

ATELIER > L'évaluation de l'école par les standards internationaux

Un laboratoire de formation à petite échelle

Le cas de la Suisse

Matthis Behrens

IRDP de Neuchâtel

matthis.behrens@irdp.ch

Le système d'éducation de la Suisse est d'une complexité impressionnante du fait que les cantons confédérés restent souverains en matière d'éducation. Ainsi, on trouve non seulement quatre langues nationales, mais vingt-six systèmes cantonaux de formation différents avec une à trois langues de solarisation. Chacun des cantons dispose par conséquent de sa propre législation scolaire, de son administration et de son ministre de l'éducation. En d'autres termes, on retrouve au niveau de la Suisse un fédéralisme à très petite échelle, une construction politique fortement analogue à la construction européenne.

Le niveau fédéral, appelé Confédération, n'a pas de ministère de l'Éducation à proprement parler, mais plusieurs offices et secrétariats d'État qui s'occupent de quelques tâches fédérales : la formation professionnelle, les hautes écoles spécialisées, les écoles polytechniques fédérales, la politique universitaire et celle de la recherche. Néanmoins, les cantons ne travaillent pas en autarcie absolue. La Suisse connaît une collaboration intercantonale plus ou moins formalisée au sein d'une instance politique : la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP), qui réunit les 26 ministres responsables de l'éducation. La CDIP fonde son action sur des accords (ou concordats) intercantonaux juridiquement contraignants. Le premier date de 1970 et définit un premier cadre clairement délimité. La CDIP intervient également sur des dossiers que les cantons lui délèguent dans une logique de subsidiarité, c'est-à-dire qu'elle traite des dossiers que les cantons ou les régions ne peuvent ou ne veulent assumer seuls.

Durant les quinze dernières années, le système extrêmement décentralisé a subi de nombreuses pressions externes, qui ont déclenché plusieurs réformes se caractérisant par une remise en question de l'extrême décentralisation des systèmes de formation.

Mais c'est la première enquête PISA, avec des résultats médiocres en littératie, qui ébranle la crédibilité du système suisse de formation. Sans débattre ici dans le débat de la signification effective de ces résultats, rappelons que de nombreuses initiatives sont alors lancées, y compris par la CDIP qui propose un plan d'action général. Quelques années plus tard, le parlement fédéral, également appelé à se prononcer sur la situation de formation, soumet au vote deux articles constitutionnels sur la formation. Acceptés par le peuple en 2006, ils initient un bouleversement profond du système suisse de la formation. Les autorités publiques cantonales sont désormais tenues par la Constitution fédérale de régler de manière uniforme certains paramètres fondamentaux du système éducatif, au risque de se les faire imposer. La CDIP se doit donc de présenter un projet capable de modifier les structures cantonales, au risque de perdre sa crédibilité dans ce processus d'ajustement.

Elle propose un paquet de mesures appelé *HarmoS*, pour mettre en œuvre une vaste réforme des structures éducatives cantonales. Le projet définit les principaux paramètres du système : 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 deux idées radicalement nouvelles : l'harmonisation des objectifs de formation par des standards scientifiquement validés d'une part, et des plans d'études au niveau régional, de l'autre. Les performances de l'école obligatoire seront régulièrement évaluées dans le cadre d'un monitoring. La vérification de l'atteinte des standards nationaux de formation fera partie intégrante de cette évaluation (CDIP, 2008, articles 7, 8 et 10 du concordat *HarmoS*).

Cette contribution porte sur les standards nationaux de formation, une des composantes clés du projet. Ils sont prévus au terme des 2^e et 6^e années du cycle CITE 1, et de la 3^e année du cycle CITE 2A dans les disciplines langue 1, langues 2, mathématiques et sciences naturelles. Le but est de fixer les performances minimales que les élèves doivent atteindre dans chacun des systèmes éducatifs cantonaux.

L'élaboration des standards a été confiée fin 2004 à quatre consortiums de recherche nationaux regroupant des chercheurs provenant de toutes les régions linguistiques. Les travaux ont commencé début 2005 et les résultats ont été remis aux autorités fin 2008. Parallèlement, le texte du concordat a été mis au point et soumis pour ratification en mars 2008 aux 26 cantons. La question posée aux parlements cantonaux est de savoir si, en ratifiant le concordat, ils renoncent à une partie de leur souveraineté éducative en faveur de la CDIP. En Suisse romande, le concordat HarmoS est complété par une convention scolaire qui précise l'application des articles du concordat HarmoS et les complète, entre autres, par des dispositifs d'évaluation du travail des élèves (CIIP, 2007). À l'heure actuelle, la Suisse romande a d'ores et déjà réussi à rassembler le nombre nécessaire de cantons pour l'acceptation de HarmoS et de la Convention scolaire romande; les résultats sont plutôt positifs pour les cantons alémaniques.

Sur le plan conceptuel, le terme standard est un terme polysémique qui, selon les contextes et finalités, s'insère dans des dispositifs de nature très différents (Behrens, 2005; Berner et Stolz, 2005; Huber *et al.*, 2006). Le rapport Klieme (2004), texte de référence du projet distingue entre les « standards de contenu » (*content standards*) qui correspondent aux plans d'études et les « standards de performance » (*performance standards*) qui définissent les critères de suffisance de la performance de l'école et des élèves. Sur cette première distinction se greffent des précisions concernant le seuil de suffisance attendu. On entend par « standards minimums » ou « standards de base » la description d'un seuil au-dessous duquel aucun apprenant ne doit demeurer; les « standards réguliers » précisent un niveau qui doit être atteint par la moyenne d'une population scolaire; et les « standards maximums » définissent un idéal éducatif.

Les standards suisses complètent un large éventail de dispositifs déjà existants: plans d'études cantonaux, moyens d'enseignement cantonaux, voire régionaux, évaluations bilans et évaluations cantonales des élèves pour mentionner les plus importants. On peut donc s'interroger sur l'articulation entre standards et plans d'études. La Suisse choisit, pour des raisons pragmatiques que ce seront les standards qui conditionneront les plans d'études, en raison d'un paysage éducatif peu coordonné.

L'approche est inhabituelle, car la littérature insiste sur la nécessité d'une bonne articulation entre les standards et les plans d'études, ou du moins d'un consensus fort sur les contenus d'enseignement. Les standards doivent donc découler des finalités éducatives qui doivent être légitimées par un débat public, et non être le résultat d'avis d'experts. Cela dit, on observe néanmoins le développement indépendant de standards à côté des curricula existants lorsqu'il est possible de développer un modèle de compétences, qu'il dépende ou non du curriculum. Ce dernier cas de figure a été testé dans le cadre du projet allemand VERA (*Vergleichsarbeiten* – travaux comparatifs) qui mise directement sur le développement d'une banque d'items pour l'évaluation des élèves au terme du 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 2004, p. 135) sur la base des données empiriques récoltées et organisées en modèle de compétences. Cette démarche s'inspire des approches méthodologies mises en œuvre dans le cadre des grandes enquêtes internationales, TIMSS, PISA et PIRLS.

Des experts de toutes les régions linguistiques du pays engagés dans les consortiums nationaux développent donc des modèles de compétences pour chaque discipline, articulant des activités cognitives avec les sous-domaines de la discipline. Cette façon de procéder permet de contourner habilement l'épineuse question d'un accord consensuel quant aux finalités et aux objectifs des différents systèmes éducatifs cantonaux, en particulier des cantons alémaniques. C'est la raison pour laquelle je pense que la démarche pourrait se révéler un modèle opératoire pour la construction européenne.

La CDIP s'inspire des travaux du Conseil d'Europe pour le *Cadre européen commun de référence pour les langues*, estimant qu'il est possible de le transposer aux disciplines prévues par le projet (L1, L2, mathématiques, sciences naturelles). En fixant ainsi les performances attendues à la fin des cycles scolaires, la CDIP cherche à imposer une norme de performance afin d'obtenir un effet harmonisant indirect sur les curricula cantonaux. Elle choisit des standards minimums. Ils doivent être contraignants, c'est-à-dire que les acteurs du système scolaire sont tenus d'amener toute leur population d'élèves à la maîtrise d'un socle minimal de compétences qui sera vérifié par des épreuves qui s'y rattacheront. Autrement dit, les items de test et le taux de réussite vont illustrer concrètement ce qui est attendu, mais leur transposition en objectifs de réussite sera du ressort du politique.

Sur le plan méthodologique, plusieurs observations peuvent être faites. Comme suggéré plus haut, un standard n'est pas un objet isolé. Il est d'abord indicateur de suffisance et ensuite point de départ pour des évaluations très différentes qui permettent soit de piloter le système, soit de réguler les apprentissages.

La fixation des seuils est un problème méthodologique délicat qui n'a pas été traité avec suffisamment d'attention. Selon Blais (2008), il se construit en étalonnant des valeurs subjectives sur une échelle de mesure. Le détour par un modèle de compétence ne confie pas plus d'objectivité à cette fixation normative, si ce n'est que l'analyse des compétences à évaluer paraît plus systématique.

Dans le projet HarmoS, un modèle de compétence est une grille bidimensionnelle organisant une discipline scolaire en sous-domaines et aspects cognitifs sur l'autre. Les compétences sont inférées à partir d'items de test placés dans la grille ainsi obtenue. L'utilisation de la théorie de réponse aux items (*item response theory* – IRT) a permis de construire des items distinctifs sur une échelle de difficulté permettant de prédire la probabilité avec laquelle un candidat est capable de répondre correctement à des items de difficulté similaire. Chaque item se positionne donc sur une courbe représentant les réussites de la population étudiée. C'est ainsi qu'émerge la troisième dimension du modèle qui donne des niveaux de performances à l'instar des niveaux de difficultés dégagés dans l'enquête PISA.

Afin de renforcer ces modèles, d'importants travaux de validation ont été réalisés : 12 000 élèves de 6^e et de 9^e année ont été testés dans quatre disciplines. Cependant, les résultats obtenus montrent qu'il n'était possible que de vérifier les propriétés métriques des items et de les hiérarchiser sur une courbe de difficulté progressive. Dans ce sens, les compétences postulées et la description des disciplines en sous-domaines et en aspects cognitifs restent invalidées.

Par ailleurs, la stabilité du modèle à travers les différentes cultures linguistiques demeure faible, à cause d'une sous-estimation de l'investissement à fournir pour tenir compte d'une situation multilingue. Des travaux supplémentaires importants sont nécessaires pour déterminer la signification réelle des performances observées. Un échec aux items présentés révèle-t-il nécessairement l'absence de la compétence contrôlée ? Peut-on contrôler une acquisition particulière indépendamment d'un contexte d'apprentissage ? Voici des questions que posait déjà Cardinet (1988), alors confronté à l'évaluation dans le cadre de la pédagogie de la maîtrise.

Par ailleurs, la double finalité visant l'harmonisation curriculaire et la mesure des performances introduit un brouillage fonctionnel avec de fortes répercussions méthodologiques. J'estime le premier objectif inatteignable, car, si l'on postule la primauté des standards sur les curricula, tout en attendant un effet d'harmonisation sur les objectifs de formation, au moins l'une des deux conditions doit être remplie pour qu'ils puissent répondre à la demande. Ou bien l'univers des standards est au moins équivalent à la somme des objectifs de formation, ou bien les standards se reportent à des compétences et des contenus que l'on juge centraux et prédictifs de la future progression des apprentissages. Or aucune des deux exigences ne semble véritablement remplie (Behrens 2008).

Les standards ont comme particularité de se limiter à une partie du curriculum ou d'une discipline en faisant « ressortir clairement les principes fondamentaux de la discipline et de la branche d'enseignement » (Klieme 2004, p. 22). Se pose tout de suite la question de la mesurabilité de ces éléments. Ensuite, l'exigence de standards insensibles aux variations linguistiques conduit à l'élimination de nombreux items et resserre davantage le champ des compétences importantes et mesurables. Les modèles de compétences en filigrane des standards sont donc d'abord des modèles de mesure et n'incluent pas suffisamment des considérations didactiques ou psychopédagogiques qui auraient peut-être permis d'identifier de tels principes fondamentaux. Des travaux supplémentaires seraient nécessaires, validant la valeur cognitive des compétences du modèle. On voit donc mal comment les modèles de compétences développés pourraient servir directement à l'harmonisation des objectifs, si ce n'est en remplaçant la normativité du curriculum par celle des modèles de compétences.

Il est intéressant de noter qu'en langue de soclarisation, des domaines tels que la structuration du langage, la grammaire et l'orthographe n'ont pas été traités initialement à cause des variations linguistiques existantes. Or cette option est inacceptable pour la dimension curriculaire du projet. De ce fait, les modèles de compétence ont du être complétés par des compétences non-validées et différentes selon les régions linguistiques.

L'effort nécessaire pour le développement de standards nationaux valables pour trois régions linguistiques a donc été fortement sous-estimé. Une application rigoureuse des critères d'acceptation des items de tests aurait même signifié l'échec de la validation nationale. En revanche, le fait d'avoir passé par une validation confie aux modèles de compétences et aux standards une légitimité symbolique, feignant l'existence d'une mesure « naturelle » dans laquelle la compétence deviendrait l'unité de mesure, un référent objectif et stable, comme un mètre. Ce processus que j'ai appelé une « naturalisation de la mesure » (Behrens 2008) interpelle sur l'utilisation des résultats de la recherche, dès lors que cette dernière est trop fortement instrumentalisée par le politique.

Sur le plan de l'utilisation des standards, les dispositifs connexes n'existent pas encore et sont en train d'être planifiés. Le texte du concordat se limite à rappeler l'alignement entre standards, plans d'études et moyens d'enseignement et d'établir le lien entre les standards et un projet de monitoring qui portera sur l'évaluation du système, mais dont les contours restent à définir.

Pour le moment, les interrogations se posent avec la formulation définitive des standards destinés à une large consultation. Si l'on se tenait exclusivement à l'objectif d'évaluation du système, l'on pourrait se contenter d'un score global, tel que le produit l'enquête PISA. Cependant, sous cette forme, les indicateurs perdraient leur sens pour les acteurs directement concernés, les enseignants, les directions d'école et/ou les corps d'inspection, les élèves et les parents. Si l'on se tenait exclusivement à l'objectif d'harmonisation curriculaire, l'on devrait formuler de manière détaillée chaque standard. Toutefois, sous cette forme, l'édifice deviendrait trop complexe pour une consultation politique et il faudrait craindre un glissement de l'évaluation système à l'évaluation de l'élève. La CDIP privilégie la deuxième variante et essaye de formuler les standards de la même manière pour les quatre disciplines, malgré l'inconsistance des modèles.



L'alignement demandé entre standards, plans d'étude, moyens d'enseignement et épreuves à l'intention des élèves a d'ores et déjà provoqué une multitude d'initiatives tout à fait nouvelles. En Suisse allemande, 22 cantons ont enfin réussi à lancer un projet de plan d'études commun, une première dans les annales éducatives suisses. En Suisse romande, où les travaux sur le plan d'études romand sont quasiment terminés, la question se pose de savoir comment vérifier l'alignement de cet ouvrage avec les standards nationaux. En même temps, les ministres de l'éducation des cantons romands ont décidé de développer des épreuves de référence romandes pour vérifier l'atteinte des objectifs du plan d'étude. Ces travaux sont en cours. Ils nécessitent une coordination accrue entre différentes pratiques cantonales qui, à chaque moment, relancent le débat sur la souveraineté des cantons. ■

BIBLIOGRAPHIE

BEHRENS M. (éd.) (2005) : *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 Harnos*, Neuchâtel : IRDP.

BEHRENS M. (2008) : « Les standards de formation, l'anneau de Moebius de l'école suisse ? », *Revue internationale d'éducation de Sèvres* (pp. 127-136), n° 48 (L'école et son contrôle), septembre 2008.

BERNER E., STOLZ S. (2005) : Literaturanalyse zu Entwicklung, Anwendung und insbesondere Implementation von Standards in Schulsystemen: Nordamerika: Im Auftrag der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren EDK http://www.cdep.ch/PDF_Downloads/Harnos/Literaturanalyse_3.pdf (pages consultées en mai 2006).

BLAIS J.G. (2008) : « Les standards de performance en éducation », *Mesure et évaluation en éducation*, 2008, vol. 31, n° 2, pp. 93-105

CARDINET J. (1988) : « La maîtrise, communication réussie », dans Huberman, M. (Éd), *Assurer la réussite des apprentissages scolaires*, Delauchaux et Niestlé, Neuchâtel (pp. 127-154).

CDIP (2007) : *Accord intercantonal du 14 juin sur l'harmonisation de la scolarité obligatoire (Concordat HarmoS)* http://www.edudoc.ch/static/web/arbeiten/harnos/harnos_dfi.pdf (consulté le 18 juillet 2007).

CIIP (2007) : *Convention scolaire romande, texte adopté par la CIIP le 21 juin 2007*, Neuchâtel, http://www.ciip.ch/pages/actualite/fichiers/CSR_210607.pdf (consulté le 18 juillet 2008).

HELMKE A., HOSENFELD I. (2004) : Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven. In M. Wosnitza, A. Frey & R.S. Jäger (Hrsg.), *Lernprozess, Lernumgebung und Lerndiagnostik: Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (pp. 56-75). Landau: Empirische Pädagogik.

KLIEME E. et al. (2004) : *Le développement de standards nationaux de formation : une expertise*. Bonn : Ministère fédéral de l'éducation et de la recherche (BMBF).

MÜLLER K., SILVER R. (2006) : « Standards in Education – Review of US Literature ». In M. Behrens (éd.), *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif* (pp. 9-68). Neuchâtel : IRDP.