsforschung Band 27*. Berlin : Bundesministerium für Bildung und Forschung.
- PIÉRON H. (1963). *Examen de docimologie*. Paris : PUF.
- RESNICK L. & BESTERFIELD-SACRE M. (2007). « A framework for effective management of school system performance ». In P. Moss (dir.), *Evidence and decision making. Yearbook of the National Society for the Study of Education*, vol. 106, n° 1. Massachusetts : Blackwell Publishing Malden.
- RESNICK L. & ZURAWSKY C. (2005). « Standards-based reform and accountability: Getting back on course ». *American Educator*, vol. 29, n° 1, p. 1-13.
- REY B. (1996). *Les compétences transversales en question*. Paris : ESF.
- REY B. & CARETTE V. (2005). « Référentiels de compétences scolaires et outils d'évaluation de compétences : la situation en Communauté française de Belgique ». In M. Behrens (dir.), *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif, réalisé dans le cadre du projet HarMoS*. Neuchâtel : Institut de recherche et de documentation pédagogique, p. 101-109.
- RÖNNBERG L. (2008). *The re-emerging state? Motives and arguments underlying the réintroduction of Swedish national school inspections*. Policy studies and politics of education, network n° 23. Göteborg : European Educational Research Association.
- RÜRUP M. (2003). « Ausländische und internationale Bildungsberichte als Orientierung für die nationale Bildungsberichterstattung in Deutschland ». In H. Avenarius, H. Ditton, H. Döbert et al., *Bildungsbericht-erstattung für Deutschland: Konzeption*. Frankfurt am Main : Deutsches Institut für Internationale Pädagogische Forschung.
- THÉLOT C. (2002). « Évaluer l'école ». *Études*, vol. 397, n° 10, p. 323-334.
- THÉLOT C. (2004). *Pour la réussite de tous les élèves. Rapport de la commission du débat national sur l'avenir de l'École*. Paris : La Documentation française. Disponible sur Internet à l'adresse : <<http://www.ladocumentationfrancaise.fr/rapports-publics/044000483/>> (consulté le 26 janvier 2011).
- THRUPP M. (2001). « Sociological and political concerns about schools effectiveness research: Time for a new research agenda ». *School Effectiveness and School Improvement*, vol. 12, n° 1, p. 7-40.
- TIMES (2008). « No Child Left Behind: Doomed to fail ». Disponible sur Internet à l'adresse : <<http://www.time.com/time/nation/article/0,8599,1812758,00.html>> (consulté le 26 janvier 2011).
- TOCHON F. (2006). « La formation des enseignants aux États-Unis : enjeux socio-économiques et professionnels actuels ». *Formation et pratiques d'enseignement en question*, n° 5, p. 29-41.
- VERMOT B., BEHRENS M. & MARC V. (2011). « Étude comparative de deux tests assistés par ordinateur en vue de l'élaboration d'épreuves commune en Suisse romande ». In J.-G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication. Le futur est à notre porte*. Québec : Presses de l'Université Laval.
- WEINERT F. (2001). « Vergleichende Leistungsmessung in Schulen: eine umstrittene Selbstverständlichkeit ». In F. Weinert (dir.), *Leistungsmessung in Schulen*. Weinheim : Beltz, p. 17-31.