L'évaluation de l'école par les standards internationaux

Evaluating schools by international standards

La evaluación de la escuela por las normas internacionales

Jean-Marie De Ketele

Édition électronique

URL : <https://journals.openedition.org/ries/743>

DOI : 10.4000/ries.743

ISSN : 2261-4265

Éditeur

France Education international

Édition imprimée

Date de publication : 1 décembre 2009

Pagination : 69-76

ISBN : 978-2-8542-0577-0

ISSN : 1254-4590

Référence électronique

Jean-Marie De Ketele, « L'évaluation de l'école par les standards internationaux », *Revue internationale d'éducation de Sèvres* [En ligne], 52 | décembre 2009, mis en ligne le 01 décembre 2012, consulté le 26 avril 2022. URL : <http://journals.openedition.org/ries/743> ; DOI : <https://doi.org/10.4000/ries.743>

L'évaluation de l'école par les standards internationaux¹

Jean-Marie De Ketele

Selon le *Dictionnaire de l'éducation* (van Zanten 2008), le concept de standard prend son origine en 1862, lors du Code révisé dans l'enseignement anglais, et désigne « chaque degré d'aptitude reconnu, testé par un examen, d'après lequel les élèves étaient classés » (p. 634), tandis qu'aujourd'hui, « les standards scolaires sont des objectifs relatifs aux savoirs que les enseignants visent à faire atteindre à leurs élèves » (p. 634). Dans son article de synthèse sur les standards de performances en éducation, Blais (2008) souligne une différence fondamentale entre le sens originel du concept selon qu'il est utilisé par les anglo-saxons et les francophones. Chez les premiers, il désigne « le profil de performance de l'élève idéal » ; chez les seconds, « il sert à qualifier des produits, des situations, des processus normaux dans le sens de ce qui est correct ou adéquat pour un contexte donné (un prix standard, le français standard, une performance standard, etc. » (p. 93). Comme d'autres concepts, celui de compétence, par exemple, est torturé dans divers sens selon les lieux et les personnes qui en parlent. Ne citons que deux exemples tirés de la *Revue internationale d'éducation* sur le thème « Que savent les élèves ? » (2006, n° 43). Parlant du cas des États-Unis, Reville (2006) affirme que les standards de formation sont « des référentiels clairs en matière de connaissances et de compétences... dont le niveau est fixé », tandis que Sroka *et al.* (2006), parlant du cas de l'Allemagne, parlent de « profils de compétences, qui impliquent chaque fois une construction de la compétence ancrée simultanément dans la pédagogie d'une discipline et dans une progression ».

Dans son *Dictionnaire actuel de l'éducation*, Legendre (1993) pointait déjà ce qui apparaît comme le dénominateur commun à l'ensemble des sens attribués au concept en affirmant que le standard est « ce qui est conforme à une norme ». Le standard est présenté, à tort ou à raison, comme une « norme » et

---

1. Cet article prend sa source dans l'atelier que nous avons coordonné au CIEP le 13 mars 2009 sur le thème « L'évaluation de l'école par les standards internationaux » et où sont intervenus Matthias Behrens (Suisse), Jean-Pierre Jarousse et Jean-Claude Mantes (Afrique), Nathalie Mons (comparaison européenne), Marcos Reigota (Brésil) et Thierry Rocher (France). Les communications disponibles peuvent être consultées en ligne sur : <http://www.ciep.fr/ries/colloque-2009/contributions.php/>. Cet article n'est pas un résumé, mais une synthèse personnelle intégrant leurs propos et le fruit de notre réflexion sur la base de notre expérience, de nos travaux et de nos lectures.

l'adjectif « conforme » véhicule bien des connotations comme en témoigne le tableau suivant.

Les connotations véhiculées par le concept de standard

<i>Standard</i>	<i>versus</i>	<i>Spécifique</i>
Conforme		Différent
Uniforme		Multiforme
Universel		Contextualisé
Homogène		Hétérogène
...		
<i>Normes</i>	<i>versus</i>	<i>Orientations</i>
Indicateurs		Critères
Contrôle		Production de sens
Sanction		Valorisation
Classement		Etape du développement
Maîtrise		Développement
Produit		Processus
...		

Plus que de longs discours, l'examen de ces connotations nous amène à nous poser une série de questions fondamentales. Quels standards ? Que ne trouve-t-on pas (et pourquoi ?) dans les standards proclamés, dans les évaluations effectuées ? Qui participe à l'identification des standards et à quel niveau : local, national, régional, international ? Comment se fait le passage du standard aux questions d'évaluation ? Comment tient-on compte des cultures et des contextes ? Quels usages fait-on des standards et des évaluations ?

QUELS STANDARDS ?

Force de pression importante sur les systèmes éducatifs, l'OCDE a défini, avec l'appui de nombreux experts et techniciens, deux types de standards : les « standards de contenus » et les « standards de performance ». La désormais célèbre épreuve internationale PISA distingue trois grands types de standards de contenus : la compréhension de l'écrit ; la culture mathématique ; la culture scientifique. Il serait faux de croire que les responsables des enquêtes internationales ont accordé peu d'importance à la question du choix de ce qui doit être évalué, comme le montrent Baye, Demonty, Fagnant, Lafontaine, Matoul et Monseur (2004). Ainsi, l'évaluation de « la culture mathématique chez les jeunes de quinze ans » est définie à travers la combinaison de trois dimensions : la complexité conceptuelle, le contenu à mobiliser et les réalités ou situations envisagées. La complexité conceptuelle est envisagée à travers trois groupes de compétences de plus en plus complexes : la reproduction de connaissances acquises ; les connexions à établir entre l'énoncé d'un problème et les connaissances mathématiques pour traduire le problème de départ en un

problème mathématique ; la « réflexion élaborée » qui implique des capacités de modélisation, d'argumentation, d'abstraction et de généralisation. Cette dimension « groupes de compétences » va être croisée avec quatre types potentiels de contenus : le domaine des « quantités », celui de « l'espace et des formes », celui de « l'incertitude » (les statistiques et les probabilités) et celui des « variations et relations ». Ces deux dimensions seront aussi combinées avec des situations issues de quatre champs : les réalités personnelles de l'élève ; le champ éducatif ou professionnel ; les réalités dites publiques ; le champ des réalités scientifiques. Cet aperçu montre bien que la conception des items du test PISA repose sur un cadre conceptuel élaboré. Il en va de même pour chacune des « disciplines » évaluées : la résolution de problèmes, la culture scientifique, la compréhension de l'écrit. Ainsi, pour donner un second exemple, la compréhension de l'écrit repose également sur trois dimensions à combiner : le type de processus de lecture dont les modalités sont « retrouver des informations », « interpréter le texte », « réfléchir sur le contenu du texte » ; le type de textes, à savoir « les textes continus » et « les textes non continus » ; l'usage pour lequel le texte a été conçu, à savoir « à des fins personnelles », « à usage public », « à usage éducatif » et « à usage professionnel ». Quant aux standards de performances, ils sont conjugués en six niveaux après une transformation mathématique afin de rendre, autant que possible, les résultats comparables.

Le travail effectué par les experts de l'OCDE, fruit d'une grande technicité, donne l'impression de sérieux et de rigueur. Que penser alors du travail effectué par le système éducatif japonais qui semble d'une nature bien différente

Les standards selon le système éducatif japonais

Axes de la « force de vivre »	Piliers de chaque axe
1. L'initiative et l'autonomie de la vie en général	1. L'amour propre 2. La santé et la capacité physique 3. La capacité de prendre des décisions et de faire des choix 4. La capacité à planifier son avenir, à saisir son rôle, à approfondir ses connaissances et à passer à l'exécution
2. La relation entre soi et les autres dans la vie privée et à l'école	1. La concertation, le savoir partager avec les autres 2. La capacité de s'entendre, de communiquer 3. Le savoir coordonner les intérêts des uns et des autres en respectant les règlements, la justice et l'équité
3. La relation entre soi et la société dans la communauté et sur le lieu de travail	La participation et la contribution sociale La sensibilité, la créativité et l'appréciation de l'art et de la nature La maîtrise de la langue et de l'informatique, la maîtrise des techniques et des connaissances L'assiduité au travail, la compréhension de la culture nationale et des autres cultures La capacité de découvrir, de résoudre le problème, de planifier, de pratiquer, d'évaluer

Source : Soyonama, 2006

(Soyonama, 2006) ? Dans le Rapport de Chukyoshim du 26 octobre 2005, on y définit les standards comme suit : « Les connaissances et les contenus que les élèves doivent acquérir et qui seront évalués par le test national à l'âge de douze et quinze ans seront structurés à partir de l'objectif éducatif de la force de vivre, c'est-à-dire l'implantation de la compétition basée sur les connaissances. Les connaissances seront structurées en trois grands axes : l'initiative et l'autonomie de la vie en général ; la relation entre soi et les autres dans la vie privée et à l'école ; la relation entre soi et la société dans la communauté et sur le lieu de travail ».

L'analyse des standards selon l'OCDE et selon le système éducatif japonais montre bien deux philosophies profondément différentes aboutissant à des produits différents, à des évaluations de nature et de formes différentes et, sans doute aussi, à des interprétations différentes de certains produits opérationnels qui pourraient paraître semblables. On pourrait faire état d'autres différences liées aux contextes : de plus en plus, les pays en voie de développement accordent une importance grandissante aux « compétences de vie » (*life skills*), comme la prévention du sida, la promotion de l'allaitement maternel, le traitement de l'eau, etc. (l'UNICEF parle des douze comportements prioritaires à installer dans les pays en développement dans le curriculum scolaire). C'est pourquoi la comparaison des systèmes éducatifs est si enrichissante : elle nous révèle non seulement les standards communs à plusieurs systèmes, mais aussi incommensurables les standards que l'on ne retrouve pas dans nos systèmes éducatifs.

LES STANDARDS PROCLAMÉS SONT-ILS ÉVALUABLES DANS LES ÉVALUATIONS EXTERNES ?

Dans l'article « Mais qu'évaluent donc les enquêtes internationales ? » (De Ketele, 2006), nous introduisons notre propos par ces mots : « Comme le souligne Lise Demailly dans ses travaux récents (2006a et 2006b), les offreurs d'évaluation se sont multipliés, tant au niveau national qu'international. Elle en conclut avec raison qu'il est urgent d'évaluer ces évaluations ». Il est intéressant de relever que les enquêtes internationales de l'*International Association for Educational Achievement* (IEA) se référaient à un cadre de référence qualifié de « dénominateur commun » aux programmes des pays soumis aux tests. L'expérience a montré combien il était difficile de se mettre d'accord sur un tel dénominateur et que cela conduisait parfois à la non prise en considération de sections entières dans certaines disciplines, comme les sciences. Le cas de la Suisse rapporté par Behrens lors de l'atelier « L'évaluation de l'école par les standards internationaux » est particulièrement éloquent et intéressant, car la Suisse est une fédération composée de 26 cantons, de 26 ministères de l'Éducation mais aucun ministère fédéral de l'Éducation, de 26 cadres législatifs différents et de quatre langues, dont trois de scolarisation. Suite à la grande diversité des

résultats obtenus lors de la première enquête PISA, le besoin d'harmoniser les systèmes éducatifs est apparu nécessaire. Le projet HarmoS était né avec comme objectifs, notamment, « l'harmonisation des objectifs de formation par des standards scientifiquement validés d'une part et des plans d'études au niveau régional d'autre part », ainsi que la construction d'épreuve pour « la vérification de l'atteinte des standards nationaux de formation » (Behrens, 2008). À la distinction énoncée plus haut (standards de contenus et standards de performances), le projet greffe une distinction issue d'une réflexion sur la notion de seuil : les « standards de base ou minimum » sont établis sur la base d'un seuil au dessous duquel les élèves ne peuvent rester dans le système en question ; les « standards réguliers » précisent un niveau qui doit être atteint par la moyenne d'une population scolaire donnée ; les « standards maximum » définissent un idéal éducatif. C'est donc au départ un projet d'autant plus ambitieux qu'il veut s'appuyer, au plan méthodologique, sur la théorie de réponse à l'item (TRI), outil mathématique sophistiqué qui sert de référence dans les grandes enquêtes internationales pour valider les épreuves d'évaluation. Que retirer de la mise en œuvre de ce vaste projet ? Behrens rapporte que, malgré les travaux importants de validation (douze mille élèves de 6^e et de 9^e), il n'a pas été possible de vérifier les propriétés métriques et les courbes de difficultés attendues par le modèle de validation (TRI) pour une majorité d'items ; d'autre part, 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. Ceci nous montre une nouvelle fois les limites d'un tel modèle de validation. De Ketele et Gérard (2005) avaient déjà attiré l'attention sur le fait que, si ce modèle convenait bien pour valider des items relatifs à la maîtrise de connaissances et de savoir-faire de base (désignés par le vocable de « ressources »), il devenait hasardeux de l'utiliser pour valider des épreuves ou parties d'épreuves portant sur des compétences. En effet, toute compétence est par nature multidimensionnelle ; or le modèle TRI repose sur le postulat de l'unidimensionnalité de l'objet évalué, comme le rappelait l'exposé de Rocher. Et ce n'est pas un hasard si l'analyse des contenus des épreuves internationales révèle le plus souvent que seules les ressources sont évaluées et non les compétences (mobilisation de ressources en situation complexe). Or les standards sont formulés davantage en termes de compétences que de simples ressources !

QUELS SONT LES ACTEURS IMPLIQUÉS ?

L'identification des standards est de plus en plus souvent confiée à des experts et passe maintenant par des organisations ou des consortiums internationaux, même si ceux-ci prennent souvent le soin de consulter des panels d'utilisateurs. Il faut d'ailleurs remarquer deux phénomènes paradoxaux : les standards de contenus tendent à être formulés de façon générale (par exemple :

la compréhension de l'écrit, l'utilisation des TIC, le travail individuel autonome et le travail en équipe...) ; les standards de performance tendent au contraire à être définis en termes de formules complexes au point de nécessiter un glossaire et un mode opératoire détaillé. Les premiers sont donc polysémiques et sujets à des compréhensions multiples ; les seconds deviennent monosémiques au point de poser des difficultés si l'on veut les faire évoluer, tout en restant opaques pour ceux qui ne disposent pas des glossaires et des modes opératoires. Les premiers sont ouverts facilement à des nécessités de contextualisation ; les seconds sont plus difficilement contextualisables, au point qu'un même résultat peut avoir des sens foncièrement différents selon les contextes d'application.

La conception des épreuves standardisées internationales et nationales sont également le fait d'experts, certes issus de différents secteurs, mais où la place prise par les spécialistes de la mesure et du traitement des données devient de plus en plus importante car ils sont peu nombreux à maîtriser ces deux champs. On conçoit donc que les enseignants, pratiquant librement différentes formes et fonctions de l'évaluation dans leurs classes, se sentent dépossédés de l'évaluation ou s'en méfient, les dénigrent ou même les ignorent. Ceci est d'autant plus vrai qu'ils ont l'impression que les épreuves standardisées proposées s'écartent de leurs pratiques ou leurs cultures professionnelles.

L'utilisation des résultats est sans doute la question la plus sensible. Comme le montre Mons (2008) et plusieurs intervenants lors de l'atelier, une culture politique de l'évaluation se développe de plus en plus :

- par souci des autorités politiques et éducatives de se situer par rapport à des standards internationaux, mais surtout par rapport à d'autres pays (les pays européens veulent occuper les premières places dans les classements ; les pays en développement ou émergents veulent se rapprocher des pays européens) ;
- par souci d'asseoir la légitimité des politiques prescrites par les résultats des évaluations internationales ou nationales, lorsque cela arrange les autorités politiques qui les ont prescrites ou tentent de les prescrire ;
- en raison de la tendance à responsabiliser les enseignants et les acteurs locaux et à instaurer des procédures de reddition de comptes (montée en puissance de l'évaluation par les résultats plutôt que par les processus mis en œuvre, évaluation des établissements, évaluation des enseignants) ; mais tendance aussi de nombreux établissements à « monter dans les classements » par recrutement d'une certaine catégorie d'élèves (avec l'appui de certaines catégories de parents) et élimination d'autres catégories d'élèves (utilisation du redoublement et de stratégies d'orientation).

C'est donc plus une culture politique de l'évaluation qu'une culture pédagogique de l'évaluation. Cette dernière signifierait que l'évaluation sert à : montrer ce qu'il faut apprendre et pourquoi (les épreuves devraient donc évaluer ce que l'on désire faire apprendre, comme les standards qui sont le plus souvent des compétences complexes, et pas seulement ce que l'on sait évaluer avec les

modèles de validation) ; indiquer comment enseigner et faire apprendre ; repérer et promouvoir des pratiques obtenant des plus-values dans des contextes particuliers ou avec des publics spécifiques. En effet, si penser globalement est nécessaire, agir localement est la seule façon de faire évoluer les systèmes éducatifs. Dans son intervention lors de l'atelier, Thierry Rocher faisait état, à partir de son expérience française, de la nécessité de promouvoir des évaluations nationales (et même régionales) à côté des évaluations internationales, non seulement en retravaillant les données des élèves français aux épreuves internationales à partir de certaines hypothèses de travail, mais aussi en créant d'autres protocoles ou d'autres formes d'évaluation.

L'attention aux usages contextualisés de l'évaluation et à l'agir local est l'une des préoccupations centrales des interventions de Jarousse et Mantes, faisant état de la situation en Afrique, ainsi que de Reigota, parlant du Brésil. Dans de tels pays où les systèmes éducatifs sont moins ou très peu « normés » sur le plan de leur organisation (mode de groupement des élèves, taille des classes, formation et accompagnement des enseignants, volume et calendrier d'activité...), les évaluations externes (peu ou moins nombreuses que dans les pays du Nord) devraient conduire, si l'on en reste qu'aux résultats globaux, à une gestion reposant sur des arbitrages d'une grande brutalité, car les contraintes ne permettent d'agir que sur l'un ou l'autre facteur au détriment des autres. Dans ces pays, les activités d'évaluation standardisées ont encore moins d'impact sur le quotidien des acteurs locaux. Il est donc plus que nécessaire d'accorder une place à une évaluation qui s'intègre aux pratiques courantes des équipes enseignants, des chefs d'établissement, des accompagnateurs locaux, et qui accorde une place au questionnement, par les acteurs locaux, des indicateurs de valeur ajoutée adéquats au contexte et des relations entre les plus-values observées et les facteurs explicatifs plausibles.

N'y a-t-il pas là des indications à méditer pour les systèmes éducatifs du Nord ?

BIBLIOGRAPHIE

BAYE A., DEMONTY I., FAGNANT A., LAFONTAINE D., MATOULA A. & MONSEUR Ch. (2004) : « Les compétences des jeunes de 15 ans en Communauté française de Belgique en mathématiques, en lecture et en sciences », *Les Cahiers du service de pédagogie expérimentale*, n° 19 et 20, 1-224.

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

BEHRENS M. (2009) : « Un laboratoire de formation à petite échelle. Le cas de la Suisse ». *Revue internationale d'éducation de Sèvres* [www.ciep.fr/ries/colloque-2009/contributions.php/atelier-B.php].

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

DE KETELE J.-M. (2006) : « Mais qu'évaluent donc les enquêtes internationales ? », *Revue internationale d'éducation de Sèvres*, n° 43, 141-157.

DE KETELE J.-M. & GÉRARD F.-M. (2005) : « La validation des épreuves d'évaluation selon l'approche par compétences », *Mesure et Évaluation en Éducation*, vol. 28, n° 3, 1-26.

DEMAILLY L. (2006a) : « En Europe : l'évaluation contre la crise des systèmes scolaires, l'évaluation en crise ». *Éducation et Sociétés*, n° 17, 105-120.

DEMAILLY L. (2006b) : « La nécessaire méta-évaluation ». In : G. Figari & L. Mottier-Lopez (Eds). *Recherche sur l'évaluation*. Paris : L'Harmattan, pp. 203-210.

JAROUSSE J.-P. (2009) : De l'évaluation externe des systèmes éducatifs à la gestion locale de la qualité des apprentissages. Un regard sur les systèmes éducatifs africains. *Revue internationale d'éducation de Sèvres* [www.ciep.fr/ries/colloque-2009/contributions.php/atelier-B.php].

LEGENDRE R. (1993) : *Dictionnaire actuel de l'éducation*. Montréal/Paris : Guérin/Eska (2^e édition).

MONS N. (2008) : « Autonomie et contrôle des enseignants. Une enquête dans les pays européens », *Revue internationale d'éducation de Sèvres* n° 48, 109-117.

REIGOTA M. (2009) : « Les élèves brésiliens ne savent-ils donc rien ? ». *Revue internationale d'éducation de Sèvres* [www.ciep.fr/ries/colloque-2009/contributions.php/atelier-B.php].

REVILLE S.P. (2006) : « Les standards de formation. Le cas des États-Unis ». *Revue internationale d'éducation de Sèvres*, n° 43, 23-32.

ROCHER T. (2009) : « Nécessité et complémentarité des évaluations internationales et nationales. Le cas de la France ». *Revue internationale d'éducation de Sèvres* [www.ciep.fr/ries/colloque-2009/contributions.php/atelier-B.php].

SOYONAMA D. (2006) : « Une politique volontariste des savoirs et des compétences basées sur l'évaluation ». *Revue internationale d'éducation de Sèvres*, n° 43, 47-55.

SROKA W., ABS H.J., STECHER L. (2006) : « Que savent les élèves en Allemagne ? », *Revue internationale d'éducation de Sèvres*, n° 43, 77-90.

VAN ZANTEN A. (2008) : *Dictionnaire de l'éducation*. Paris : Dunod.