

D'autres façons de concevoir et de faire l'évaluation des apprentissages

Jacques Laliberté

Membre du Groupe de recherche-action
PERFORMA
Université de Sherbrooke

Aux États-Unis, les limites inhérentes à l'utilisation, fort répandue, des tests standardisés ont amené des spécialistes et des éducateurs à chercher d'autres façons d'évaluer¹ les apprentissages des élèves. À cela s'ajoutent l'influence grandissante d'une conception cognitiviste et constructiviste de l'apprentissage de même que celle qui découle d'une orientation du curriculum dans le sens de l'éducation basée sur le développement de compétences (*Competence-based education*) ou, en termes plus larges, l'éducation basée sur les résultats recherchés (*Outcome-based education*). Tout cela a eu des effets profonds sur la manière de concevoir et de mettre en œuvre l'évaluation des apprentissages chez nos voisins du Sud.

En février 1991, dans un article intitulé « "Évaluation" dites-vous ? Non, "assessment"... » et publié dans *Pédagogie collégiale* (vol. 4, n° 3, p. 36-39), Paul Forcier analysait les caractéristiques essentielles d'un renouveau des pratiques d'évaluation

des apprentissages aux États-Unis. Cet article conserve toute sa pertinence et il est d'une actualité encore plus vive, puisque ce qui apparaissait alors comme une tendance naissante a, depuis ce temps, pris de l'ampleur.

Notons, au passage, que dans la documentation américaine subsiste un problème de terminologie puisque, dans plusieurs écrits, le terme *assessment* a une acception générique et désigne pratiquement toute forme d'évaluation, tandis que, dans d'autres textes, leurs auteurs l'emploient pour faire référence à ces « nouvelles » façons d'évaluer les apprentissages des élèves dont il vient d'être question.

Sous-jacentes à ces formes nouvelles se retrouvent les préoccupations suivantes : évaluer ce que les élèves sont capables de faire avec leur savoir, leurs habiletés et leurs capacités, leurs attitudes et dispositions (souci d'intégration et de transfert des

apprentissages) ; évaluer leur capacité de faire la démonstration de leur maîtrise d'une compétence, leur capacité de résoudre un problème difficile ou de réaliser une tâche complexe ; évaluer leurs capacités supérieures de pensée ; évaluer leur degré d'atteinte de standards précis et publics ; rapprocher l'évaluation du monde réel par la contextualisation des épreuves, des tâches ou des problèmes servant à l'évaluation.

Les trois notices descriptives présentées ici portent sur des articles et un ouvrage, récemment publiés aux États-Unis, qui s'inscrivent dans ce mouvement de recherche de nouvelles formes d'évaluation et qui pourraient sans aucun doute inspirer notre réflexion et notre action, à un moment où, dans le cadre du renouveau de l'enseignement collégial, on s'interroge sur des nouvelles façons de concevoir et de faire l'évaluation des apprentissages.

TAYLOR, Catherine, « Assessment for Measurement or Standards : The Peril and Promise of Large-Scale Assessment Reform » dans *American Educational Research Journal*, vol. 31, n° 2, été 1994, p. 231-262.

Sous un titre un peu rébarbatif et qui semble annoncer un propos d'allure très technique, Catherine Taylor, dans cet article très récent, nous convie à une réflexion en profondeur sur les finalités de l'évaluation des apprentissages et, en dernière analyse, sur la philosophie de l'éducation qui sous-tend les pratiques d'évaluation

auxquelles on a recours dans les établissements scolaires ou à l'échelle de tout un système d'enseignement. Vers la fin de son texte, elle formule clairement l'enjeu fondamental en ces termes : « Allons-nous continuer de créer des instruments qui sont conçus pour classer et comparer les élèves, ou voulons-nous des systèmes d'éva-

luation qui nous permettent de savoir si les élèves atteignent des cibles d'apprentissage complexes ? » (p. 254) Et un peu plus loin, elle ajoute : « Nous devons commencer à croire que *la plupart* des élèves sont tout à fait capables d'apprendre et de réussir, que les différences considérables [*dramatic differences*] que nous observons

dans les performances étudiantes sont la résultante de conditions qui n'ont aucun rapport avec la capacité d'apprendre des élèves. » (p. 255) Conditions et différences dont il faut bien tenir compte cependant...

Dans l'expression de cette conviction, on peut voir une autre façon de mettre de l'avant le postulat d'éducabilité proposé par Meirieu, notamment, comme moteur de l'activité professionnelle de l'enseignante et de l'enseignant et qui exige, comme une espèce de corollaire naturel, la conception et la mise en œuvre d'une pédagogie différenciée, si l'on veut vraiment tenir compte de l'hétérogénéité des groupes-classe.

Mais, ne nous y trompons pas, l'article de Taylor, du début à la fin, a bien pour objet l'évaluation des apprentissages et porte essentiellement sur des considérations d'ordre méthodologique d'une grande portée : les exigences relatives à la validité et à la fidélité dans la conception des instruments et l'interprétation des résultats ; le choix et l'utilisation des critères et des standards ; la nature et les connotations du jugement professionnel à exercer.

Taylor part du constat qu'aux États-Unis les éducateurs et les législateurs sont à la recherche de systèmes d'évaluation (*assessment*) qui vont requérir des élèves qu'ils s'engagent dans la résolution de problèmes et dans des tâches complexes exigeant le recours à des capacités supérieures de pensée plutôt que de faire simplement la démonstration d'un savoir cloisonné (*discrete knowledge*)² et de l'habileté à appliquer ce savoir (p. 232). Parmi les méthodes, moyens et modalités d'évaluation qui sont de plus en plus utilisés ou recherchés, on note l'évaluation *authentique*, l'évaluation de *performances* et les évaluations à l'aide du *portfolio*.

Selon l'auteure, il faut prendre conscience que, dans le cas des entreprises d'évaluation sur une large échelle, les éducateurs et les législateurs demandent que l'on conçoive des épreuves ou des instruments qui vont servir deux finalités incompatibles :

- a) déterminer si les élèves maîtrisent les standards ou les performances désirés ou du moins progressent en ce sens ;

- b) fournir des évaluations relatives (*relative measurements*) des élèves, des écoles, des districts scolaires ou des États par rapport à des échelles de rendement (p. 232). À ce sujet, Taylor craint que le fait d'appliquer au développement d'évaluations de la performance un modèle fondé sur la mesure, et ce sur une large échelle, en vienne ultimement à saper les efforts qu'on fait sur le plan national pour améliorer la qualité de l'éducation de tous les élèves (p. 233).

Dans son article, Taylor compare les caractéristiques essentielles, les visées profondes, les conséquences pratiques et pédagogiques de deux modèles d'évaluation des apprentissages : *le modèle fondé sur la mesure* (qui est le modèle dominant aux États-Unis depuis une soixantaine d'années) et *le modèle fondé sur des standards* (actuellement en forte émergence chez nos voisins du Sud).

Dans le premier modèle, on est à la recherche des différences observables chez les personnes. On postule qu'on peut situer l'individu, pour une caractéristique ou un trait donné, par rapport à une courbe de distribution « normale », cette fameuse courbe qui a donné lieu à l'établissement de procédures psychométriques servant à établir la fiabilité des tests, à assurer la rigueur et la validité de l'interprétation des résultats, etc. (p. 236-242). Dans le *modèle fondé sur la mesure* c'est la différenciation et le classement des individus qui priment sur la détermination d'attentes précises à l'égard des élèves. « L'excellence, écrit Taylor (p. 242), est déterminée par le fait que quelqu'un a une note supérieure à tous les autres qui ont passé le même examen », le même test ou la même épreuve, pourrait-on ajouter.

Quant au *modèle fondé sur des standards*, il repose sur quatre postulats :

- ◆ nous pouvons déterminer des standards publics et travailler à les atteindre ;
- ◆ la plupart des élèves peuvent assimiler et réaliser les standards ;
- ◆ des performances et des démonstrations très différentes de la part des élèves peuvent refléter les mêmes standards ;

- ◆ les éducateurs peuvent être formés à assimiler les standards et à devenir des juges fiables et constants de diverses performances étudiantes (p. 243).

On voit combien ces deux modèles sont profondément différents l'un de l'autre. Et l'auteure, tout en rappelant et en soulignant les limites du testing traditionnel, ne fait pas mystère des défis qu'ont à relever les théoriciens et les praticiens d'une évaluation basée sur la performance : assurer la fiabilité et la fidélité d'un pareil type d'évaluation ; déterminer les performances essentielles dans des disciplines données ; établir des standards et des critères relatifs à ces performances ; obtenir des exemples de performances reflétant ces standards et ces critères ; communiquer le tout au public ; etc. (p. 247-253). Ce qui entraîne plusieurs conséquences et exigences sur le plan pédagogique (p. 254-259).

On peut, me semble-t-il, tirer une double leçon des propos de Taylor :

- d'une part, le type d'évaluation qu'on préconise et qu'on utilise doit être en harmonie avec les visées éducatives qu'on poursuit ;
- d'autre part, une évaluation basée sur la performance, pour donner des fruits et être crédible, ne doit pas simplement avoir un caractère d'authenticité par rapport au monde réel et par rapport à ce qui est fondamental dans une discipline ; elle doit également être finalisée par des critères et des standards exigeants et intervenir dans un environnement éducatif où l'on a tout mis en œuvre pour aider les élèves à atteindre ces standards, à respecter ces critères, à travers des productions ou des démonstrations, variables d'un élève à l'autre, mais toutes adéquates.

Rien donc qui ressemble à un automatisme, ou à un nivellement par le bas...

Plusieurs commentateurs ont souligné que, par une utilisation judicieuse des critères et des standards, l'approche par compétences pourrait avoir pour effet de contribuer au développement, chez les élèves, de la capacité d'auto-évaluer, avec rigueur et précision et sans complaisance, leurs apprentissages et leurs comportements. Celles et ceux que cette thématique de l'auto-évaluation intéresse liront avec grand profit ce numéro que le périodique *New Directions for Teaching and Learning* consacre à la question. Non seulement y trouveront-ils des éléments de justification de cette pratique pédagogique ainsi que des possibles modalités d'application, mais également une évocation des retombées qu'elle peut avoir pour les élèves, les enseignantes et les enseignants. Tout le numéro est à lire. En plus de la notice préliminaire, il rassemble sept textes d'auteurs différents. Chaque contribution mériterait d'être commentée. Je me contenterai ici de donner du relief à quelques facettes seulement de ce recueil de textes.

Pour Edith Kusnic et Mary Lou Finley, « l'expression auto-évaluation étudiante (*student self-evaluation*) renvoie tout à la fois à une production écrite, qui peut prendre une diversité de formes, et au processus conduisant à de telles productions par les élèves. En tant que processus, l'exercice de s'auto-évaluer demande aux élèves de réfléchir sur ce qu'ils ont appris et d'écrire sur le sujet. L'auto-évaluation étudiante, comme processus, est un puissant outil d'apprentissage. En tant que productions, les auto-évaluations étudiantes décrivent et analysent l'apprentissage des élèves pour eux et pour nous » (p. 8).

Deux caractéristiques fondamentales ressortent avec force dans cette citation : l'importance de l'écriture dans l'auto-évaluation étudiante et le rapport que celle-ci entretient avec la qualité de l'apprentissage des élèves. À travers tout le numéro, on retrouve exprimée, sous diverses formes, cette double articulation.

Dans leur article (p. 5-14), Kusnic et Finley indiquent ou nous permettent d'inférer quelques-unes des retombées positives que peuvent avoir des exercices ou des tâches amenant les élèves à rédiger des auto-évaluations de leurs apprentissages. Cela peut :

- aider les élèves à utiliser leur savoir ;
- aider les élèves à développer des capacités de réflexion sur soi (*self-reflection*) et à établir des relations actives et signifiantes avec la matière qu'ils étudient ;
- amener les élèves à renforcer leurs capacités d'analyse, de synthèse et d'évaluation ; à trouver du sens à ce qu'ils ont appris et à explorer la relation que ce nouveau savoir entretient avec leurs connaissances et leurs idées antérieures ; à devenir plus conscients de leurs valeurs et des façons dont se forment leurs valeurs ; à approfondir ce qu'ils ont appris et à établir des liens entre eux et le contenu de leurs études ; à développer les capacités, la compétence et l'assurance requises pour un apprentissage efficace tout au long de leur vie ;
- apporter une nouvelle forme de rétroaction aux élèves sur leur apprentissage et fournir des données utiles pour une évaluation des résultats de l'éducation et de l'enseignement ;
- amener les élèves à se mettre au centre et aux commandes de leur expérience d'apprentissage (voir p. 5 à 9 tout particulièrement).

De son côté, Carl J. Waluconis (p. 15-33) décrit différents contextes pour des auto-évaluations étudiantes. On peut concevoir des exercices d'auto-évaluation :

- centrés sur une brève période ;
- portant sur un travail s'étalant sur plusieurs semaines ;
- portant sur tout un cours ;
- ayant pour objet plus d'un cours.

L'auteur parseme son article d'extraits de textes rédigés par des étudiantes et des étudiants.

Pour terminer cette trop rapide présentation, je voudrais attirer l'attention sur l'annexe (p. 101-117) dans laquelle on retrouve :

- des exemples de tâches, de conseils et de directives concernant des exercices d'auto-évaluation par les étudiantes et les étudiants ;
- des exemples d'essais d'auto-évaluations étudiantes d'une certaine ampleur ;
- une liste de ressources additionnelles : des ouvrages à caractère théorique touchant le développement des étudiants et la valeur de la réflexion sur soi ; des approches pratiques et des études se rapportant spécifiquement à l'auto-évaluation étudiante.

Il faut noter que la totalité des exemples précis qui sont fournis dans cette annexe sont de niveau postsecondaire, la plupart sinon tous référant vraisemblablement à des études du premier cycle universitaire.

Trois caractéristiques de la problématique actuelle de l'évaluation des apprentissages aux États-Unis :

- ◆ Le testing est remis en cause ;
- ◆ plusieurs chantiers sont présentement ouverts touchant la détermination des standards ;
- ◆ on recherche et on expérimente d'autres façons d'évaluer les apprentissages des élèves.

À quelqu'un qui me demanderait : « Dans la documentation américaine que tu as lue récemment, y a-t-il un livre que tu me recommanderais pour ce qu'il pourrait m'apporter touchant l'évaluation des apprentissages des élèves dans une approche par compétences ? », je répondrais sans hésiter : « L'ouvrage de Grant P. Wiggins, *Assessing Student Performance...* ». J'estime, en effet, que ce livre peut, de bien des manières, nous aider à approfondir notre compréhension de ce que des spécialistes américains appellent l'*assessment* lorsqu'ils veulent faire référence à de nouvelles tendances dans les façons de concevoir et de mettre en œuvre l'évaluation des apprentissages des élèves. De plus, *Assessing Student Performance...* peut être une source d'inspiration extrêmement riche pour la pratique professionnelle des enseignantes et des enseignants, et pour celle des conseillères et des conseillers pédagogiques qui travaillent avec eux.

Il n'est pas aisé de rendre justice ici au contenu de cet ouvrage. Tout de même, pour en donner un aperçu aussi concret que possible, j'ai choisi d'attirer l'attention sur un certain nombre de thèmes traités par Wiggins et dont chaque lecteur, selon ses centres d'intérêt, ses préoccupations et ses convictions, peut tirer profit.

- 1° On y trouve une critique du testing traditionnel largement répandu aux États-Unis. Les tests standardisés ne sont pas sans mérites mais ils comportent plusieurs limites : un accent abusif sur la simple connaissance de faits ; une simplification et une décontextualisation des tâches afin d'assurer plus de précision à la notation ; une conception fixiste de l'intelligence se traduisant par des pratiques évaluatives où le classement des élèves les uns par rapport aux autres apparaît plus important que la qualification d'une performance en regard de standards clairement identifiés ; etc. (voir les chapitres 1, 3, 4 et 5 principalement).
- 2° On y fait la promotion d'une conception large et exigeante de l'*assessment* défini comme « une analyse complète d'une performance, analyse personnelle, basée sur le jugement et comportant plusieurs facettes. Comme l'écrivait, il y a une trentaine d'années, Lee Cronbach, professeur à l'Université Stanford et doyen des psychométriciens américains : « l'*assessment* nécessite l'utilisation d'une variété de techniques, repose principalement sur des observations (de la performance), et exige l'intégration d'une information diversifiée dans un jugement synthèse [*summary judgment*]. » (p. 13)
- 3° Dans une perspective d'*assessment*, le professeur devient l'allié de l'élève davantage qu'un juge (p. 14). Wiggins soutient que « de judicieux *assessments* ne diffèrent pas des tests simplement parce qu'ils sont plus complexes. Les questions relatives aux droits et aux responsabilités sont cruciales : dans un *assessment* adéquat, nous faisons passer les droits des élèves en premier » (p. 22). De quelle nature sont ces droits ? Wiggins nous en donne une idée à l'aide de deux documents : le premier est un ensemble de principes adoptés par le ministère de l'Éducation de la Nouvelle-Zélande et qui visent à mettre l'*assessment* au service d'un meilleur apprentissage (p. 26-27) ; le second est une Déclaration des droits des élèves en matière d'*assessment* que Wiggins lui-même a rédigée (p. 28), mais dont il a le regret de dire qu'elle a été mal reçue par plusieurs enseignants auxquels il l'a présentée dans des ateliers...
- 4° Lorsqu'on veut évaluer le progrès intellectuel des élèves, on est confronté à huit dilemmes que Wiggins décrit sommairement (p. 37-45). Par exemple :
 - ◆ nous devons nous préoccuper de ce que les élèves savent mais nous devons aussi nous préoccuper d'évaluer si ce qu'ils savent a de la signification pour eux ;
- 5° Wiggins dégage neuf postulats à respecter, si l'on veut mettre en place un système d'*assessment* sérieux ; parmi ceux-ci :
 - ◆ nous devons établir un équilibre entre une évaluation de la maîtrise que les élèves ont des idées et des productions d'autrui et une évaluation de leur maîtrise de leurs propres idées et productions.
- 6° Soucieux que l'évaluation ait un effet positif sur la motivation des élèves à apprendre, Wiggins y va de quelques recommandations en ce sens, notamment les deux suivantes :
 - ◆ un système d'évaluation authentique doit reposer sur des critères et des standards connus, clairs, publics et non arbitraires ;
 - ◆ on évalue mieux le degré de compréhension des élèves en donnant suite aux questions qu'ils posent qu'en se limitant à une notation de leurs réponses ;
 - ◆ on devrait évaluer l'honnêteté intellectuelle des élèves et d'autres dispositions d'esprit ; au-delà des cas de tricherie, on devrait aussi prendre en considération la capacité des élèves d'admettre leur ignorance ou d'exprimer leur perplexité face à une question ou un problème.
- 7° évaluez le progrès et les réalisations des élèves ; pour cela que votre notation soit basée sur des modèles d'une performance exemplaire vers laquelle les élèves tendent à travers des cheminements différents (p. 171-172) ;
- 8° concevez un système d'évaluation en vertu duquel les proportions de ce qui compte dans la notation peuvent varier : au début, on peut ac-

corder une pondération plus forte à l'effort et au progrès, mais en faisant porter l'accent, par la suite et de plus en plus, sur le rendement et la réalisation (p. 172-173).

7° L'auteur accorde une grande importance à la rétroaction qu'on peut donner aux élèves. Il compare les caractéristiques respectives d'une rétroaction efficace et d'une rétroaction inefficace (p. 198-199). Il avait auparavant rappelé à quelles exigences doit satisfaire tout système d'information conçu pour donner le maximum de support à la performance. Ces exigences, au nombre de huit, ont été proposées par T. F. Gilbert dans son ouvrage intitulé *Human Competence* (New York, McGraw Hill, 1978, p. 178-179). Elles sont présentées sous la forme de huit étapes allant de l'identification des réalisations attendues (n° 1) à l'identification d'actions visant à apporter un correctif spécifique à des aspects d'une piètre performance (n° 8) en passant par la description de la manière dont la performance sera évaluée et pourquoi on procédera ainsi (n° 3) et l'identification de personnes dont les performances sont exemplaires de même que l'identification de ressources disponibles dont on peut se servir pour devenir comme ces personnes exemplaires dans leurs performances (n° 5).

8° Dernier élément que je soulignerai : l'ouvrage est intéressant et utile par les exemples et les outils qu'il renferme :

- ◆ des exemples de tâches complexes ;
- ◆ une liste de critères formulés par Lauren Resnick et qu'on peut associer aux formes supérieures de la pensée (*higher order thinking* – p. 215) ;
- ◆ une liste de critères pour juger de l'authenticité de tests ou d'épreuves visant à évaluer le rendement intellectuel des élèves (p. 239-240) ;
- ◆ une liste de standards de performance (p. 286-288).

Quand on termine la lecture de cet ouvrage, on ne s'étonne pas que Grant Wiggins ait écrit un article, à large diffusion, intitulé : « Creating Tests Worth

Taking » (*Educational Leadership*, vol. 49, n° 8, mai 1992, p. 26-33) et qu'il soit considéré comme une figure de proue au sein du mouvement faisant la promotion d'une évaluation dite « authentique ». ■

Source : *Le Relais. Journal pédagogique de l'Assemblée générale*, PERFORMA collégial, Université de Sherbrooke, vol. 4, n° 1, janvier 1995, p. 37-49.

NOTES

1. Dans la documentation américaine, on emploie l'expression « *alternative assessment* » pour désigner des pratiques qui ne sont pas forcément inédites, mais qui apparaissent nouvelles au regard des pratiques conventionnelles, très largement répandues. C'est dans ce sens que je parle de « nouvelles façons d'évaluer ».
2. J'ai eu beaucoup de difficulté à traduire cette expression. S'agit-il d'un savoir restreint, circonscrit, limité, ou d'un savoir cloisonné dans le sens qu'il serait considéré comme un en-soi, qui n'est pas mis en relation avec autre chose, qui n'est pas mobilisé dans le cadre d'une tâche complexe ?