

Mesure et évaluation en éducation



Le savoir-évaluer comme politique éducative Vers une évaluation plus profonde

François Victor Tochon

Volume 34, Number 3, 2011

URI: <https://id.erudit.org/iderudit/1024798ar>

DOI: <https://doi.org/10.7202/1024798ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Tochon, F. V. (2011). Le savoir-évaluer comme politique éducative : vers une évaluation plus profonde. *Mesure et évaluation en éducation*, 34(3), 137–160. <https://doi.org/10.7202/1024798ar>

Article abstract

Assessment as it is lived in American schools has changed its role. The humane dimension of teacher know-how, in assessment, has been replaced by quantitative performance assessments that play a political function, as it is the case in a number of countries. This essay discusses it from a critical perspective considering research on the validity of standardized performance tests. Teachers' field assessments tend to integrate the human factor as a major consideration, to the contrary of standardized measures that re-shape teaching to match the demand of the law makers. The current tendency in the field of school assessment places Ins tru mental Reason to its extreme, as if machines and bureaucratic criteria could replace the human agents. Because it enforces assimilation, it an carry a wide prejudice to linguistic minor ities. Assess ment is always relative enterprise, as it depends upon a concep tual interpretation and the context of its application. A more humane and deeper approach of evaluation is proposed.

Le savoir-évaluer comme politique éducative : vers une évaluation plus profonde

François Victor Tochon

University of Wisconsin–Madison

MOTS CLÉS : Raison instrumentale, tests de performance, politique éducative, évaluation profonde, sagesse

L'évaluation vécue dans les écoles américaines a changé de rôle. De l'évaluation à dimension humaine, nous sommes passés, dans de nombreux pays, à une évaluation de performance standardisée et quantitative dont la fonction est politique, que cet essai analyse de façon critique au vu des recherches sur la validité des tests standardisés. Dans leur évaluation de terrain, les enseignants tendent à tenir compte des facteurs humains, au contraire des mesures standardisées qui contribuent à refaçonner l'enseignement selon les prescriptions des législateurs. La tendance actuelle en matière d'évaluation scolaire porte la raison instrumentale à son extrême, comme si les machines et les critères bureaucratiques pouvaient remplacer les agents humains. Elle peut porter un large préjudice aux minorités linguistiques. L'évaluation est toujours relative car elle dépend d'une interprétation conceptuelle et du contexte d'application. Une approche plus humaine est proposée.

KEY WORDS: Instrumental Reason, performance testing, educational policy, linguistic genocide, racial segregation, deep evaluation, wisdom

Assessment as it is lived in American schools has changed its role. The humane dimension of teacher know-how, in assessment, has been replaced by quantitative performance assessments that play a political function, as it is the case in a number of countries. This essay discusses it from a critical perspective considering research on the validity of standardized performance tests. Teachers' field assessments tend to integrate the human factor as a major consideration, to the contrary of standardized measures that re-shape teaching to match the demand of the law makers. The current tendency in the field of school assessment places Instrumental Reason to its extreme, as if machines and bureaucratic criteria could replace the human agents. Because it enforces assimilation, it can carry a wide prejudice to linguistic minorities. Assessment is always relative enterprise, as it depends upon a conceptual interpretation and the context of its application. A more humane and deeper approach of evaluation is proposed.

PALAVRAS-CHAVE: Razão Instrumental, testes de performance, política educativa, avaliação profunda, sabedoria

A avaliação vivida nas escolas americanas mudou de papel. De avaliação à dimensão humana, passámos, em muitos países, a uma avaliação das performances estandardizadas e quantitativas cuja função é política, o que este artigo analisa de modo crítico tendo em conta as investigações sobre a validade dos testes estandardizados. Na sua avaliação do terreno, os professores tendem a ter em conta os fatores humanos, ao contrário das medidas estandardizadas que contribuem para reconfigurar o ensino segundo o prescrições dos legisladores. A tendência atual em matéria de avaliação escolar leva a razão instrumental ao extremo, como se as máquinas e os critérios burocráticos pudessem substituir os agentes humanos, pode levar um prejuízo grande para minorias linguísticas. A avaliação é sempre relativa porque depende de uma interpretação concetual e do contexto de aplicação. Uma abordagem mais humana é proposta.

Note de l'auteur – L'auteur tient à remercier Marie-Françoise Legendre et le comité éditorial pour leurs rétroactions détaillées sur deux versions antérieures de l'article. Toute correspondance peut être adressée comme suit : François Victor Tochon, University of Wisconsin–Madison, Teacher Education Building, 225 North Mills Street, Madison, Wisconsin 53706-1795 USA, téléphone (domicile) : (608) 437-7669, télécopieur (bureau) : (608) 263-9992, ou par courriel à l'adresse suivante : [ftochon@education.wisc.edu].

Une vingtaine d'années de recherche sur la pensée professionnelle des enseignants¹ avait indiqué l'importance de respecter les nuances de leurs décisions en matière d'évaluation et la dimension humaine de leur savoir-évaluer (Tochon, 2000). Malgré cette ample documentation, l'évaluation des contenus scolaires a été prise d'assaut dans certains pays anglosaxons et dans les pays qui ont suivi leur modèle, par des politiques agressives visant à remplacer les enseignants dans leur rôle d'appréciation des apprentissages. L'évaluation impose des formes et des contenus d'enseignement. Sa technicisation sans sagesse pratique ni théorique se révèle dangereuse. Dans plusieurs États, des tests quantitatifs sont imposés au gré d'une argumentation centrée sur la validité de construit de l'évaluation standardisée qui cache en fait d'autres arguments, économiques. Le savoir-évaluer des enseignants passe alors à l'arrière-plan.

Dans cet essai critique, j'examine en quoi cette façon de voir est mise en question dans les milieux de recherche. Je donne des exemples des effets pervers de la généralisation d'une évaluation standardisée quantitative qui, malgré une certaine utilité dans certains contextes au vu de buts spécifiques, peut porter atteinte aux libertés individuelles et au principe d'égalité des chances pour tous, en particulier pour les minorités linguistiques. J'analyse ensuite en quoi le savoir-évaluer enseignant est spécifique et en quoi sa dimension humaine s'avère indispensable pour l'évaluation dans un cadre scolaire. J'aborde aussi les logiques sous-jacentes à l'abandon du savoir-évaluer enseignant comme référence première en indiquant combien cette question est devenue politique. J'explore enfin certaines caractéristiques de ce savoir pratique, les modèles qui peuvent lui correspondre, et quelques tentatives de revaloriser ce savoir-évaluer dans un cadre de formation continue.

Prudence pratique et sagesse théorique dans le savoir-évaluer

Les instruments d'évaluation utilisés dans un cadre scolaire ne doivent pas être séparés de la connaissance pratique et de la conceptualisation théorique dont ils sont issus, faute de quoi ils sont «déracinés» (Tochon, 2011a). Cette démonstration avait été faite par Aristote (350 AC/2004) : la technique (*techné*) n'a de sens que si elle est guidée par la connaissance et la prudence pratiques (*phronesis*) dans un but de sagesse théorique (*sophia*). Kant (1784) avait

réinvesti cette démonstration ; il déplorait l'absence de raison théorique derrière les buts techniques de la logique instrumentale, une argumentation reprise par Habermas (2003) : la raison instrumentale n'a de sens que si elle est liée à la raison pratique et à la sagesse théorique. Ensemble, elles permettent la critique sociale nécessaire au maintien de la démocratie. Cette démonstration, corroborée par la théorie critique des systèmes (Fuenmayor, 2006), accorde une priorité à la raison humaine et contextualisée dans les procédures imposées par une logique instrumentale. En évaluation scolaire, par exemple, elle devrait conduire à privilégier le rapport de l'enseignant à l'élève dans le processus évaluatif, plutôt que d'accorder un pouvoir excessif à des instruments appliqués indépendamment du contexte comme les tests standardisés quantitatifs.

Bain (1987), dans une recherche sur le temps scolaire, avait remarqué qu'au niveau de l'école moyenne (11-15 ans), 300 heures étaient consacrées chaque année à l'évaluation, manifestement perdues pour l'apprentissage. Aujourd'hui, ces heures se sont encore multipliées en raison de l'imposition, par les milieux politiques, d'une forme de reddition de comptes influencée par les modèles des sciences économiques. Les orientations évaluatives imposées aux milieux scolaires par les politiques influencées par le *New Public Management*, aux États-Unis notamment, donnent à l'évaluation standardisée un rôle central de responsabilisation institutionnelle fondée sur quatre principes :

- 1) la production éducative est mesurable ;
- 2) elle doit être mesurée grâce à des instruments scientifiquement valides ;
- 3) les acteurs qui bénéficient d'autonomie sont redevables de leurs actions ;
- 4) les écoles doivent être régulées par leurs résultats.

Dès lors, les logiques d'évaluation et de construction de tests sont déconnectées de la réalité scolaire et de la réalité des personnes humaines qui sont la cible de l'évaluation. L'évaluation standardisée, qui s'appuie sur un modèle de management industriel, acquiert un statut politique de régulation macro-systémique qui n'est guère compatible avec les microprocessus humains propres à l'enseignement et à l'apprentissage. Or, « aucun consensus empirique net ne se dessine autour d'un bénéfice de ces réformes » (Mons, 2009, p. 6). De plus, on peut remarquer l'impact de la standardisation des performances sur les élèves de milieux ethniques, immigrants ou à faible revenus, ou souffrant de handicaps.

Le *New Public Management* imposé aux États-Unis, notamment dans l'analyse du fonctionnement des institutions, répond à des besoins politiques et économiques qui peuvent être en contradiction flagrante avec la dimension de rapport humain nécessaire en éducation. L'association entre praxis, instrumentation et sagesse est rompue dès lors que ce ne sont plus les enseignants qui évaluent. Les formes d'évaluation introduites dans les classes sous le prétexte de ne pas laisser d'enfants pour compte (*No Child Left Behind*) donnent une image déplorable de leurs limites quand ces évaluations sont associées à des mesures qui créent une panique chez les enfants, le personnel scolaire et les parents. Les tests scolaires imposés à des fins administratives ont pour conséquence de modifier l'enseignement dans une direction dépourvue de sens pédagogique et prennent un temps considérable au détriment de l'apprentissage. En bref, on peut se demander si la philosophie de l'évaluation centrée sur les performances attendues ne devrait pas être revisitée pour s'assurer qu'elle n'engendre pas des abus, voire une des formes de ce qu'Edgar Morin (1982) a nommé le terrorisme d'État, quand la raison instrumentale s'impose.

Les tests de performance standardisés : position des enseignants

La recherche sur les pratiques évaluatives des enseignants a surtout été menée par des équipes de recherche éloignées du terrain. On a peu eu recours à la recherche action participative, aux modalités de l'autoethnographie ou à la recherche collaborative. La nature sociale du *testing* a été analysée en termes de relations de pouvoir. Brookhart (2004) par exemple étudie les tensions relatives à l'application des principes de l'évaluation chez des stagiaires, préoccupés d'être socialement justes et de soutenir l'apprentissage. Sarrazin, Tessier et Trouilloud (2006) étudient le climat positif créé par l'enseignant et la façon dont il stimule la motivation des élèves. Black & William (2006) établissent les modalités appropriées d'application de l'évaluation ; McMillan (2007) tente de circonscrire le concept d'équité en évaluation. Harris et Brown (2008) examinent l'*accountability* du point de vue de l'enseignant et de l'élève et ils sont parmi les rares à utiliser l'introspection pour étudier ce qu'en pensent les enseignants. Simon, Chitpin et Yahya (2010) remarquent combien les théories sous-jacentes au processus d'évaluation sont influencées par des implications théoriques et techniques, psychologiques et sociales, dont les effets ne peuvent être sondés si l'on ne consulte pas les personnes les plus

directement concernées que sont les enseignants. Dans la plupart de ces recherches, on ignore ou néglige les effets des nouvelles formes d'évaluation standardisées introduites sur la vie de la classe, la perception qu'en ont les principaux acteurs, élèves et enseignants.

Dans son article sur « la fin de l'évaluation par les tests » qui vise à revaloriser le savoir-évaluer des enseignants, Eva Baker, alors présidente de l'Association américaine de recherche en éducation, remarquait dans son allocution présidentielle de 2007 que l'accent placé sur les tests crée un déséquilibre dans la vie scolaire. Elle notait que, pour rééquilibrer la situation en se fondant sur des principes d'équité et de justice, il conviendrait de dépasser la logique institutionnelle de reddition de comptes parce que les humains ont besoin d'une zone de confort pour apprendre. Bien des enfants et des enseignants craignent et même détestent les tests car leur application ignore le plus souvent les principes issus des recherches sur l'apprentissage et la motivation. Il découle de cette absence de lien conceptuel entre politique imposée et besoin éducatif une réalité très inconfortable : les tests ne mesurent généralement pas ce que les écoles veulent enseigner – ce qui contrarie le postulat de reddition de comptes. Nous avons très peu d'indices qui puissent suggérer que l'évaluation scolaire standardisée quantitative atteint ses fins et fournit une guidance utile dans la prise de décisions. « Nous agissons comme si les tests étaient valides, alors que nous n'en avons que de faibles indices, fort limités » (Baker, 2007, p. 310). Les restrictions budgétaires et les horaires de travail interdisent le plus souvent une interprétation *qualitative* des performances des enfants, alors que l'évaluation qualitative faite par les enseignants eux-mêmes a plus de validité externe, écologique.

On remarque dans les sondages que les enseignants ne sont pas *a priori* opposés à l'évaluation dans son principe, si elle soutient et stimule l'apprentissage. Ainsi des expériences fructueuses sont menées en Angleterre et en Nouvelle-Zélande dans lesquelles le professionnalisme enseignant est reconnu (Annan, 2011 ; Timperley, 2011). Quand les réformes sont associées à un développement professionnel, que leurs buts sont clairs et que les enseignants participent à la réflexion nécessaire au changement, les réformes du système évaluatif peuvent empêcher l'apparition de fortes inégalités entre programmes locaux et favoriser la réflexion en équipe autour de l'analyse des évaluations (Behrens, 2006). En revanche, les enseignants dénoncent souvent les dispositifs perçus comme négatifs parce qu'ils ne tiennent pas compte des caractéristiques sociales des publics accueillis ou qu'ils entraînent un « rétrécissement

des curricula» (Jones, 2007). Ces dispositifs, lorsqu'ils sont appliqués sans nuances, ont conduit certains enseignants et certaines écoles à tricher systématiquement pour élever les résultats, ce qui a été relevé tant à Chicago qu'en Ontario ou en Suède (Bélair, 2005). On a constaté, dans plusieurs États ayant généralisé des évaluations standardisées, une démotivation du corps enseignant dont la perception du métier devenait négative (DeBard & Kubow, 2002), phénomène d'autant plus important que les écoles accueilleraient des enfants de milieux à bas profil salarial, dont les résultats paraissaient faibles, faute de prendre en compte les caractéristiques sociales des élèves (Jones, 2007). Cette démotivation s'expliquerait par un sentiment de déprofessionnalisation, les tests enlevant aux enseignants une part majeure de leur pouvoir d'améliorer la situation de leurs élèves (Maroy & Cattonar, 2002; Osborn, 2006); la standardisation était alors perçue comme une déqualification.

Recalé dans une position de semi-professionnel, l'enseignant ne serait plus jugé au regard de sa tâche qui lui confère la légitimité d'évaluer le progrès de ses élèves. C'est vraisemblablement la raison qui pousse beaucoup d'enseignants à dénoncer les mécanismes à forts enjeux qui ont un impact jugé, en dernière analyse, négatif sur les démarches pédagogiques.

Les effets pervers de tests de performance standardisés

Les tests standardisés, provinciaux et nationaux, ont des effets pervers : ils modifient par exemple les contenus effectivement enseignés et abaissent le niveau général et l'interprétation qui est faite des compétences standards (Koretz, 2002). Ils engendrent dans la profession une hyperconscience que Weinberger (2007, p. 54) a nommée « *accountabalism* », une forme de « cannibalisme évaluatif ». Les enseignants en viennent à enseigner en fonction des tests de performance standardisés qui ont ainsi une force normalisatrice (Igen-Igaenr, 2005). Cependant, cela peut avoir dans certains cas un effet formateur (Demailly, 2001), si l'évaluation se développe de manière participative avec une forte association des enseignants, sur la base d'objectifs démocratiques (par opposition à une évaluation autoritaire), et si les concepteurs du changement ont un lien direct avec le terrain, une grande capacité de conviction et un volontarisme. Ce n'est que rarement le cas.

Tant sur le plan des contenus que de leurs modalités d'intégration, les tests limitent l'espace de liberté et de compétence de l'enseignant, déjà soumis à de multiples contraintes d'horaires, de programmes, et de vérification à plusieurs

niveaux sans oublier la pression parentale. Certains enseignants expriment le sentiment que les contenus sont presque relégués à l'arrière-plan tant la pression de l'évaluation standardisée quantitative est forte : il paraît important que l'enfant entre dans le moule de la production prescrite. La créativité des enfants est de plus en plus bridée par des buts précis, définis en termes mesurables, qui assurent un étalonnage commun des compétences acquises. Ce modèle rassure les politiciens car il paraît logique de penser que si l'enfant échoue dans l'accomplissement des buts scolaires, des mesures doivent être mises en place pour évaluer les causes possibles, avec des voies de remédiation pour l'enfant, des ateliers de perfectionnement pour l'enseignant, et des contraintes économiques en cas de faible rendement pour l'école.

Mais la mécanisation de l'évaluation peut entraîner la perte de valeurs humaines. Élèves et enseignants manifestent de nombreux problèmes de santé liée à la situation de test à fort enjeux (Gregory & Clarke, 2003). Dans certains pays comme les États-Unis (encore que la situation puisse varier d'un État à l'autre), ces mesures peuvent être poussées à l'extrême. L'interprétation des résultats de l'évaluation quantitative est naturalisée comme un fait objectif qui a des conséquences sérieuses, au point d'entraîner un avertissement à la direction d'école en cas de baisse de rendement aux tests officiels puis, après une période de renforcement des mesures censées renflouer l'école en déficit, une fermeture de l'école si elle ne répond pas aux normes minimales de rendement. Tout cela, pour un enfant « qu'on ne veut pas laisser en arrière », mais dont le trajet quotidien en bus scolaire, si son école ferme, sera souvent doublé pour atteindre, dans un autre quartier ou village, une école plus « performante ». L'évaluation mécanisée, dénuée de sens pratique et de sagesse, peut engendrer, dans un contexte comme celui-ci, une diaspora des enfants les plus pauvres. Le rapport scientifique sur ces phénomènes est bridé par les impératifs d'objectivité qui interdisent de témoigner de l'indignation que peuvent susciter de telles mesures, préjudiciables à un grand nombre d'enfants, d'enseignants et d'écoles.

L'école, qui devrait être un milieu de vie, est en train d'être récupérée à des fins non éducatives. L'évaluation standardisée quantitative joue un rôle majeur dans cette opération qui, sous l'apparence d'une markétisation du savoir, pourrait en fait ressortir d'une déstabilisation délibérée du système public pour imposer sa privatisation (Ravitch, 2010; Tochon, 2011b). Ce constat est pénible pour les professionnels de l'éducation. Certes, l'analyse qui est faite ici force le trait pour faire ressortir les aspects problématiques,

mais ceux-ci forment le substrat des problèmes scolaires contemporains de plusieurs États. La réaction actuelle des milieux conservateurs américains, qui cherchent à privatiser la scolarisation tout en éradiquant le pouvoir syndical, prend pour prétexte la coexistence forcée, en milieu scolaire public, d'enfants de classes sociales différentes et le risque de « contamination » qui s'ensuit. On trouve ces arguments sous diverses formes, qui vont de l'*English Only movement* à la fermeture des écoles peu performantes. Le lien avec l'enseignant, qui avait une valeur pacifiante, s'atténue alors que son rôle est confiné à la gestion de contenus préprogrammés. La normalisation de pratiques nouvelles entraîne une modification dans l'appréciation des valeurs.

Des limites de la validité de construit en évaluation

L'argument principal qui justifie l'intégration des tests standardisés dans les écoles est leur crédibilité en regard de l'évaluation des enseignants, car ces tests sont censés avoir une meilleure validité interne et de construit. Alors que la validité interne soutient la relation de cause à effet d'un test au détriment d'autres variables possibles, la validité de construit détermine si l'on mesure bien ce que l'on est censé mesurer et assure ainsi le potentiel de généralisation du résultat. Toutefois, Cronbach et Meehl, en 1955 (p. 297) avaient spécifié que la généralisabilité des tests repose sur la validation d'un « principe sur lequel fonder des inférences ». Alderson et Banerjee (2001), dans leur recension des théories relatives à la construction des tests, démontrent dès lors que l'important n'est plus tant la validité interne que la validité écologique, de nature externe. La validité d'une évaluation repose sur une interprétation subjective des priorités à privilégier, dont dépend la généralisation. Au cours des dernières années, la communauté de recherche est devenue toujours plus consciente des conséquences de l'utilisation des tests et de leur interprétation, et parfois de leur mauvais usage. On se pose maintenant la question de savoir si les développeurs de tests ne devraient pas être tenus responsables de l'impact de leurs tests sur la société, ce qui constituerait une forme nouvelle de validité, « implicative » (*consequential* dans les termes d'Alderson et Banerjee, 2002).

Les chercheurs qui ont mis en lumière cette nouvelle manière de penser l'évaluation et l'utilisation des tests en particulier, comme Lissitz et Samuelson (2007) ou McNamara et Roever (2006), en sont arrivés à la conclusion que la validité de construit a de multiples facettes. Il n'existe pas de réponse simple à la question de savoir si un test mesure vraiment ce qu'il est censé mesurer. À l'instar des instruments créés par les enseignants pour leur usage en classe,

les tests standardisés doivent s'inscrire dans un contexte spécifique et ne devraient pas être utilisés en dehors de la conceptualisation et du raisonnement qui justifient leur usage localement. Ceci revient à dire que le fondement conceptuel du test – le sens perçu de son utilité – est plus important que la validité de construit. Les chercheurs, pour cette raison, doivent investiguer les modalités interprétatives de l'évaluation dans son contexte, sa pertinence dans une situation d'apprentissage particulière, par exemple, plutôt que de se fixer sur les scores. Il s'agit là d'un usage plus réflexif et conceptuel de l'évaluation. «Ce qui résulte de cette perspective unifiée est que la validation est maintenant vue comme la supervision continue et la constante mise à jour de données pertinentes, un processus qui n'est jamais complet» et, dans les faits, qui ne peut être assumé que par l'enseignant (Alderson & Banerjee, 2002, p. 79). L'évaluation est valide si elle est pratique et si ses inférences sont utiles en pratique. L'utilité de l'évaluation dans son contexte d'utilisation est ainsi le premier critère selon lequel un test doit être jugé. Les conséquences de l'utilisation du test, son authenticité ou adéquation au contexte, son potentiel d'interaction et de communication et sa dimension pratique deviennent les qualités majeures et les critères attendus d'un test, qui contribuent à sa fiabilité.

Une évaluation bien pensée a pour but de proposer une rétroaction constructive, une émancipation des apprenants dans la réalisation de buts qui leur sont propres, et sur lesquels une communication a lieu (Fetterman, 2001). Ce second constat valorise le savoir-évaluer enseignant. En effet, l'enseignant est la seule personne au sein du système à connaître l'histoire personnelle de l'enfant et à pouvoir établir un pont entre ses besoins et les exigences du système scolaire. Dans le processus d'évaluation, la dimension interpersonnelle devient essentielle (McNamara & Lumley, 1997). Le choix de la personne qui teste est crucial : quelles sont ses perceptions des personnes évaluées ? Si l'évaluation compare les apprenants de façon normative, avec qui la comparaison a-t-elle lieu, et quelle influence cela a-t-il sur l'interprétation de l'évaluation ? Le sentiment de compétence est coconstruit au cours d'une évaluation pragmatique qui évolue au fil de l'expérience (Tochon, 1991). De plus, l'évaluation ne peut plus être comprise comme unidimensionnelle. Quelle que soit l'approche adoptée, les forces et faiblesses de l'action évaluative doivent viser l'utilité pratique pour l'apprenant. Cette révolution dans la manière dont nous concevons l'évaluation a entraîné plusieurs publications majeures chez *Cambridge University Press*, amenant une ouverture vers plus de bon sens dans un domaine qui avait jusqu'alors perpétué des notions extrêmement limitées de ce que l'évaluation devait représenter dans la salle de classe.

Les tests de performance comme violence symbolique et sociale

Giroux (2008) critique la politique des institutions éducatives et indique en quoi les logiques économiques issues du néolibéralisme ont introduit dans les écoles et les universités, sous le couvert de l'évaluation standardisée, une forme de contrôle ayant des implications qui vont bien au-delà des cadres scolaire et académique. Dans son livre intitulé «Contre le terrorisme néolibéral», Giroux redéfinit le néolibéralisme au sein de l'histoire de l'autoritarisme dans l'héritage du fascisme. Il estime que le néolibéralisme doit être compris dans la mouvance d'un processus sociohistorique caractérisé par des formes d'organisation indépendantes de tout éthos démocratique et fondé sur la foi dans la consommation et les lois du marché. Cette doctrine réduirait au silence le débat public qui est l'essence d'une culture démocratique, déplacerait les ressources en direction d'intérêts privés tout en généralisant des politiques de non-tolérance dans un cadre scolaire. La gestion statistique des populations et l'individualisation du risque témoignent «de méthodes par lesquelles des pédagogies néolibérales hégémoniques organisent les investissements affectifs, les désirs, et les identités en un réseau de contrôle social» (p. 173) auquel les individus doivent consentir faute de quoi des mesures punitives sont prises. L'évaluation standardisée quantitative peut être ainsi perçue comme une violence symbolique. On qualifie de violence symbolique «tout pouvoir qui parvient à imposer des significations et à les imposer comme légitimes en dissimulant les rapports de force qui sont au fondement de sa force» (Bourdieu, 1972, p.18). Ainsi, lorsque l'évaluation de performances standardisée, comparative et quantitative, imposée par l'État, devient omniprésente, elle peut conduire les enseignants comme les élèves à s'ajuster au profil exigé par l'autorité. La crise économique actuelle engendre des mesures évaluatives sans précédent dont les implications morales mériteraient d'être questionnées, notamment par les enseignants, qui doivent être en mesure d'adopter une distance critique sur la forme de société que l'école peut contribuer à promouvoir.

Il est vrai qu'une large variété de courants idéologiques peuvent coexister dans un État et que la situation est toujours complexe et nuancée. Rogers (1969) proposait la liberté d'apprendre. Certaines de nos écoles entraveraient-elles, de façon subtile, cette liberté? L'évaluation bien sûr ne constitue qu'un aspect de cet ensemble idéologique. Le *testing* de performance imposé tant aux élèves qu'aux enseignants et aux districts scolaires semble devenir une

des formes de la violence sociale. Quand l'évaluation des performances conduit à restreindre les chances de réussite d'une classe sociale ou d'un groupe racial, elle s'inscrit en porte-à-faux des valeurs démocratiques socialement promues par les accords et conventions de l'UNESCO et de l'Organisation des Nations Unies.

Les groupes minoritaires sont culturellement désavantagés par rapport à la culture dominante qui fabrique des tests standardisés et est à l'origine du système dans le cadre duquel ils sont employés. Le stress engendré par les tests de performance atteint les populations minoritaires en particulier. La souffrance mentale qui leur est infligée entraîne le transfert des enfants de milieux minoritaires à la culture dominante (Skutnabb-Kangas, 2000). L'intentionnalité apparaît par exemple dans les documents apportés au procès intenté à l'État de Pennsylvanie par la *National Parents' Commission*, en grande partie mené par des personnalités de droite au cours de la précédente administration (Luksik, 1996, 2005). L'article 7, par. 2, sous-par. g indique que la « persécution » définit une privation sérieuse et intentionnelle de droits fondamentaux. Par exemple, le droit fondamental à une égalité des chances en éducation et le droit des citoyens de refuser les tests et leur administration. Le sous-par. h couvre toute forme de « persécution contre un groupe identifiable ou une collectivité sur une base raciale, nationale, ethnique, culturelle, religieuse, etc., ou autres raisons qui sont universellement reconnues comme étant interdites par la loi internationale ». Il y a trente ans, Gould (1981) avait déjà indiqué ces problèmes avec pertinence dans son livre « Mismeasure of Man ». Les tests de performance ont un impact racial aux États-Unis dont témoignent les statistiques officielles de ces quinze dernières années. L'Union des maires de comtés avait déposé plainte en 1999 arguant du fait que les politiques scolaires d'évaluation contribuent de façon systémique aux performances basses des élèves, aux problèmes disciplinaires et aux problèmes de dépression des enseignants dans les écoles de milieux à salaire minimum. On constate une réduction systématique des performances scolaires des enfants africains-américains année après année ; ces différences ne sont pas attribuables à une déficience cognitive mais à d'autres facteurs comme les expériences liées au groupe dans le cadre du système éducatif (Finch *et al.*, 2002 ; Darling-Hammond, 2007). On évoque un déficit linguistique et les caractéristiques différentes de la langue ébonique (Sarzyniak, 2009), autant de façons de masquer le préjugé racial (Perry & Delpit, 1998). Les cas rares de succès (mitigés) des politiques de *No Child Left Behind*, s'expliquent par les mesures de redoublement systématique des enfants, en Floride en 2003-2004 notamment (Haney, 2006).

En bref, la situation créée par les tests de performance standardisés, dans certains pays, est telle que les bases nécessaires à un procès collectif sur le plan national et à un recours aux organes de justice internationaux sont présentes. Dans ces pays, les tests de performance standardisés peuvent constituer des instruments de ségrégation engendrant une souffrance mentale que certains qualifient même de génocide (Alexander, 2010). Selon ces hypothèses, l'évaluation de performance imposée en fonction de normes standardisées, quantitatives et non contextualisées serait devenue un fléau. Nombre d'enseignants résistent à la culture de l'évaluation de performance (Behrens, 2006). En raison du pouvoir qu'a l'évaluation de modifier les approches spontanées de l'enfant vis-à-vis du savoir et, en d'autres termes, de faciliter ou d'empêcher certaines aspirations et d'aiguiller les trajectoires désirées, il s'avère important de comprendre en profondeur et de respecter la manière dont les professionnels du terrain scolaire perçoivent et conçoivent les pratiques évaluatives (Harris & Brown, 2008).

Vers une conception plus humaine et profonde du savoir-évaluer

Le *testing* de performances à forts enjeux, sur fond idéologique néolibéral, place les élèves, les enseignants et les écoles dans une compétition malsaine qui tend à pourrir le climat social des districts scolaires en créant des conflits sans fin. La publicité donnée à ces tests se révèle souvent préjudiciable et contraire aux buts de la réforme. Même si l'on s'accorde sur le fait que les modèles d'*accountability* peuvent varier, le *testing* systématique des performances a des effets pervers qui influencent négativement le processus d'apprentissage. Les tests ne peuvent se justifier hors contexte. Ils peuvent être nuisibles si les enseignants ne sont pas impliqués dans leur conception, leur gestion, et la gestion de leur utilisation à des fins constructives. On ne peut faire fi du savoir-évaluer des enseignants, qui se fonde sur une pratique réfléchie et une sagesse de l'action bien contextualisée. Quels que soient les dispositifs adoptés, ils doivent être articulés avec le développement professionnel, le soutien financier des écoles en difficulté, et des programmes de soutien entre établissements scolaires, sous forme de réseaux d'échanges professionnels par exemple. La manière dont les enseignants participent à ces évaluations doit être revue en profondeur : « Plus ceux-ci participent à la fois à la conception, l'administration et l'analyse des résultats, plus leur implication dans le processus s'en trouve développée et la culture d'évaluation assimilée » (Mons, 2009, p. 36).

Quand les perspectives sont mécanistes et superficielles et que la formation n'est pas vraiment réfléchie, réflexive et approfondie, des distorsions apparaissent ; chaque fois, on peut localiser les problèmes dans l'absence de prise en compte du savoir enseignant car les enseignants ne sont guère engagés dans les processus de changement et son évaluation (Stobart, 2011). Confrontés à des politiques parfois contraignantes, les praticiens de l'éducation doivent devenir plus actifs. Ils peuvent mener des recherches collaboratives avec les formateurs et les décideurs pour stimuler une amélioration de l'éducation en continu (Annan, 2011). L'évaluation permet alors de piloter le dispositif, de clarifier les rôles et d'envisager un développement durable en faveur d'enfants de milieux désavantagés. La formation continue des enseignants est un moyen sûr d'améliorer les résultats des élèves. La recherche action participative dans de tels réseaux permet d'analyser les difficultés des élèves et de localiser les secteurs de besoins (Timperley, 2011).

Les conceptions qu'ont les enseignants influent sur leur pratique et agissent sur les résultats des élèves (Muijs, 2006) ; toutefois, bien des enseignants savent pondérer leur rigueur évaluative avec un peu d'humanité dans des évaluations authentiques. Celles-ci impliquent d'observer, de prendre note et de documenter le travail de l'enfant afin de guider des décisions adaptées à chaque situation. Cette approche se fonde sur les événements éducatifs de la vie courante, permettant la mise sur pied de projets évolutifs dont l'évaluation est intégrée et continue. Elle se centre sur des problèmes de la vie réelle au sein d'une recherche ouverte impliquant la réflexion, des échanges sur l'activité et des apprentissages sociaux (Rule, 2006). L'évaluation authentique est une alternative à l'évaluation standardisée quantitative (Gardner, 1991). Elle répond à plusieurs fonctions du savoir-évaluer des enseignants car ceux-ci sont avant tout enseignants parce qu'ils veulent faire apprendre et non sanctionner (Tochon, 2010). Ces fonctions représentent le cœur du métier :

- 1) individualisation face à la standardisation ;
- 2) encouragement face au découragement engendré par les tests ;
- 3) dialogue face au monologue administratif ;
- 4) apprentissage personnel face à la mesure publique.

Face aux logiques instrumentales imposées dans certains pays, une solution est ainsi de revaloriser le savoir-évaluer enseignant. La documentation de recherche indique en effet des changements dans la représentation que se font les acteurs de leur profession, de leur action sociale et de leur pouvoir

d'action, dès lors largement réduit. Les recherches sur la motivation à l'accomplissement et à l'autodétermination font ressortir les conséquences positives d'un climat dans lequel les élèves sont les maîtres de leurs apprentissages et de leur évaluation, et l'impact négatif d'un climat de compétition fondé sur les rangs relatifs à des tests standardisés ; l'aménagement d'un contexte d'apprentissage qui soutient l'autonomie des élèves s'avère bénéfique (Sarrazin, Tessier, & Trouilloud, 2006). Pour atténuer les effets pervers de l'évaluation standardisée quantitative, les enseignants peuvent utiliser des stratégies visant à accroître l'autonomie des apprenants, intégrer des évaluations formatives, des technologies ou un matériel d'instruction que les enfants apprécient et avec lesquels ils se sentent à l'aise ; proposer un tutorat ; assurer une rotation des contenus ; établir des priorités ; utiliser des mesures multiples ; intégrer dans leur évaluation des contenus et des activités qui ont une valeur humaine. Ce sont là autant d'aspects du savoir-évaluer enseignant.

Les enseignants ont pour la plupart des valeurs socialement utiles et une éthique de vie (Noddings, 2003). Ils contribuent à une évaluation équitable qui s'appuie sur la connaissance des élèves dans leur contexte, car ils sont experts sur leur terrain professionnel (Tochon, 1993, 2004). Leur apport est essentiel pour concevoir une société plus humaine. L'évaluation émancipatrice (*empowerment evaluation* – Fetterman, 2001) semble être une des rares formes d'évaluation qui stimule l'apprentissage et reconnaît l'enseignant comme expert sur son terrain. « Il en résulte une hausse de l'autodétermination et une émancipation » (Fetterman, 2005, p. 213). Wandersman *et al.* (2005) soulignent les principes de cette évaluation : elle doit porter sur des améliorations à dimension humaine ; les acteurs locaux doivent assumer la responsabilité de son *design*, gérer l'évaluation et mettre en œuvre les mesures nécessaires ; la prise de décision est partagée, fondée sur une communication et une collaboration authentiques ; des buts d'équité sociale visent le bien plus large de la communauté en accroissant la capacité des personnes ; les instruments sont adaptés à l'environnement, à la culture et aux conditions locales.

Une telle évaluation implique activement les enseignants qui peuvent témoigner de leur savoir local et le rendre utile. Ce qui est notable, si l'on compare ces principes à ceux d'une standardisation quantitative du rendement scolaire, est l'implication intelligente des participants locaux, un engagement qui remplit des buts d'apprentissage organisationnel et professionnel. Cette approche suscite une discussion honnête sur les points forts et les faiblesses des programmes évalués. Il s'ensuit une planification concertée des remédia-

tions, l'identification de la documentation utile au progrès. L'évaluation devient partie intégrante de la gestion de programmes, non pas à distance mais sur place (Fetterman & Wandersman, 2004). Lorsqu'elle se centre sur l'approfondissement des contenus et relève d'une sagesse en action bien contextualisée, elle peut constituer ce que j'appellerai une «évaluation profonde»: une évaluation qui stimule des apprentissages en profondeur.

Des écoles communautaires vont dans cette direction, au nom d'un apprentissage de vie autonome. Dans ces modèles alternatifs, l'éducation se fonde sur des projets. Des écoles de sagesse pratique, locales, échangent et génèrent des savoirs, savoir-faire et savoir-être locaux, uniques et particuliers. Le savoir-évaluer vise alors des dimensions intégrées d'épanouissement, d'amour et de respect. L'apprentissage de la sagesse remplace la logique de production; on revient à la philosophie, qui constitue l'apprentissage premier (Fuenmayor, 2006). Grâce à ces savoirs locaux et autonomes, et à cette sagesse, chaque personne peut croître dans la liberté de connaître et d'agir selon une éthique de vie en société. Toda (2000) a ainsi proposé un ensemble de sociétés régionales plus petites et humaines, fondées sur le soin mutuel. Korten (2006) démontre combien ces communautés d'apprentissage élargies peuvent changer la société. Les apprenants voyagent dans ce réseau d'écoles établies dans plusieurs pays, pour comprendre d'autres pratiques et d'autres cultures. La liberté de créer d'autres manières de faire pourrait être un antidote aux abus administratifs des ministères gouvernementaux.

En guise de conclusion

Changer la nature de l'évaluation scolaire et valoriser le savoir-évaluer des enseignants

En bref, le début de cet article a mis en valeur deux forces contradictoires en évaluation éducative : celle qui soutient une pratique et une sagesse réflexives, et celle dont la visée est économique et politique, qui se centre sur la reddition des comptes. Le problème est que la seconde s'effectue en général au détriment de la première, et qu'elle dépossède l'enseignant d'une dimension cruciale de sa compétence professionnelle qui lui permet de réguler l'apprentissage de façon individualisée, en encourageant l'enfant par un dialogue qui personnalise l'apprentissage. Cette dimension humaine risque d'être rejetée par les mesures de performance standardisées imposées de façon autoritaire, car elles modifient la nature de l'enseignement, qui perd en grande part sa dimension éducative et est relégué au rang d'instruction technique.

Nous avons vu que, parmi les acteurs impliqués dans le système évaluatif mis en place dans ces réformes, cadres administratifs, enseignants, apprenants et parents sont conditionnés par l'environnement psychologique et social que créent les tests de performance et la comparaison quantitative des scores entre élèves, enseignants, écoles et districts scolaires. Bien que la plupart des partenaires du changement ne voient pas *a priori* grand mal à généraliser ce qui auparavant était une pratique locale bien acceptée, on s'aperçoit que l'extension du dispositif évaluatif hors du contrôle des praticiens de l'éducation engendre des biais systématiques dans la conception du travail de l'enseignant qui sont potentiellement dommageables pour l'apprentissage, pour le climat de la classe et de l'école, et pour la dimension proprement humaine de ce que l'éducation représentait auparavant. Ainsi, certains chercheurs (Baker, 2006 ; Baker, 2007 ; Stavenhagen, 1990) mettent en garde contre le risque que les tests de performance deviennent dans certains contextes des facteurs de sélection sociale, voire raciale, et de mise à l'écart de certaines minorités linguistiques et culturelles. Toute évaluation est socialement marquée, car la valorisation de certains critères dévalorise d'autres aspects d'une manière potentiellement discriminatrice. La culture sous-jacente à ce processus de valorisation et de dévalorisation est rarement discutée, comme si l'évaluation portait sur des faits alors qu'elle est un acte politique. On remet rarement en question ce qui est dévalorisé dans le processus proposé.

Une réflexion critique peut déboucher sur des solutions et une action, dont la portée est déjà, en fait, partagée par beaucoup. Dans plusieurs pays, des chercheurs s'accordent sur certaines constantes. Comme le mentionnent Morrisette, Loye et Legendre (2011, p. 2), « L'évaluation des apprentissages dépasse la simple administration d'instruments » ; leur utilisation doit être contextualisée. Le jugement professionnel des acteurs de terrain est essentiel à cet égard (Bélair, 2006 ; Morrisette, 2010). Les politiques scolaires doivent intégrer une pensée du complexe qui soit respectueuse de la personne humaine. Nous devons confronter les logiques qui font de nous des machines à produire. L'évaluation de performance semble être devenue la seule politique éducative dans certains pays. Cette situation, associée à de multiples autres contraintes nouvelles, revêt certaines des caractéristiques d'un excès de bureaucratie. La société occidentale n'est pas prête à adopter des systèmes « post-bureaucratiques » (Maroy, 2008) comme le recommandait Ivan Illich (1969/2000). Dans plusieurs pays, l'omniprésence des mesures techniques d'évaluation quantitative tend à déposséder les humains de leur pouvoir sur leur propre action. Leur liberté de penser, de croire et de s'exprimer est en jeu. Le danger est bien réel ; l'évaluation sociométrique n'est plus liée à un idéal de ce que l'humanité pourrait devenir, mais devient une opération de contrôle et de gestion des masses humaines répondant surtout à des impératifs économiques et sécuritaires. Le risque qui naît du savoir-évaluer économique – un savoir qui n'a rien d'éducatif – est toutefois compensé par la rééquilibration des économies planétaires et les agents, au sein de l'occident comme dans les autres parties du monde, qui souhaitent un changement pacifique et négocié plutôt qu'une révolution pour modifier ces pratiques abusives.

Les hypothèses fortes proposées ici mériteraient d'être davantage étayées et argumentées, ce que l'espace imparti ne permet pas de faire. Le danger d'un propos trop général sur l'évaluation est bien évidemment d'en confondre les formes, les niveaux, les visées, alors que ce qui est ciblé ici est le principe d'atteinte à l'autonomie des personnes dans un système de masse. Mais ces questions sont importantes parce qu'elles indiquent *a contrario* le rôle de l'enseignant en manifestant son inverse. L'analyse critique porte sur le principe d'une évaluation de performance standardisée, quantitative, imposée à des masses d'individus, sans leur laisser d'autre choix que l'obéissance au système. Personne n'y trouve à redire *a priori*, même les enseignants, jusqu'au moment où ils réalisent l'ampleur et les conséquences des modifications du système de valeurs. Car c'est là que se situe l'enjeu : l'évaluation de masse, telle qu'imposée, a modifié les valeurs. L'évaluation qui étymologiquement devrait porter

sur la valorisation de ce qui est unique chez une personne devient un outil de mesure et d'uniformisation dans l'industrialisation sociale du savoir. En conséquence, le savoir-évaluer propre à l'enseignant ne devrait-il pas aussi, dès lors, porter sur la forme de société qui est imposée ? Dans certains de mes écrits (Tochon 2004, 2011), j'expose pourquoi la conscientisation doit devenir un des rôles des enseignants et des formateurs, un point sur lequel Freire (1992) avait insisté.

Repenser l'évaluation en termes de valorisation pourrait nous amener à reconceptualiser son pouvoir créateur. La production devrait cibler la création, le projet de vie, le geste d'intelligence, l'expression unique, l'art de vivre, la recherche singulière individuelle ou partagée, plutôt que de tous nous lier à des normes standards. Ainsi, afin de résoudre en partie tout au moins certains des problèmes susmentionnés, nous pourrions passer d'une évaluation de produits qui limite l'apprenant à un seul type de résultat à une évaluation de processus menant à des résultats variés. Il s'agirait de jouer la différence plutôt que l'homogénéité et la normalisation, ce qui rejoint l'idée d'une évaluation comme soutien à l'apprentissage et au développement de la personne.

NOTES

1. Dans cet article, le masculin est utilisé de façon générique et sans intention de discrimination.
2. E793, 1948; 78 U.N.T.S. 277, entrée en application le 12 janvier 1951; voir [<http://www1.umn.edu/humanrts/instree/x1cppcg.htm>].
3. U.S. Department of Education Quota News: [http://www.adversity.net/fedagency_news.htm#1_fed_agency].

RÉFÉRENCES

- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching*, 34(4), 213-236.
- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35(1), 79-113.
- Alexander, M. (2010). *The new Jim Crow: Mass incarceration in the age of colorblindness*. New York: New Press.
- Annan, B. (2011). L'amélioration continue des établissements scolaires : l'ingénierie mise en place en Nouvelle-Zélande. *Revue française de pédagogie*, 174, 21-30.
- Aristotle (350 BC/2004). *The Nicomachean ethics* (trad. & notes de J. A. K. Thompson; éd. rév. H. Tredennick). Londres : Penguin.
- Bain, D. (1987, septembre). *La formation en évaluation formative, argumentation et plaidoyer pour une entrée par la didactique*. Genève : Centre de recherche psychopédagogique. Exposé présenté lors du congrès international de l'Association de mesure et évaluation en éducation (ADMEE), Fribourg, Suisse.
- Baker, D. (2006). The political economy of fascism: Myth or reality, or myth and reality? *New Political Economy*, 11(2), 227-250.
- Baker, E. L. (2007). 2007 presidential address: The end(s) of testing. *Educational Researcher*, 36(6), 309-317.
- Behrens, M. (2006). *Analyse de la littérature critique sur le développement, l'usage et l'implémentation de standards dans un système éducatif*. Neuchâtel : IRDP.
- Bélair, L. M. (2005). Les dérives de l'obligation de résultats ou l'art de surfer sans planche. Dans C. Lessard & P. Meirieu (éds), *L'obligation de résultats en éducation* (pp. 179-188). Bruxelles : De Boeck Université.
- Bélair, L. M. (2006). L'évaluation au quotidien : conjuguer processus et produit. *Mesure et évaluation en éducation*, 29(1), 19-30.
- Black, P., & William, D. (2006). Assessment for learning in the classroom. Dans John Gardner (éd.), *Assessment and learning* (pp. 9-25). Thousand Oaks, CA : Sage.
- Bourdieu, P. (1972). *Esquisse d'une théorie de la pratique*. Paris : Droz.

- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Darling-Hammond, L. (2007). Evaluating 'No Child Left Behind'. *The Nation*, 21 mai. Extrait de : [<http://www.thenation.com/node/22962>].
- DeBard, R., & Kubow, P. K. (2002). From compliance to commitment: The need for constituent discourse in implementing testing policy. *Educational Policy*, 16(3), 387-405.
- Demilly L. (2001), Enjeux de l'évaluation et régulation des systèmes scolaires. Dans L. Demilly (éd.), *Évaluer les politiques éducatives* (pp. 181-195). Bruxelles : Éditions De Boeck Université.
- Fetterman, D. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage.
- Fetterman, D. (2005). Conclusion: conceptualizing empowerment evaluation in terms of sequential time and space. Dans D. Fetterman & A. Wandersman (éds), *Empowerment evaluation: principles in practice* (pp. 209-214). New York, NY: Guilford.
- Fetterman, D. M., & Wandersman, A. (2004). *Empowerment evaluation principles in practice*. New York: Guilford.
- Finch, S. J., Farberman, H. A., Neus, J., Adams, R. E., & Price-Baker, D. (2002). Differential test performance in the American educational system: the impact of race and gender. *Journal of Sociology and Social Welfare*, Sept. 2002. Extrait de : [http://findarticles.com/p/articles/mi_m0CYZ/is_3_29/?tag=content;col1].
- Freire, P. (1992). *Pedagogy of the oppressed*. New York: Continuum.
- Fuenmayor, R. (2006, juillet). *El camino de la sistemología interpretativa (La voie de la systéologie interprétative)*. VI Congreso Científico de la Universidad de Oriente. Núcleo Anzoategui, 28 juillet 2006.
- Gardner, H. (1991). *The alternative to standardize testing*. New York: Basic Books.
- Giroux, H. A. (2008). *Against the Terror of Neoliberalism: Politics beyond the age of greed*. Boulder: Paradigm.
- Gould, S. J. (1981). *Mismeasure of Man*. New York: W.W. Norton.
- Gregory, K., & Clarke, M. (2003). High-stakes assessment in England and Singapore. *Theory into Practice*, 42(1), 66-74.
- Habermas, J. (2003). *Truth and justification* (trad. B. Fultner). Cambridge, MA: MIT Press.
- Haney, W. M. (2006). *Evidence on education under NCLB (and How Florida boosted NAEP scores and reduced the race gap)*. Boston College, Center for the Study of Testing, Evaluation and Education Policy. Extrait de : [<http://www.bc.edu/research/nbetpp/statements/nbr6.pdf>].
- Harris, L. R., & Brown, G. T. L. (2008). *New Zealand teachers' conceptions of the purpose of assessment: Phenomenographic analyses of teachers' thinking*. Paper presented at the Australian association for research in education conference. Brisbane, Australia.
- Holman, S. (2006, août). *The effects of dialogue journaling on second language acquisition*. (Mémoire de Maîtrise). University of Wisconsin-Madison, Madison, WI.
- Igen-Igaenr (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école ?* Paris : IGEN-IGAENR.

- Illich, I. (1969/2000). *Deschooling society*. New York: Marion Boyars.
- Jones, B. (2007). The unintended outcomes of High-Stakes Testing. *Journal of Applied School Psychology*, 23(2), 65-86.
- Kant, I. (1784). What is enlightenment? Dans Noms (éds), *Foundations of the metaphysics of morals* (pp. 80-92). New York: Bobbs-Merrill.
- Koretz, D. (2002, April). *Using multiple measures to address perverse incentives and score inflation*. Communication présentée lors du symposium Multiple perspectives on Multiple Measures, au congrès annuel du National Council on Measurement in Education. La Nouvelle-Orléans, Louisiane.
- Korten, D. C. (2006). *The great turning. From empire to earth community*. San Francisco: Berrett-Koehler, Kumarian.
- Lissitz, R. W., & Samuelson, K. (2007). Dialogue on validity. A suggested change in terminology and emphasis regarding validity and Education. *Educational Researcher*, 36(8), 437-448.
- Luksik, P., & Hoffercker, P. H. (1996). *Outcome based education: The state's assault on our children's values*. Philadelphia, PA: Vital Issues.
- Luksik, P. (2005). *National parents' commission report*. Johnstown, PA.
- Magnusson, J. L. (2001). *Mad scientists and technosavants: exploring the cultural icons of industrial capitalist science*. Article présenté lors du congrès «Teaching as if the World Mattered». Toronto, OISE/University of Toronto.
- Maroy, C. (2008). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? *Sociologie et Société*, 40(1), 31-55.
- Maroy, C., & Cattonar, B. (2002). Professionnalisation ou déprofessionnalisation des enseignants? Le cas de la Communauté française de Belgique. *Cahier de recherche en éducation et formation*, 18. Louvain-la-Neuve: Université de Louvain.
- McMillan, J. H. (2007). *Classroom assessment: Principles and practice for effective teaching* (4th ed.). Boston, MA: Pearson Education.
- McNamara, T. F., & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing*, 14(2), 140-156.
- McNamara, T. F., & Roever, C. (2006). *Language testing: the social dimension*. Malden, MA: Blackwell.
- Mons, J. (2009). *Les effets théoriques et réels de l'évaluation standardisée*. Compléments à l'étude «Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats EACEA». Université Pierre-Mendès-France, Grenoble 2, Eurydice.
- Morin, E. (1982). *Science avec conscience*. Paris: Fayard.
- Morrisette, J. (2010). *Manières de faire l'évaluation formative des apprentissages : analyse interactionniste du savoir-faire d'enseignantes du primaire*. Sarrebruck, CH: Les Éditions universitaires européennes.
- Morrisette, J., Loye, N., & Legendre, M.-F. (2011, janvier). *Autour du savoir-évaluer des enseignantes et des enseignants en salle de classe*. Montréal, QC: Université de Montréal.
- Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12(1), 53-74.

- Noddings, N. (2003). *Caring: A feminine approach to ethics and moral education* (2nd ed.). Berkeley: University of California Press.
- Osborn, M. (2006). Changing the context of teachers' work and professional development: a European perspective. *International Journal of Educational Research*, 45, 242-253.
- Perry, T., & Delpit, L. (1998). *The real Ebonics debate: power, language, and the education of African-American children*. Milwaukee, WI: Rethinking Schools.
- Ravitch, D. (2010). *The death and life of the great American school system*. New York: Basic Books.
- Rogers, C. R. (1969). *Freedom to learn*. Columbus, OH: Charles E. Merrill.
- Rule, A. C. (2006). The components of authentic learning. *Journal of Authentic Learning*, 3(1), 1-10.
- Sarrazin, Ph., Tessier, D., & Trouilloud, D. (2006). Climat motivationnel instauré par l'enseignant et implication des élèves en classe : l'état des recherches. *Revue française de pédagogie*, 157, 147-177.
- Sarzyniak, K. (2009). *Attitudes that affect the Black-White achievement gap: how do cultural identity and the use of Ebonics affect African American students' success in the classroom?* New York: Lambert Academic.
- Simon, M., Chitpin, S., & Yahya, R. (2010). Pre-service teachers' thinking about student assessment issues. *International Journal of Education*, 2(2), E5.
- Skutnabb-Kangas, T. (2000). *Linguistic genocide in education – or worldwide diversity and human rights*. Mahwah, NJ: Lawrence Erlbaum.
- Stavenhagen, R. (1990). *The ethnic question. Conflicts, development, and human rights*. Tokyo: United Nations University Press.
- Stobart, G. (2011). « L'évaluation pour les apprentissages » : d'une expérimentation locale à une politique nationale. *Revue française de pédagogie*, 174, 41-48.
- Timperley, H. (2011). Le développement professionnel des enseignants et ses effets positifs sur les apprentissages des élèves. *Revue française de pédagogie*, 174, 31-40.
- Tochon, F. V. (1991). Évaluation pragmatique et situations spontanées : l'indexation des activités en classe. *Mesure et évaluation en éducation*, 14(4), 42-47.
- Tochon, F. V. (1993). *L'enseignant expert/L'enseignante experte*. Paris : Nathan.
- Tochon, F. V. (2000). Recherche sur la pensée des enseignants: un paradigme à maturité. *Revue française de pédagogie*, 133, 1-23.
- Tochon, F. V. (2004). Le nouveau visage de l'enseignant expert. *Recherche et formation*, 47, 89-103.
- Tochon, F. V. (2010). Deep education. *Journal for Educators, Teachers and Trainers*, 1, 1-12.
- Tochon, F. V. (2011a). Deep education: Assigning a moral role to academic work. *Educação, Sociedade & Culturas*, 33, novembre.
- Tochon, F. V. (2011b). Propos libres sur la formation des enseignants : Une entrevue avec François Victor Tochon. Formations et pratiques d'enseignement en questions. *Revue des Hautes écoles pédagogiques et institutions de formation de Suisse romande et du Tessin*, 12.
- Toda, M. (2000). Emotion and social interaction: A theoretical overview. Dans G. Hatano, N. Okada & H. Tanabe (éds), *Affective minds* (pp. 1-17). Amsterdam: Elsevier Science & Technology.

- Wandersman, A., *et al.* (2005). The principles of empowerment evaluation. Dans D. Fetterman & A. Wandersman (éds), *Empowerment evaluation: principles in practice* (pp. 27–41). New York, NY: Guilford.
- Weinberger, D. (2007, February). The HBR list: Breakthrough ideas for 2007. *Harvard Business Review*, 85(2), 21-36, 44-48, 50, 52, 54.

Date de réception : 2 juin 2011

Date de réception de la version finale : 2 juin 2011

Date d'acceptation : 30 novembre 2011