

## Les politiques d'évaluation en éducation. Et après?

Dany Laveault

Volume 42, numéro 3, automne 2014

Les politiques d'évaluation en éducation. Et après?

URI : <https://id.erudit.org/iderudit/1027402ar>

DOI : <https://doi.org/10.7202/1027402ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association canadienne d'éducation de langue française

ISSN

1916-8659 (numérique)

[Découvrir la revue](#)

Citer ce document

Laveault, D. (2014). Les politiques d'évaluation en éducation. Et après? *Éducation et francophonie*, 42(3), 1–14. <https://doi.org/10.7202/1027402ar>

# Les politiques d'évaluation en éducation. Et après?

**Dany LAVEAULT**

Université d'Ottawa, Ontario, Canada

## Buts et conséquences des politiques: des résultats pas toujours au rendez-vous

L'éducation occupe une place importante dans nos sociétés et ses coûts figurent, avec ceux de la santé, parmi les dépenses les plus importantes des États modernes. Il n'est donc pas étonnant qu'une organisation internationale telle que l'OCDE se soucie d'évaluer l'efficacité des systèmes éducatifs de ses pays membres, et plus particulièrement l'efficacité de leur cadre d'évaluation (OCDE, 2009, 2013). Sans évaluation, l'éducation est un navire sans pilote. Sans évaluation valide, les systèmes d'éducation sont à risque de se méprendre quant au degré d'atteinte de leurs objectifs ainsi que sur l'efficacité des moyens mis en œuvre (Laveault, 2009a).

Au cours des vingt dernières années, deux principaux phénomènes ont accru l'impact de l'évaluation dans nos systèmes d'éducation. Le premier est la participation d'un nombre toujours plus grand de pays aux enquêtes internationales sur le rendement des élèves (PISA, TIMSS, PIRLS). Le second est la prise en compte grandissante du rôle de l'évaluation non seulement pour informer sur le rendement des élèves, mais aussi pour former ces derniers. L'évaluation formative s'est vue ainsi introduite comme composante nécessaire de toute politique d'évaluation visant, d'une part, à certifier les apprentissages des élèves et, d'autre part, à soutenir les enseignants, les élèves et les autorités scolaires dans leurs prises de décision.

Dans le cas des enquêtes internationales, les résultats sont largement diffusés auprès du public et font l'objet d'études approfondies tant par les chercheurs que par les autorités scolaires. Pour plusieurs pays, ces études comparatives du rendement

des élèves ont eu pour effet de remettre en question l'efficacité de leur système d'éducation. De plus, à l'intérieur de chaque territoire scolaire (districts, cantons, États ou provinces), il s'est avéré que le rendement des écoles ou des instances locales pouvait varier considérablement, toutes choses étant égales par ailleurs. Cela a fait en sorte que de nombreuses organisations scolaires ont mis en place leur propre système d'évaluation à grande échelle afin de faire une surveillance rapprochée du rendement de leur population scolaire.

Dans ce nouveau contexte d'évaluations extérieures à la salle de classe, le rôle des enseignants en matière d'évaluation s'est trouvé transformé, et ce, pour deux principales raisons. La première est la reconnaissance par la communauté scientifique de la contribution majeure de l'évaluation pour soutenir l'apprentissage des élèves (Black et Wiliam, 1998; Wiliam et Thompson, 2008; Wiliam, 2011; Scallon, 2000). Quoique l'ampleur de l'effet rapporté puisse varier selon les études, cet effet n'en demeure pas moins considérable, d'où l'importance que l'évaluation formative a prise peu à peu dans les politiques d'évaluation. De pratique facultative et artisanale qu'elle était à l'origine, l'évaluation formative est devenue une composante essentielle de la planification de l'enseignement et d'une véritable politique d'évaluation. La seconde raison est probablement plus subtile. Comparativement aux évaluations externes et uniformes, les évaluations sommatives réalisées par les enseignants – celles qui ont pour objectif de dresser le bilan des apprentissages et éventuellement de les certifier au moyen d'un document officiel tel qu'un bulletin – sont trop différentes et leurs niveaux d'exigence trop variables pour renseigner sur l'efficacité des systèmes d'éducation. C'est ainsi que les évaluations externes ont fait en sorte de diminuer le rôle des enseignants en matière d'évaluation sommative, alors que les attentes à l'égard de leur rôle lié à l'évaluation formative devenaient plus élevées.

Ce réaménagement des rôles n'est pas terminé. Il se traduit par des tensions qui peuvent être saines, surtout si elles donnent lieu à des synergies nouvelles qui contribuent ultimement à améliorer l'apprentissage des élèves et à de meilleures prises de décision aux différents niveaux des systèmes éducatifs (Laveault, 2009b; Laveault, Dionne, Lafontaine, Tessaro et Allal, 2014). Ces tensions peuvent aussi être contre-productives. De fait, les changements apportés par les évaluations externes et l'évaluation formative ne sont pas toujours à la hauteur des espérances. Il y a cependant lieu d'être optimiste. Comme annoncé par Fullan (2001, 2009), nous en sommes aux balbutiements d'une « professionnalisation » des réformes éducatives. Les systèmes d'éducation n'ont jamais fait l'objet d'autant d'études, de recherches et, bien entendu, d'évaluations. Il est à souhaiter qu'en assurant un meilleur suivi des systèmes éducatifs, le besoin de réformes importantes sera remplacé par des aménagements graduels et continus, tels que la révision périodique des programmes d'études (Legendre, 2002).

## L'évaluation des apprentissages: des concepts en évolution

Pour l'ensemble des raisons précédentes, la mobilisation des connaissances dans le domaine de l'évaluation scolaire est de plus en plus cruciale. Les changements intervenus récemment dans les systèmes d'éducation ont non seulement suivi l'évolution des concepts en évaluation, mais dans certains cas les ont précédés et même provoqués. Ces changements ont fait apparaître des besoins nouveaux en matière de formation initiale et continue des enseignants en évaluation afin de mieux refléter et actualiser dans la pratique les objectifs des politiques d'évaluation et, aussi, de tenir compte des changements dans le rôle des enseignants. Pour que les attentes élevées à l'égard des systèmes éducatifs se matérialisent, force est de reconnaître que non seulement les politiques d'évaluation, leurs fondements scientifiques et les outils d'évaluation doivent évoluer, mais aussi que les enseignants doivent être informés et formés en plus d'être partie prenante de ces changements.

Parmi les concepts en évaluation appelés à évoluer, figurent l'évaluation formative et la validité de conséquence (Messick, 1988). Ces deux concepts sont par ailleurs liés indirectement: à quoi peut servir l'évaluation si celle-ci ne se traduit pas par des effets positifs ou si l'ensemble des conséquences positives ne l'emporte pas sur les inconvénients? Comme le disait déjà Cardinet, «l'évaluation scolaire est formative par nature et sommative par nécessité».

### Le formatif et le sommatif: des objectifs qui se rejoignent

Historiquement, la responsabilité de soutenir l'apprentissage a été confiée à l'évaluation formative, l'évaluation sommative étant limitée au bilan des apprentissages et à la certification des apprentissages. Tourneur (1985, p.7) mettait déjà en garde contre cette fausse simplification à outrance:

L'évaluation sommative se trouve réduite au jugement de valeur sur la personne de l'élève ou à la notation des étudiants en termes de pourcentage ou de classement, quand elle n'est pas jugée responsable des abus liés à l'exercice du «pouvoir magistral».

Malgré le fait que les buts assignés à l'évaluation formative et à l'évaluation sommative soient différents, rien n'empêche a priori l'information recueillie initialement pour effectuer le bilan des apprentissages d'être utilisée dans un but formatif. De la même manière, l'information recueillie dans le but de soutenir l'apprentissage peut aussi renseigner sur la progression de l'élève et contribuer à son bilan, particulièrement dans les zones grises de prise de décision (Lafortune et Allal, 2008). Un groupe d'experts internationaux, principalement anglo-saxons, s'est entendu sur la définition suivante de l'expression évaluation-soutien d'apprentissage (EsA) («*assessment for learning*»), fréquemment utilisée en remplacement d'évaluation formative:

L'évaluation-soutien d'apprentissage fait partie des pratiques quotidiennes des élèves et des enseignants qui, individuellement et en interaction, réfléchissent sur et réagissent à l'information provenant d'échanges,

démonstrations et observations afin de favoriser les apprentissages en cours (Allal et Laveault, 2009).

Cette définition globale, tout comme les distinctions apportées par Stiggins (2005), cherche à établir un meilleur équilibre entre le rôle de l'enseignant et celui de l'élève. De plus, dans cette définition, l'éventail des situations d'évaluation pouvant servir à soutenir l'apprentissage est considérablement élargi pour y inclure autant des processus formels, comme des tests, qu'informels, comme les entretiens avec les élèves et les observations anecdotiques.

Ce qui importe selon Good (2011), ce n'est pas tant le qualificatif de l'évaluation que l'usage qui sera éventuellement fait de l'information recueillie. Par exemple, un bilan des apprentissages réalisé dans le cadre de l'évaluation sommative peut être réinvesti pour alimenter un entretien de l'enseignant avec l'élève et ses parents, identifier des cibles d'apprentissage prioritaires et orienter les choix scolaires de l'élève.

Peu importe le contexte dans lequel l'information sur l'apprentissage de l'élève est recueillie, un principe fondamental devrait être celui d'en optimiser l'utilisation pour l'apprentissage de l'élève. Plusieurs auteurs (De Ketele, 2006; Harlen, 2005; Looney, 2011; Taras, 2005) sont d'avis que le formatif et le sommatif ne doivent pas être utilisés en vases clos, l'un informant l'autre. De plus, les deux sont complémentaires: le sommatif porte principalement sur des informations à « gros grains », soit des apprentissages globaux se produisant sur des cycles de longue durée, alors que le formatif porte sur des apprentissages à « grains fins » dont les cycles sont de courte durée, jusqu'à l'interaction maître-élève. Peu importe la nature, fine ou grossière, de l'information recueillie sur l'apprentissage des élèves, l'important est de s'assurer que celle-ci contribue à améliorer les décisions qui sont prises pour favoriser l'apprentissage des élèves. Nous pourrions résumer ce principe par le slogan : formatif et sommatif, même combat!

### **La validité de conséquence : plus de bien que de mal**

L'optimisation de l'usage de l'information d'évaluation répond à des impératifs d'efficacité et d'économie. Déjà, beaucoup de temps est utilisé pour tester les élèves et recueillir de l'information sur leurs acquis. Mais qu'en est-il du temps pris pour traiter ces renseignements, les analyser et prendre les décisions nécessaires pour soutenir leur progression? On ne peut indéfiniment accumuler des renseignements sur l'élève ou sur les systèmes d'éducation sans se soucier de l'impact réel de ces évaluations sur l'apprentissage.

À cet égard, l'évaluation n'en serait pas à ses premières errances. De mauvais usages de l'information recueillie peuvent avoir des conséquences néfastes, telles que le rétrécissement du curriculum, une évaluation incomplète du rendement des enseignants et des écoles qui ne tiendrait compte que des résultats des élèves et l'introduction de palmarès de rendement des établissements scolaires qui ne tiendrait pas compte du contexte propre à chacun. Même l'évaluation à intention formative – dont le rôle attendu consiste pourtant à soutenir l'élève dans ses apprentissages – a pu connaître des dérives et donner lieu à des pratiques d'évaluation qui ne vont pas

dans le sens de tels objectifs. Enfin, ajoutons que ces errances vont bien au-delà des simples aspects techniques qui se limiteraient à l'utilisation des résultats, au développement de tests et à des analyses statistiques des résultats selon les « règles de l'art ». Elles interpellent au plus haut point nos valeurs éducatives et nécessitent une réflexion en profondeur sur l'éthique de l'évaluation scolaire (Hadji, 1997, 2012).

Parallèlement à la diversification des modalités d'évaluation et au décroisement entre le formatif et le sommatif, les exigences en matière de validité de conséquence ont progressivement été élevées. Si un usage plus diversifié de l'information recueillie sur l'apprentissage des élèves peut donner lieu à une meilleure utilisation du temps accordé à l'évaluation, cela ne signifie cependant pas que toutes les modalités de collecte d'information, ce qui est décrit en anglais par « *assessment* » (Laveault, 2010; Laveault, 2013), se valent les unes les autres. Une planification à rebours s'avérera nécessaire pour tailler la situation d'évaluation sur mesure, surtout lorsque les renseignements nécessaires ne sont pas disponibles, ou encore pour aider à prendre une décision bien précise. Bref, si une variété de renseignements recueillis sur l'élève, tant à « gros grains » qu'à « petits grains », peuvent contribuer à soutenir l'élève dans ses apprentissages, tout renseignement n'est pas formatif, et il est parfois nécessaire de concevoir des tâches qui permettront d'aller chercher une information utile à des fins bien précises.

Somme toute, la question n'est pas de savoir s'il faut recueillir plus ou moins de renseignements sur les élèves et sur le fonctionnement des systèmes éducatifs. La question est plutôt de savoir comment en optimiser l'usage au service de l'apprentissage de l'élève afin de réduire les effets négatifs de l'évaluation et d'en maximiser les effets positifs. Ultimement, l'évaluation scolaire a sa raison d'être lorsque l'utilisation des renseignements sur l'apprentissage de l'élève se traduit par un accroissement des probabilités, tant chez l'enseignant que chez l'élève, de prendre les bonnes décisions et de faire les bons choix afin de réguler au mieux la progression de ce dernier.

### « Teach to the test »

La mise en place des récentes réformes éducatives a été accompagnée dans plusieurs pays industrialisés par l'introduction d'examens uniformes censés assurer un suivi et orienter les prises de décision à tous les niveaux du système éducatif. Un tel changement n'a pas eu que des effets positifs. L'une de ses conséquences négatives a été de provoquer un « rétrécissement du curriculum », les enseignants consacrant plus de temps aux matières testées, telles que la lecture, l'écriture, les mathématiques, et délaissant les autres matières. Poussée à l'extrême, cette pratique néfaste conduit les enseignants à entraîner les élèves à passer des tests similaires aux tests officiels et à ne faire porter les apprentissages que sur des bribes de connaissances qui font partie régulièrement de ces examens.

Une telle façon de faire, connue en anglais sous l'expression « *teach to the test* », relève plus de mauvaises pratiques d'enseignement que de mauvaises évaluations. S'opposer à l'utilisation d'examens uniformes et standardisés sur la base de telles façons de faire ne repose pas sur des motifs valables. Au contraire, ce serait priver les enseignants, les écoles et les systèmes scolaires dans leur ensemble de renseignements

précieux sur la qualité de l'alignement des apprentissages des élèves sur les priorités des programmes d'études dans les matières de base.

Cela étant dit, le meilleur moyen de mettre fin à ce genre de dérives consiste à améliorer non seulement l'enseignement, mais aussi les outils d'évaluation et la formation des enseignants. L'utilisation à outrance de questions à choix multiples peut contribuer à réduire la crédibilité des épreuves standardisées, notamment pour tout ce qui concerne l'évaluation des apprentissages de haut niveau, et à en réduire l'impact sur l'apprentissage des élèves. Même s'il ne faut pas sous-estimer la capacité de questionnaires à choix multiples à tester de tels apprentissages avec validité, une évaluation fine de compétences complexes nécessite beaucoup de temps et de supervision et une diversité de moyens d'évaluation, ressources rarement disponibles pour les évaluations à grande échelle (Beckers, Hirtt et Laveault, 2014). C'est pourquoi des modalités « hybrides » d'évaluation, combinant ce que les examens uniformes font de mieux – recueillir une masse de renseignements sur tous les élèves en peu de temps et à faible coût – avec ce que les enseignants font de mieux – observer de façon précise les réalisations des élèves dans des situations d'apprentissage authentiques et complexes –, constituent sans doute le meilleur compromis.

Dans les faits, plusieurs organisations scolaires optent déjà pour ce genre de modalités d'évaluation à grande échelle, dont le dosage entre partie « externe » de l'évaluation et partie « interne » sous la responsabilité de l'enseignant peut varier. Il en existe plusieurs témoignages dans ce numéro spécial d'*Éducation et francophonie* et dans plusieurs autres références récentes (OECD, 2011, 2013; Laveault, 2009b; Laveault, Dionne, Lafontaine, Tessaro et Allal, 2014). Peu importe, cependant, la modalité choisie, la clef de voûte de l'implantation réussie d'une politique d'évaluation passe dans tous les cas par une compétence accrue des enseignants à évaluer, que ce soit pour utiliser les résultats des évaluations externes à bon escient ou encore pour participer à la collecte de données probantes sur les aspects les plus complexes de l'apprentissage des élèves.

## Les composantes d'une politique d'évaluation efficace

L'une des conséquences indirectes des études sur les évaluations internationales du progrès des élèves, telles que PISA, TIMMS, PIRLS, a été de mettre à l'avant-plan le rôle déterminant des enseignants dans la réussite scolaire des élèves. Dans la plupart des études sur la question, l'effet enseignant, quoique difficile à isoler d'autres effets de système (établissement, classe, etc.), rendrait compte d'une part non négligeable de la variance dans les résultats d'apprentissage des élèves : près de 30 % selon Hattie (2003) ou encore de 10 à 20 % selon Bressoux (1994). Selon l'OECD (2005), l'amélioration de la performance des enseignants est probablement la politique éducative la plus susceptible de se traduire en gains d'apprentissage pour les élèves. Darling-Hammond (2008) a démontré que les pays qui se classent le mieux aux enquêtes internationales sur les progrès des élèves se distinguent par le niveau d'exigence et la qualité de la formation du corps professionnel enseignant.

Cependant, la capacité des enseignants à exploiter l'information recueillie sur les apprentissages des élèves semble constituer un véritable défi. L'étude de Heritage, Kim, Vendlinski et Herman (2009) montre que, si les enseignants peuvent être du même avis quant à l'identification des faiblesses d'un élève, ils le sont moins souvent lorsqu'il s'agit de décider quelle devrait être l'étape suivante pour assurer sa progression. D'après Herman, Osmundson et Silver (2010), les enseignants ont besoin de développer les compétences qui leur permettront non seulement de définir les besoins d'apprentissage des élèves, mais aussi d'y répondre adéquatement. Ces deux études soulèvent des questions importantes quant à ce qui peut être mis en place pour améliorer la capacité des enseignants à intervenir adéquatement.

Dans les deux études précédentes, les enseignants devaient se prononcer sur la base de résultats obtenus au moyen d'outils d'évaluation qu'ils n'avaient pas eux-mêmes conçus. Les résultats de ces études demandent à être vérifiés dans des contextes où les enseignants jouent un rôle important dans la conception des instruments qui serviront à évaluer les élèves. Selon Looney (2011), les enseignants sont bien conscients de l'importance pour eux d'être plus systématiques dans leur approche de l'évaluation en salle de classe, étant donné que les interactions les plus efficaces avec les élèves requièrent une planification soignée. En conséquence, l'un des meilleurs moyens pour accroître l'usage formatif de l'information recueillie sur l'apprentissage des élèves consiste à accroître l'implication des enseignants dans la conception et la planification de l'évaluation (Laveault, 2014). Selon Webb (2009), l'amélioration des compétences professionnelles des enseignants en évaluation n'est pas un problème de ressources. Elle consiste plutôt à développer chez l'enseignant le regard d'un concepteur (*designers' eye*) pour choisir, adapter et développer les tâches nécessaires pour vérifier la compréhension des élèves.

Dans plusieurs États américains, de bons résultats d'évaluation peuvent être associés à des primes salariales, avec les dérives que de telles primes peuvent entraîner. Selon Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson et Shepard (2010), la prétention selon laquelle les enseignants vont améliorer l'apprentissage des élèves s'ils sont récompensés financièrement pour les gains des élèves n'est simplement pas justifiée par la recherche. Les conséquences indésirables de telles politiques sont nombreuses : appauvrissement du curriculum, désintérêt des enseignants pour les élèves les plus en besoin, diminution de la collaboration entre les enseignants et démoralisation.

Parmi les pays industrialisés, les réussites de la Finlande aux examens internationaux ont beaucoup attiré l'attention. Selon Darling-Hammond et McCloskey (2008), ce pays attribue sa réussite à des investissements massifs en formation des maîtres, la plupart d'entre eux possédant une maîtrise en éducation et dans leur matière de prédilection. En matière d'évaluation, la formation des enseignants finlandais met l'accent sur l'évaluation formative au service de l'apprentissage. Dans un tel contexte, il n'est pas étonnant de constater que les décideurs politiques accordent beaucoup plus d'autonomie aux enseignants quant à ce qu'ils doivent enseigner qu'on ne le fait ailleurs et que le programme d'études finlandais est un document beaucoup plus «léger» que celui rencontré dans d'autres pays.

L'État australien du Queensland est souvent mentionné pour illustrer les avantages d'une implication accrue des enseignants dans les systèmes d'évaluation nationaux (Atkin, 2007; Black et William, 1998; Darling-Hammon et McCloskey, 2008). Ce système a été développé et raffiné durant une période de trente ans et a nécessité à ses débuts des investissements importants dans la formation des enseignants. Il ne tient pas pour acquis que le jugement professionnel des enseignants repose uniquement sur la communication descendante (*top-down*) de standards nationaux. Dans ce système, la notion de «modération sociale» joue un rôle capital (Maxwell, 2007, cité dans Wyatt-Smith, Klenowski et Gunn, 2010). La modération tient pour acquis que l'évaluation des enseignants possède une validité réelle, mais que la fidélité ne peut en être assurée qu'en permettant aux enseignants de se faire une représentation commune des standards et des indicateurs nécessaires à la reconnaissance des résultats qui démontrent l'atteinte de ces standards (Maxwell, 2001, p. 6). De telles pratiques font en sorte que les standards visés par les politiques prennent un sens qui se développe et se raffine avec le temps à l'intérieur d'une communauté professionnelle d'usagers (Wyatt-Smith, Klenowski et Gunn, 2010).

L'ensemble des études internationales montre bien les limites d'une approche qui considère l'enseignant comme un simple technicien de l'évaluation. Au contraire, lorsque les politiques d'évaluation misent sur le développement des compétences professionnelles des enseignants, celles-ci s'accompagnent de meilleurs résultats. Les compétences en évaluation ne sont pas qu'un outil puissant de régulation de la pratique des enseignants en salle de classe. Elles sont aussi pour l'enseignant une source importante de réflexion sur sa pratique et de perfectionnement continu. Bref, pour qu'une politique d'évaluation fasse l'objet d'une mise en œuvre réussie, il faut prévoir améliorer non seulement les évaluations, mais aussi les évaluateurs. C'est à ce titre, également, que les enseignants pourront agir comme modèles et développer chez leurs élèves les habiletés réflexives et métacognitives qui les rendront plus aptes à s'autoévaluer et à apprendre par eux-mêmes. L'indicateur ultime de la réussite de la performance d'un système éducatif ne réside pas uniquement dans les résultats d'apprentissage eux-mêmes, que l'on peut gonfler momentanément de manière artificielle, mais aussi et surtout dans la capacité des enseignants et de leurs élèves à apprendre par eux-mêmes.

Les avancées dans le domaine de la recherche en évaluation scolaire nous ont permis de prendre conscience plus que jamais des impacts, tant positifs que négatifs, que peuvent avoir certaines politiques sur l'apprentissage des élèves. Mieux informés des conséquences, de ce qui se produit après la conception et la mise en œuvre de nouvelles politiques d'évaluation, nous pouvons maintenant mieux anticiper la validité de conséquence de plusieurs de nos choix futurs. À moins de pratiquer un aveuglement volontaire, à la prochaine vague d'innovations dans le domaine de l'évaluation on ne pourra certainement pas plaider l'ignorance, du moins pas entièrement. Il existe un corpus de savoirs à mobiliser dans ce genre d'entreprise et l'élaboration de nouvelles politiques ne pourra faire l'économie de ces connaissances accumulées grâce à de nombreuses recherches en évaluation des apprentissages et en évaluation de programmes.

## Et maintenant?

Quel est l'état du corpus de connaissances en matière de politiques d'évaluation dans les pays francophones et en particulier dans les provinces et territoires du Canada? À la suite de la publication en 2011 du numéro sur la réussite scolaire et pour répondre à de nombreuses demandes de son lectorat, la revue *Éducation et francophonie* a voulu rendre compte de l'état des connaissances sur les politiques d'évaluation en milieu scolaire en faisant paraître un numéro spécial sur le sujet. Elle entend ainsi faire la lumière sur les politiques actuelles d'évaluation afin de voir comment elles influencent les systèmes d'éducation et contribuent à améliorer les apprentissages des élèves. Les auteurs du numéro ont donc pour objectif non seulement de présenter les politiques d'évaluation, mais aussi d'engager une réflexion sur leur valeur éducative. En considérant la genèse, l'historique et le contexte particulier des politiques, ils ont cherché à bien expliquer les raisons d'être et à répondre aux questions suivantes:

- Qui a voulu ces politiques et à quelles fins?
- Les résultats obtenus ont-ils transformé les systèmes d'éducation?
- Les évaluations ont-elles contribué à améliorer les niveaux d'apprentissage?

Ce numéro spécial se présente donc comme une réflexion critique sur l'évaluation afin d'alimenter le débat qui a cours sur le sujet dans le domaine de l'éducation. Son objectif est de traiter des différents rôles, formatif et sommatif, que peut jouer l'évaluation à tous les niveaux des systèmes d'éducation, de la salle de classe aux plus hautes instances locales et nationales, en passant par les établissements scolaires. Une attention particulière a été accordée à l'évaluation des apprentissages dans le cadre de la scolarité obligatoire, c'est-à-dire au primaire et au secondaire. Pour ce qui concerne la composante canadienne, la problématique de l'évaluation scolaire en milieu minoritaire et la façon dont celle-ci est prise en compte dans les provinces ont fait l'objet d'une attention particulière. Enfin, ce numéro cherche à familiariser les lecteurs de la revue *Éducation et francophonie* avec les concepts importants de l'évaluation en prenant en considération le fait que ceux-ci ont évolué depuis les dernières réformes en éducation et que, dès lors, une mise à jour s'imposait.

Parce que les politiques d'évaluation relèvent d'instances gouvernementales, les articles de ce numéro spécial se distinguent d'abord par leur origine géographique. Selon le cas, ils décrivent la situation actuelle dans les provinces de l'Est du Canada (Bourque, Lang et Labelle), au Québec (Laurier), en Ontario (Laveault et Bourgeois) et enfin dans les provinces de l'Ouest et les territoires (Rocque). Pour faire écho à la situation prévalant dans la francophonie internationale, un article décrivant la situation vécue en Suisse fait également partie de ce numéro (Mottier-Lopez). Enfin, avec la publication toute récente des résultats PISA, nous avons voulu comparer les résultats des pays francophones. Les auteurs Baye, Demeuse et Friant vous expliqueront en quoi « ce n'est pas si simple! ».

Et la question est effectivement complexe: autant d'instances gouvernementales, autant de politiques d'évaluation, mais aussi autant de leçons à tirer et d'occasions

de parfaire nos connaissances et d'enrichir notre expérience. Toutes ces politiques tirent plus ou moins leur origine des dernières réformes en éducation. Dans les provinces de l'Ouest du Canada et dans les provinces maritimes, le management axé sur les résultats de performance (*New Public Management*) a grandement influencé les politiques d'évaluation. Du côté de l'Atlantique, les chercheurs ont cherché des moyens pour que «les évaluations standardisées puissent être utilisées d'une manière optimale sans qu'elles deviennent un obstacle aux finalités plus humanistes et émancipatrices proposées par *Les enfants au premier plan*». Pour ces auteurs, une approche cohérente du management axé sur les résultats est possible à condition «de ne pas confondre l'évaluation du système éducatif et l'évaluation des apprentissages».

Au Canada central, l'Ontario et le Québec semblent avoir connu des expériences fort différentes quant aux modalités d'implantation de leurs politiques. En ce qui concerne le Québec, Laurier conclut que la politique québécoise, fortement influencée par un modèle pédagogique axé sur le socioconstructivisme, «s'est déployée avec l'appui de documents complémentaires et de modifications aux règlements qui en précisent et parfois en réduisent la portée». En Ontario, Laveault et Bourgeois décrivent une implantation qui s'est faite dans la continuité, notamment par la mise en place de nouvelles institutions, dont un ordre professionnel des enseignants, mais surtout l'Office pour la qualité et la responsabilité en éducation (OQRE), chargé de faire le suivi des apprentissages des élèves et de vérifier l'alignement de la formation des élèves avec les attentes des programmes d'études. Ils en concluent que «c'est la refonte en profondeur des programmes-cadres en Ontario qui a le plus influencé les pratiques d'évaluation des enseignants et a soutenu une mise en œuvre progressive de la politique d'évaluation actuelle».

Une comparaison rapide entre les provinces canadiennes permet d'observer une différence importante dans l'implantation des politiques d'évaluation entre l'Est et l'Ouest du Canada et les deux provinces du centre. Dans le premier cas, la mise en place des politiques semble motivée et dirigée par des considérations d'ordre administratif, alors que dans le deuxième cas ce sont des considérations pédagogiques et curriculaires qui priment. Dans un cas comme dans l'autre, un équilibre optimal est sans doute souhaitable, car les deux aspects sont indissociables dans la pratique, même si les discours des théoriciens et des praticiens peuvent mettre l'accent sur l'un ou l'autre aspect.

L'article portant sur la Suisse apporte un éclairage différent. Devant l'objectif d'harmonisation de la scolarité actuelle obligatoire poursuivi par la Confédération helvétique, Mottier-Lopez se pose la question suivante: «Comment les fonctions formative et certificative de l'évaluation des apprentissages des élèves ont-elles été traitées dans les réformes en Suisse romande?» La méthode employée pour répondre à cette question est ingénieuse. En examinant les offres de formation aux enseignants genevois entre 1976 et 2013, Mottier-Lopez parvient à décrire une sorte de «cours des valeurs» que prennent celles-ci dans le perfectionnement continu des enseignants. Elle observe «des déséquilibres successifs entre les fonctions formative et certificative de l'évaluation, ainsi que des éléments d'incompréhension et de débats sociaux qui en ont résulté».

Les difficultés de mise en œuvre des politiques d'évaluation ont ceci en commun qu'elles font toutes ressortir l'importance de bien former les enseignants en matière d'évaluation. De toute évidence, nous n'y parvenons pas très bien, puisque, plusieurs décennies plus tard, nous devons encore définir en quoi consistent évaluation formative et sommative et qu'«il apparaît qu'un enjeu crucial est de penser un modèle d'une évaluation pédagogique intégrative et systémique parvenant à trouver un équilibre entre ses différentes fonctions qui, toutes, devraient tendre vers une même finalité de justice et de professionnalité» (Mottier-Lopez).

Un autre défi est particulièrement saillant dans les pays francophones, et en particulier là où l'enseignement est donné dans la langue de la minorité. La mondialité des échanges et la mobilité accrue des personnes font en sorte que les systèmes d'éducation doivent accueillir et intégrer de plus en plus de nouveaux arrivants. Baye, Demeuse et Friant démontrent que «certains systèmes semblent mieux que d'autres permettre aux élèves ne parlant pas la langue d'enseignement à la maison de la maîtriser en réduisant les écarts de performances entre groupes d'élèves». À cet égard, il existe de grandes différences entre les pays quant au risque qu'un élève immigré fasse partie du quartile des élèves les plus faibles que les natifs. Ce genre de données intéresse particulièrement les éducateurs et les éducatrices qui enseignent le français en milieu minoritaire, tant aux personnes natives qu'aux nouveaux arrivants.

Naturellement, il aurait été intéressant d'inclure plus de pays francophones dans ce numéro spécial, mais des contraintes d'espace et de temps nous en ont empêchés. Cependant, la coopération internationale en matière d'évaluation entre francophones est particulièrement active, et il est important de mentionner la collaboration qui existe entre l'Association pour le développement de la mesure et des méthodologies de l'évaluation au Canada (ADMÉE-Canada) et son association sœur, l'ADMÉE-Europe. Les deux ont en partage une revue internationale de langue française, *Mesure et évaluation en éducation*, qui constitue une référence incontournable pour tout ce qui concerne la recherche en matière d'évaluation en français. Cette revue a d'ailleurs publié un numéro spécial intitulé «L'évaluation : des politiques aux pratiques» (Laveault et Paquay, 2009). Ajoutons que le symposium de l'ADMÉE-Europe tenu au Luxembourg en 2012 a donné lieu à un chapitre de livre effectuant une mise à jour sur la situation qui prévaut au Canada et dans les pays francophones d'Europe (Laveault, Dionne, Lafontaine, Tessaro et Allal, 2014). Enfin, mentionnons que *l'International Encyclopedia of Education* a ajouté une nouvelle entrée sur la question de l'évaluation des politiques dans les pays francophones. L'article, rédigé en anglais, s'intitule «Classroom evaluation in a policy context: French sources» (Laveault, 2009).

Somme toute, nos connaissances en matière de politiques d'évaluation connaissent un essor considérable depuis une décennie. Ce numéro spécial d'*Éducation et francophonie* en est un autre témoignage. Il ne reste qu'à espérer qu'à l'avenir les leçons que nous tirons de l'expérience actuelle inspireront les prochaines générations d'éducateurs et qu'ils contribueront non seulement à améliorer l'évaluation des élèves, mais aussi à favoriser la réussite du plus grand nombre.

## Références bibliographiques

- ALLAL, L. et LAVEAULT, D. (2009). Évaluation-soutien d'apprentissage. Prise de position formulée par la Troisième Conférence internationale sur l'évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation*, 32(2), 99-107.
- ATKIN, J. M. (2007). Swimming upstream. Relying on teachers' summative assessments. *Measurement*, 5(1), 54-67.
- BAKER, E.L., BARTON, P. E., DARLING-HAMMOND, L., HAERTEL, E., LADD, H. F., LINN, R. L., RAVITCH, D., ROTHSTEIN, R., SHAVELSON, R. J. et SHEPARD, L.A. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers*. Economic Policy Institute: Briefing Paper no. 278. Récupéré le 20 septembre 2014 de <http://www.epi.org/publication/bp278/>
- BECKERS, J., HIRTT, N. et LAVEAULT, D. (2014). Les compétences à l'école : controverses et défis actuels. Dans C. Dierendonck, E. Loarer et B. Rey (dir.), *L'évaluation des compétences en milieu scolaire et en milieu professionnel* (p. 7-70). Bruxelles, De Boeck.
- BLACK, P. et WILLIAM, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-75.
- BRESSOUX, P. (1994). Les recherches sur les effets-écoles et les effets-maîtres. *Revue française de pédagogie*, 108, 91-137.
- DARLING-HAMMOND, L. et MCCLOSKEY, L. (2008). Assessment for learning around the world: What would it mean to be internationally competitive? *Phi Delta Kappan*, 90(4), 263-272.
- DE KETELE, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, 29(1), 99-118.
- FULLAN, M. (2001). *The New Meaning of Educational Change* (3<sup>e</sup> éd.). New York: Teachers College Press.
- FULLAN, M. (2009). Large-scale reform comes of age. *Journal of Educational Change*, 10, 101-113.
- Good, R. (2011). Formative « use » of assessment information: It's a process, so let's say what we mean. *Practical Assessment Research & Evaluation*, 16(3), 1-6.
- HADJI, C. (1997). Pour une éthique de l'agir évaluatif. *Mesure et évaluation en éducation*, 20(2), 7-26.
- HADJI, C. (2012). *Faut-il avoir peur de l'évaluation?* Bruxelles, De Boeck.
- HARLEN, W. (2005). Teachers' summative practices and assessment for learning – Tensions and synergies. *The Curriculum Journal*, 16(2), 207-223.
- HATTIE, J. (2003). *Teachers Make a Difference. What Is the Research Evidence?* Australian Council for Educational Research, octobre 2003. Récupéré de [http://research.acer.edu.au/research\\_conference\\_2003/4/](http://research.acer.edu.au/research_conference_2003/4/)

- HERITAGE, M., KIM, J., VENDLINSKI, T. et HERMAN, J. (2009). From evidence to action: A seamless process in formative assessment. *Educational Measurements. Issues and Practice*, 28(3), 24-31.
- HERMAN, J. L., OSMUNDSON, E. et SILVER, E. (2010). *Capturing Quality in Formative Assessment Practice. Measurement Challenges*. CRESST Report 770. Los Angeles: CRESST/University of California. Récupéré de <http://files.eric.ed.gov/fulltext/ED512648.pdf>
- LAFORTUNE, L. et ALLAL, L. (dir.) (2008). *Jugement professionnel en évaluation: pratiques enseignantes au Québec et à Genève*. Québec: Presses de l'Université de Québec.
- LAVEAULT, D. (2009a). L'amélioration de l'efficacité du système éducatif: sur quels indicateurs s'appuyer? Dans X. Dumay et V. Dupriez (dir.), *L'efficacité dans l'enseignement: promesses et zones d'ombre* (p. 177-194). Bruxelles: De Boeck.
- LAVEAULT, D. (2009b). L'évaluation en classe: des politiques aux pratiques. *Mesure et évaluation en éducation*, 32(3), 1-22.
- LAVEAULT, D. (2010). Classroom assessment in a policy context (French sources). Dans P. Peterson, E. Baker et B. McGaw (dir.), *International Encyclopedia of Education*
- LAVEAULT, D. (2013). Plus de convergences que de divergences: rencontre des conceptions francophones et anglo-saxonnes de l'évaluation des apprentissages. *Recherches et applications – Le français dans le monde*, 53, 31-41.
- LAVEAULT, D. (2014). The power of learning-centered task design: An exercise in the application of the variation principle. Dans C. Wyatt-Smith, V. Klenowski et P. Colbert (dir.), *Designing Assessment for Quality Learning* (p. 109-121), Berlin: Springer.
- LAVEAULT, D., DIONNE, E., LAFONTAINE, D., TESSARO, W. et ALLAL, L. (2014). L'évaluation des compétences: régulations entre politiques et pratiques. Dans C. DIERENDONCK, E. LOARER et B. REY (dir.), *L'évaluation des compétences en milieu scolaire et en milieu professionnel* (p. 85-96). Bruxelles: De Boeck.
- LEGENDRE, R. (2002). *Stop aux réformes scolaires*. Montréal: Guérin.
- LOONEY, J. W. (2011). *Integrating Formative and Summative Assessment*. OECD Education Working Papers, No. 58. Paris: OECD Publishing.
- MAXWELL, G. 2001. *Moderation of assessments in vocational education and training*. Brisbane, Queensland: Department of Employment and Training.
- MESSICK, S. (1988). The once and the future issues of validity: Assessing the meaning and consequences of measurement. Dans H. Wainer et H. I. Braun, *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum.

- OCDE (2009). *Examen par l'OCDE des cadres d'évaluation en vue d'améliorer les résultats scolaires*. Paris: OCDE, Direction de l'éducation. Récupéré le 20 septembre 2014 de <http://www.oecd.org/dataoecd/17/4/44568061.pdf>
- OCDE (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. Résumé en français de OECD Reviews of Evaluation and Assessment in Education. Paris: OECD Publishing. Récupéré de [http://www.oecd.org/edu/school/Synergies\\_for\\_Better\\_Learning\\_Résumé.pdf](http://www.oecd.org/edu/school/Synergies_for_Better_Learning_Résumé.pdf)
- OECD (2005). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD. Récupéré le 20 septembre 2014 de [http://www.oecd.org/document/52/0,3746,en\\_2649\\_39263231\\_34991988\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/52/0,3746,en_2649_39263231_34991988_1_1_1_1,00.html)
- OECD (2011). *Evaluation and Assessment Frameworks for Improving School Outcomes. Common Policy Challenges*. Paris: OECD. Récupéré le 20 septembre 2014 de <http://www.oecd.org/education/school/46927511.pdf>
- REEVES, D. B. (2006). Of hubs, bridges, and networks. *Educational Leadership*, 63(8), 32-37.
- SCALLON, G. (2000). *L'évaluation formative*, Saint-Laurent, Québec: Éditions du renouveau pédagogique.
- STIGGINS, R. J. (2005). *Assessment For Learning Defined*. An international forum on promoting sound assessment practice in every classroom. Portland, OR, Assessment Training Institute. Récupéré le 20 septembre 2014 de <http://ati.pearson.com/downloads/afldefined.pdf>
- TARAS, M. (2005). Assessment – Summative and formative – Some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466-478.
- TOURNEUR, Y. (1985). La certification des compétences. *Mesure et évaluation en éducation*, 7(4), 5-20.
- WEBB, D. C. (2009). Designing professional development for assessment. *Educational Designer*, 1(2). Récupéré le 20 septembre 2014 de [www.educationaldesigner.org/ed/volume1/issue2/article6](http://www.educationaldesigner.org/ed/volume1/issue2/article6)
- WILIAM, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14.
- WILIAM, D. et M. THOMPSON (2008). Integrating assessment with learning: What will it take to make it work? Dans C. A. Dwyer (dir.), *The Future of Assessment. Shaping Teaching and Learning*. New York: Erlbaum.
- WYATT-SMITH, C. M., KLENOWSKI, V. et GUNN, S. (2010). The centrality of teachers' judgment practice in assessment. A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.