

## 1. ROLE ACCRU DES STANDARDS ET DES REFERENTIELS

### 1.1 LES STANDARDS

#### Standards : quand le politique s'empare de l'évaluation

*Matthis Behrens*

L'onde de choc consécutive aux premiers résultats de l'enquête PISA<sup>1</sup> a provoqué un large débat sur l'école, ses structures et ses performances dans de nombreux pays, en Europe et ailleurs. En réponse à ce débat, la Conférence intercantonale des directeurs de l'instruction publique (CDIP), organe de coordination de l'éducation et de la formation en Suisse, a habilement proposé une série de mesures pour harmoniser les structures éducatives décentralisées. Des standards nationaux de formation se trouveront au coeur de ces projets de réformes.

Très courant dans le monde anglo-saxon, le terme « standard » ne rencontre que peu ou pas d'écho dans le débat scientifique francophone. Les quelques productions significatives regorgent de références anglophones ou cherchent à rattacher les standards à des problématiques connues par ailleurs, celles de l'évaluation et de la mesure des performances, du pilotage des systèmes éducatifs par les résultats ou des référentiels. Dans le projet suisse de standards nationaux de formation, un emprunt culturel semble être en cours, il importera de voir comment ce concept migrera d'un contexte culturel à un autre, et d'observer les métissages qui s'y opèreront (Maroy, 2004).

#### Les standards de formation

##### *Clarification conceptuelle*

Il est significatif de noter que Lam (2004) établit un lien direct entre l'enseignement par objectifs et les standards. Rappelons que, selon Mager (1971), un objectif pédagogique a) identifie et décrit le

---

<sup>1</sup> Enquête de l'OCDE : Programme international pour le suivi des acquis des élèves.

comportement de l'étudiant, b) précise les conditions dans lesquelles ce comportement doit se manifester et c) définit les critères d'une performance acceptable. La description et l'identification des contenus correspondraient aux « standards de contenu » (*content standards*) et les critères de suffisance aux « standards de performance » (*performance standards*). Sur cette première distinction se greffent des indications concernant les conditions et les processus nécessaires à l'enseignement et à l'apprentissage, c'est-à-dire des « standards définissant la possibilité qui est donnée d'apprendre » (*opportunity to learn standards*) et des précisions en matière d'évaluation, sous la forme d'une sorte de charte déontologique qui correspond à des « standards de l'évaluation » (*assessment standards*).

Les *Principles and Standards* du *National Council of Teachers of Mathematics* (NCTM) présentés en 2002 sont un dispositif qui illustre bien quelques-unes de ces différences. Réalisé par une association professionnelle, le NCTM propose des standards de contenu qui « se rapportent plus spécifiquement à la branche et concernent aussi les contenus concrets du programme tandis que les principes (*principles*) sont des lignes directrices interdisciplinaires pour un enseignement de qualité, dont la formulation se rapporte aux contenus [...], ils peuvent donc être vus comme une sorte d'*opportunity-to-learn standard* » (Klieme et al. 2004, p. 35).

Cependant, avant la publication des standards NCTM, « le terme de *standard* a été essentiellement utilisé [aux Etats-Unis] comme synonyme de standard de performance [...]. La mesure de la performance figurait donc au premier plan et l'on se référerait avant tout aux aptitudes de base » (p. 31). On se préoccupait peu de la nature du référentiel d'évaluation et de la capacité ou non des items à positionner les performances mesurées de chaque individu sur une seule échelle de compétences.

### ***Les standards et la définition de seuils de performance***

Alors que l'enseignement par objectifs insiste sur une définition minutieuse de ces derniers dans le but de construire l'évaluation de l'apprenant selon des références critériées, les standards de performance fournissent également une référence critériée mais pour établir de façon normative des seuils de performance que le système est censé atteindre.

La fixation de ce niveau cible peut varier en fonction de l'étendue de la population que l'on cherche à y amener. Klieme et al. distinguent les « standards minimum » qui correspondent à un seuil en dessous duquel aucun apprenant ne doit demeurer, des « standards réguliers » qui doivent être atteints par la moyenne d'une population scolaire, et les « standards maximum » définissant un idéal éducatif.

En Suisse, les standards nationaux de formation sont des standards minimum. Ils doivent être contraignants, c'est-à-dire les acteurs du système scolaire sont tenus d'amener toute leur population d'élèves à la maîtrise d'un socle minimal de compétences. « Les modèles de compétences et les épreuves qui s'y rattachent [...] décrivent ensuite clairement les performances qu'un(e) élève doit réaliser [...] ce qui n'exclut pas de 'tester' des exigences plus élevées, en tant qu'objectifs de processus d'apprentissage subséquents ou pour la différenciation des groupes d'apprenants ». L'idée centrale d'un standard minimum est de « ne pas 'laisser en arrière' les élèves les plus faibles. Tout établissement scolaire, tout apprenant ou enseignant doit avoir clairement conscience des exigences minimum attendues » (p. 24). Cependant, bien que séduisantes pour satisfaire à l'impératif d'équité de l'école publique, les expériences nord-américaines n'ont pas réussi à démontrer que l'évaluation de compétences minimum (*minimum competency testing* - MCT) influence de manière positive les performances des élèves (Müller & Silver, 2006).

### ***Critères de qualité d'un standard***

L'analyse de la littérature anglophone et germanophone (Behrens, 2006 ; Berner & Stolz, 2006 ; Huber, Spähni, Schmellentin & Criblez, 2006) fait apparaître une diversité extrême dans les dispositifs de standards existants au niveau international. Chaque standard est le fruit d'une démarche particulière, conditionnée par des facteurs contextuels et historiques. Sur la base de ses analyses, Klieme et al. recommandent un certain nombre de caractéristiques pour assurer la qualité des standards :

- les standards doivent refléter aussi clairement que possible les objectifs fixés et les compétences exigées,
- les standards doivent garder un lien étroit avec les disciplines ou les branches d'enseignement et tenir compte des principes qui les

régissent,

- l'approche par les standards, en règle générale, ne couvre qu'une partie d'un domaine d'étude, qu'elle considère comme le 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 des distinctions entre différents niveaux de compétences sur un même continuum d'apprentissage,
- les standards doivent être applicables, c'est-à-dire s'appuyer sur des ressources disponibles jugées réalistes et suffisantes.

Müller et Silver (2006) en s'inspirant du contexte américain insistent sur le fait que les standards se doivent d'être cohérents d'un cycle scolaire à l'autre, en particulier au passage entre l'enseignement secondaire et l'enseignement tertiaire.

Cependant ces recommandations, certes précieuses, peuvent faire oublier que les critères de qualité des standards se déclinent avant tout selon la fonction qu'on leur attribue. Ce qui est souhaitable pour le monitoring risque d'être contreproductif dans la gestion des apprentissages. En effet, « des années d'expérience aux Etats-Unis ont prouvé que l'importance accordée aux tests basés sur des standards finit par conditionner les curriculums et l'enseignement, qu'on le veuille ou non. Les tests peuvent aussi avoir pour effet de réduire les curriculums pour n'inclure que des éléments destinés à être testés. Il est toujours plus facile de mettre sur pied un examen simple qui n'évalue que les tâches non complexes » (Müller & Silver, 2005, p. 59). Il faut également s'attendre à ce que les enseignants et les établissements conditionnent leur travail en fonction des enjeux en favorisant les « bons élèves » pour obtenir de « bons résultats », au détriment de ceux qui auraient davantage besoin de leur attention.

### ***Les standards comme référentiels pour l'évaluation des réformes***

La mise en relation de la logique d'enseignement par objectifs avec les modèles d'évaluation de programmes (Tyler, 1949), notamment le modèle d'évaluation de Provas (1971) (*Discrepancy Evaluation Model*) a

donné naissance au courant des réformes éducatives basées sur des standards (*standard based reforms*). Selon Lam (2004), ces réformes peuvent être modélisées par l'agencement cyclique de quatre étapes :

1. La définition de standards de performance pour les élèves, comme phase de référentialisation,
2. la responsabilisation des écoles et des enseignants quant à l'atteinte de ces standards,
3. l'évaluation à large échelle pour accélérer les changements dans les écoles, soutenue par des politiques d'encouragement,
4. l'utilisation des résultats de cette évaluation pour juger de la pertinence des réformes engagées et accompagner celles-ci.

Ces phases sont évidemment à nuancer en fonction des particularités de chaque système.

Les fondements des réformes basées sur des standards suggèrent que la mise en place de standards de performance non seulement pose le problème de la définition de ce qu'est une compétence et celui de la façon de l'opérationnaliser en performances observables, mais aussi favorise certains modèles théoriques d'apprentissage au détriment d'autres. Pour de nombreux observateurs, le fait de privilégier ainsi certaines méthodes d'apprentissage constitue une régression, par la mise en péril d'approches telles que le travail avec des situations-problèmes complexes. Cette tendance risque d'être renforcée par l'insuffisance théorique liée à la mesure. A en croire de Ketele et Gerard (2005), l'éduométrie, à l'heure actuelle, ne dispose pas d'un appareil méthodologique suffisamment large pour évaluer valablement les compétences complexes. De même, la tendance psychométrique à projeter sur une échelle mathématique unidimensionnelle le degré de maîtrise d'un ensemble d'items isomorphes n'est pas tenable lorsqu'on évalue la réalité multidimensionnelle des compétences complexes.

### ***L'élaboration de standards***

Plusieurs auteurs (Dubs, 2004 ; Elley, Hall & Marsh, 2005 ; Klieme et al., 2004) insistent sur la nécessité d'une bonne articulation entre les standards et les plans d'étude, ou du moins d'un consensus fort sur les contenus d'enseignement. Les standards doivent découler des finalités

éducatives ; selon Dubs, ces dernières doivent être légitimées par un débat public et non pas découler d'avis d'experts.

Les expériences faites en Allemagne montrent que les standards s'élaborent soit en complétant les plans d'études par la définition des performances attendues, soit en rédigeant de nouveaux plans d'étude qui intègrent de telles indications. Dans ce cas, les standards font partie intégrante du curriculum. On observe également le développement indépendant de standards à côté des curriculums existants lorsqu'on développe un modèle de compétence, qu'il dépende ou non du curriculum. Ce dernier cas de figure s'applique au projet VERA (*Vergleichsarbeiten* - travaux comparatifs) qui mise directement sur le développement d'une banque d'items pour l'évaluation des élèves au terme du degré primaire (Helmke & Hosenfeld, 2004). Dans ce cas, la définition des standards intervient a posteriori : « on conçoit des situations-problèmes en collaboration avec un grand nombre d'enseignant(e)s et l'on élabore des tests selon des critères psychométriques ; ces tests seront ensuite standardisés au niveau de l'ensemble du système » (Klieme et al., 2004, p. 135) sur la base des données empiriques récoltées. Cette démarche s'inspire des approches méthodologiques mises en oeuvre dans le cadre des grandes enquêtes internationales, TIMSS, PISA et PIRLS.

En Suisse, le projet HarmoS a pris l'option de développer des standards sans passer par les curriculums. Afin de contourner l'épineuse question d'un accord consensuel quant aux finalités et aux objectifs des différents systèmes éducatifs cantonaux, les pouvoirs politiques ont préféré le détour par des standards se basant sur un modèle de compétence défini a priori et indépendamment des plans d'étude, du type cadre européen des langues. En fixant ainsi les performances attendues à la fin des cycles scolaires, le projet HarmoS cherche à obtenir un effet indirect sur les curriculums cantonaux. Il est à relever que ce projet national met sérieusement en question les travaux curriculaires réalisés en Suisse romande, qui ont abouti à un projet de plan d'études cadre pour l'ensemble des cantons francophones. Dans ce sens, la dimension évaluative n'est pas neutre. C'est d'abord un projet politique qui a conduit à l'élaboration des standards de formation dans une optique

d'efficience. Le but n'est pas nécessairement l'amélioration de l'école, mais le contrôle de ses performances dans une logique de redevabilité.

Par conséquent, pour être valables dans plusieurs contextes culturels, les compétences ainsi définies doivent être extra-curriculaires. Elles sont issues d'une modélisation empirique des apprentissages, ce qui pourrait initier une référentialisation cognitive de l'évaluation. « Cette voie pragmatique facilite certes le lancement rapide de l'évaluation systématique des écoles et du monitoring de la formation, mais elle offre peut-être des possibilités de repère plus limitées aux enseignant(e)s... » (Klieme et al., 2004, p. 135).

### **Les standards pour quoi faire**

Berner et Stolz (2006) en citant la *Mid-continent Research for Education and Learning* (McREL) constatent qu'il existe peu de consensus sur la question de savoir comment doit être conçue une éducation basée sur des standards. L'usage qui est fait des standards aux Etats-Unis fait apparaître que ceux-ci constituent des références considérées comme stables et valides, permettant d'évaluer avec précision les systèmes de formation, les établissements et les enseignants.

En règle générale, les résultats ainsi produits sont utilisés par les administrations scolaires très sensibles à la productivité et à l'efficience des écoles. L'efficacité des établissements est déterminée en termes d'*outputs* et les résultats des élèves en sont la mesure essentielle. La fixation de critères, d'objectifs et leur mise en oeuvre s'inscrit dans une rationalité économique où l'intention des politiques éducatives est d'abord de réguler et de distribuer des ressources de manière rentable et efficace (Normand, 2005) sur la base, précisément, d'évaluations fondées sur des standards. En d'autres termes, les politiques éducatives recèlent des enjeux importants, d'où l'utilisation du terme *high-stake testing* (test à enjeu élevé). Il existe une abondante littérature anglophone qui questionne les effets et la validité de ces évaluations. Dans le cadre de cet article, précisons simplement que l'existence même de ces dispositifs tend à stimuler la demande (McREL). Leur construction et leur mise en oeuvre absorbent des moyens financiers considérables au détriment des mesures de remédiation, comme les moyens d'enseignement, les curriculums et la formation des enseignants.

La perspective d'un monitoring des systèmes sur la base de résultats est certes séduisante, mais ce n'est pas une affaire simple. Goldstein (2004) et Wolf (2000) attirent l'attention sur le présupposé central, mais erroné, que les standards restent une référence stable dans la durée. Elley et al. (2005) à leur tour critiquent la forte variabilité d'une année à l'autre constatée en Nouvelle-Zélande, qui met sérieusement en question l'utilisation des standards comme base décisionnelle dans la carrière scolaire des enfants. Ces deux constats montrent bien le processus cyclique du développement des standards qui, selon Müller et Silver (2006), « ne s'arrête jamais, car les évaluations doivent sans cesse être adaptées pour refléter le curriculum, les méthodes d'enseignement doivent être adaptées pour permettre la réussite du plus grand nombre d'élèves, les standards de contenu doivent être réactualisés, etc. » (p. 58).

On peut donc s'interroger sur la manière d'utiliser ces informations dans une logique de monitoring et, avant le lancement de tels travaux, il convient de bien définir les fonctionnalités évaluatives recherchées et de préciser les rapports de redevabilité.

### **Perspectives**

La demande politique pour des standards de formation ouvre donc un large débat qui, du point de vue scientifique, se heurte aux questionnements classiques de l'évaluation. Nous pouvons d'ores et déjà affirmer que les expériences faites nuancent la visée téléologique de ces dispositifs que certains interprètent comme une transformation de l'institution scolaire pour l'intégrer à la nouvelle économie du savoir. Néanmoins, le chantier ainsi ouvert offre un immense potentiel d'investigation qui, à terme, pourrait constituer une formidable base de connaissances empiriques, laquelle représentera une avancée significative pour la recherche éducative. Si les milieux de la recherche parviennent à contenir les dérives constatées ailleurs, si les travaux aboutissent à des modèles de compétence, si ces modèles permettent des mesures récurrentes de nos systèmes éducatifs, il deviendra possible de consolider nos connaissances sur le fonctionnement de ceux-ci, de questionner certaines propositions didactiques (Schneuwly, 2005) et d'identifier enfin les éléments du système qui apportent une réelle valeur ajoutée dans les processus d'apprentissage de nos élèves.