

Mai 2016, numéro 16

L'évaluation sommative

Sélection des ressources et rédaction

Robert Howe, Consultant en pédagogie de l'enseignement supérieur et spécialiste en évaluation

Recherche documentaire

Andrée Dagenais, Bibliothécaire professionnelle au CDC

Présentation

L'évaluation sommative est un sujet presque banal. Tous les enseignants évaluent les apprentissages de leurs étudiants, obligatoirement de façon « sommative » et presque toujours de façon « formative ». Pour plusieurs, l'évaluation sommative est l'évaluation « qui compte », et l'évaluation formative, celle qui « ne compte pas ».

La présente étude documentaire pourrait paraître surannée à certains, car presque tous les textes qui traitent d'évaluation des apprentissages situent leur propos, explicitement ou implicitement, dans le cadre de l'évaluation sommative. Toutefois, il importe de prendre du recul afin de renouer avec les principes de base en mesure et en évaluation. C'est ce que nous vous proposons dans cette sélection de documents, qui constituent autant de références propices à la réflexion.

Sommaire

1. Introduction
2. La documentation fondamentale
3. Les classiques québécois
4. Les guides contemporains
5. Les articles
6. Pour aller plus loin

1. Introduction

Comme dans toute activité des sciences humaines, l'évaluation des apprentissages n'est pas une affaire de recettes ou de solutions toutes faites. Ici comme ailleurs, on fait face à des problèmes complexes, jamais identiques, et toujours à approcher avec circonspection. En évaluation des traits humains, une attitude de prudence et de nuances exige du jugement et une réflexion méthodique. Ce jugement, chacun doit l'éduquer pour soi-même en fréquentant, entre autres, des auteurs qui, quoique parfois « anciens », réfléchissent « à haute voix » sous nos yeux dans leurs publications. Ce sont ces auteurs qui nous intéressent ici. Ce sont eux qui inspirent les chercheurs d'aujourd'hui. Ils ont, justement, contribué à éduquer le jugement professionnel des artisans contemporains en mesure et évaluation tant au Québec qu'ailleurs.

Faisons donc un petit tour au pays de la réflexion fondamentale en évaluation sommative. Dans ces quelques fiches de lecture, on trouvera un échantillon de la documentation la plus pertinente de ces auteurs qui donnent du sens aux notions de base, celles qui nous servent aujourd'hui à nommer les choses et à prendre les meilleures décisions dans nos pratiques en mesure et évaluation sommative.

Certaines des références qui suivent ne sont pas récentes. Elles sont retenues ici parce qu'elles constituent assurément des sources déterminantes du savoir contemporain en mesure et évaluation. Les praticiens et chercheurs d'aujourd'hui s'y réfèrent comme un pilote maritime se reporte à ses balises. Qui peut ignorer le phare qu'est Benjamin Bloom, cet auteur et éditeur qui, il y a plus de 50 ans, nous a donné sa célèbre taxonomie et, en 1971, son indispensable *Handbook*?

Si nous parlons de littéracie pour évoquer les nécessaires fondements culturels et professionnels de l'activité humaine, on inclura ces ressources, entre autres, au chapitre de la « formation fondamentale » en mesure et évaluation.

Dans le contexte francophone québécois, nous nommerons quelques auteurs incontournables qui, parfois un peu oubliés aujourd'hui, ont toujours le mérite de constituer des havres d'une pensée analytique et méthodique en évaluation.

Bonne lecture!

2. La documentation fondamentale



■ **BLOOM, Benjamin S., J.T. HASTINGS et G.F. MADAUS (1971). *Handbook on Formative and Summative Evaluation of Student Learning*, New York, McGraw-Hill, 923 pages. (Disponible au CDC, cote [718022](#))**

Voilà un classique majeur. Un essentiel de la pédagogie et de l'évaluation. Destiné aux enseignants en salle de classe, il traite des « règles de l'art » de l'évaluation. Même s'il date de loin et que les approches ont évolué, notamment par les sciences cognitives et la focalisation sur le développement des compétences, les principes fondamentaux qu'il contient sont toujours transposables aujourd'hui. Le génie de Benjamin Bloom et de ses coéquipiers s'y lit dans le confort de leur autorité morale et professionnelle. Ces chercheurs-là appartiennent aux pionniers de la pédagogie moderne. Plusieurs des principes de bonnes pratiques en évaluation sont inspirés du *Handbook*. Ce manuel a une valeur historique : il sonne l'amarce d'un regard pédagogique sur le psychométrique, le début de la transition du docimologique¹ vers la « mesure et évaluation » – processus qui a évolué, notamment depuis la publication de l'ouvrage. Rappelons, par exemple, que c'est ce dernier qui a introduit dans la salle de classe le concept d'évaluation formative, utilisé jusqu'alors dans un autre contexte. Bien téméraire serait celui qui affirmerait que ce manuel n'a plus de valeur aujourd'hui.

Le chapitre quatre de ce manuel est dédié à l'évaluation sommative, que Bloom, Hastings et Madaus décrivent à partir de ses caractéristiques. La première caractéristique porte sur l'intention, l'utilité, le « pourquoi » de l'évaluation sommative : on génère une note, on certifie l'atteinte d'objectifs d'apprentissage.

La deuxième a trait aux moments lors desquels l'évaluation sommative intervient dans le contexte de l'apprentissage. Cette évaluation est effectuée à la fin d'apprentissages substantiels, à la fin d'un cours et à des moments charnières de ce cours. C'est là qu'on pose la question « L'étudiant a-t-il appris? ».

La troisième caractéristique concerne le caractère global, holistique de l'évaluation sommative, par opposition à l'évaluation formative, qui est plus analytique, proche des détails et des composantes des apprentissages. L'évaluation sommative porte sur un jugement beaucoup plus général, au sujet du degré d'atteinte d'objectifs substantiels, globaux, rassembleurs qui seront atteints sur l'ensemble d'un cours ou suite à une partie substantielle d'une formation.

Bref, l'évaluation sommative fait le point. La francophonie a traduit « *summative* » par « *sommative* », et certains pourraient confondre ce mot avec l'idée de « faire la somme, l'addition », ce qui évoquerait une addition de notes de plusieurs moments d'évaluation. Or, dans son esprit, dans son intention, « *summative evaluation* » se veut un moment de bilan (*to sum up*) portant sur l'essentiel, pas sur le détail.

¹ C'est en 1920 que le Français Henri Piéron et un collègue proposent le mot « docimologie » pour désigner la « science des examens » ou « science de l'épreuve ».

Dans ce chapitre, on évoque pour la première fois le concept de tableau de spécification. Cette matrice (n'oublions pas que nous sommes en 1971) présente, en abscisse et en ordonnée, une grille de rencontre entre, d'une part, les contenus à enseigner et, d'autre part, les objectifs à atteindre pour chacun de ces contenus. Évidemment, à cette époque, on n'en est pas encore à focaliser en toute priorité sur les compétences à développer. Les contenus prévalent alors, quoique le discours pédagogique de Bloom et autres mène ultimement au développement de compétences (*skills and abilities*). Très général, le tableau de spécification nomme les grandes catégories de contenu et, de là, décrit sommairement les dispositifs qui doivent, à terme, permettre à l'évaluateur d'être témoin des apprentissages. Il s'agit surtout d'instruments traditionnels de type questionnaires à choix multiples. Bloom propose le tableau de spécification comme outil d'analyse favorisant la cohérence entre ce qu'on veut enseigner et ce qu'on évaluera, à terme.

L'étape suivante dans la préparation d'un instrument d'évaluation sommative sera d'assembler les tâches ou questions dans une structure cohérente, avec consignes et directives, afin que l'étudiant compétent puisse démontrer... sa compétence. Enfin, on développera une stratégie de notation, un barème, un corrigé duquel dériveront un symbole, chiffré ou autre, ou une note, pour prise de décision, pour certification.

Si l'on accepte de lire « au travers » de la culture psychométrique de l'époque, on y relèvera avec profit les fondements de la fonction de l'évaluation sommative. Les « tests » ou dispositifs d'évaluation servent habituellement à l'attribution de notes, à la certification des habiletés et des compétences, à la prédiction de réussites dans des formations subséquentes, à la détermination du point de départ de la suite de la formation, à la rétroaction, et enfin, à la comparaison des résultats de la formation chez divers groupes.

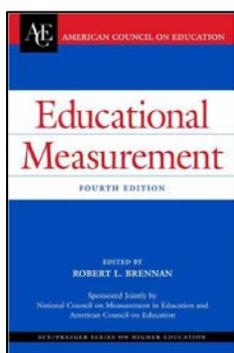
En 1971, nous étions dans une ère d'évaluation normative. Toutefois, avec une lecture consciente et éclairée, nous pouvons aisément transposer dans notre contexte actuel d'évaluation critériée les principes recommandés quant aux soins à prendre dans la sélection des questions ou tâches à soumettre aux étudiants pour en générer des notes signifiantes et crédibles. Bien entendu, nous ne parlons plus, depuis plusieurs années, d'*indices de discrimination* ou de *difficulté*. Mais la représentativité des questions quant à ce qui est mesuré doit s'appuyer sur un jugement professionnel éduqué et structuré, entre autres, par une grille (tableau de spécification²) méticuleusement établie et validée. Ce chapitre de Bloom décrit la structure et l'usage de cette grille.

En plus de générer une note, l'évaluation sommative certifie que l'étudiant possède des habiletés, des connaissances, des compétences... « à ce moment-là » (« *at least at that time* »). C'est la source de ce que Scallon décrira, plus tard, sous le concept d'évaluation certificative. Dans cette fonction, l'évaluation sommative doit permettre d'inférer que l'étudiant peut entreprendre l'étude de textes d'approfondissement, possède les bases pour commencer un cours plus poussé, a les habiletés requises pour entamer la résolution de problèmes dans ce laboratoire, etc. Elle porte sur le niveau atteint dans l'apprentissage. Ceci implique que les tâches ou questions soumises dans le dispositif d'évaluation sont

² En 1996, un groupe de travail de PERFORMA appelé « Pôle de l'Est » a publié un document intitulé *Processus de planification d'un cours centré sur le développement d'une compétence*. Au chapitre 5 de ce texte, le groupe crée, caractérise et recommande un tableau de planification qu'il nomme « vue synoptique ». Ce tableau est issu de la même préoccupation que le « tableau de spécification » de Bloom et se veut pertinent dans la recherche de la cohérence entre ce qui est enseigné et ce qui sera évalué. Voir : <http://www.cdc.qc.ca/performa/720834-deshais-et-al-planification-dev-competence-rimouski-performa-1996.pdf>

choisies soigneusement afin de pouvoir justifier de telles inférences. On se fonde ici sur l'hypothèse à l'effet qu'il existe un niveau de performance au-dessus duquel l'étudiant peut, et au-dessous duquel il ne peut pas. La définition de ce niveau, ce seuil de « réussite », requiert l'exercice du jugement professionnel.

Plus loin dans ce chapitre, Bloom et ses collègues discutent des moyens techniques servant à conférer aux questions d'examen et aux tâches prescrites les qualités métriques de validité et de fidélité dans un esprit d'équité et de crédibilité.



■ **SHEPARD, Lorrie A.** (2006). *Classroom Assessment, Educational Measurement*, R. L. Brennan, eds., Praeger Publishers, p. 623-646. (Disponible au CDC, cote 788985)

Comme les trois précédentes, la quatrième édition d'*Educational Measurement* constitue une somme capitale de tout le savoir contemporain en mesure et évaluation. Cet ouvrage de référence auquel participent les grosses pointures du monde de l'évaluation est sans conteste un incontournable dans le domaine. Ici, les auteurs américains majeurs en mesure et évaluation sont mis à contribution dans des chapitres dédiés à toutes les facettes du domaine, dont plusieurs ne concernent pas l'enseignement supérieur québécois – si l'on pense notamment au

rôle important des tests standardisés aux États-Unis. On y traite de tout : les concepts, les méthodes, les recherches, les enjeux historiques, légaux, pédagogiques et prospectifs.

Le chapitre 17 de ce manuel fondamental porte sur le *classroom assessment*, c'est-à-dire l'évaluation conçue et administrée par les enseignants eux-mêmes pour leurs étudiants. Alimentée, voire stimulée, par les recherches alors récentes en psychologie cognitive, Lorrie Shepard offre ce chapitre toujours extrêmement inspirant sur les liens entre la pédagogie et l'évaluation des apprentissages. Dans le contexte américain, on fait un usage massif de tests standardisés commerciaux. Ce chapitre se penche sur les liens entre, d'une part, l'évaluation formative et sommative et, d'autre part, l'administration massive de tests standardisés externes. D'entrée de jeu, s'appuyant sur la recherche, l'auteure dénonce les effets délétères des tests standardisés sur la qualité de l'enseignement et de l'apprentissage.

Sur le plan de ses fonctions, nous savons tous que l'évaluation sommative consiste à certifier que l'étudiant a fait les apprentissages prévus. La francophonie la désignera aussi comme *évaluation certificative* parce que, dans cette fonction, le jugement évaluatif placera l'étudiant d'un côté ou de l'autre d'un seuil. Ce dernier correspond à ce que l'on conviendra de désigner comme une réussite ou un échec à chacun des objectifs prévus dans les activités d'apprentissage. Ce qui caractérise le plus l'évaluation sommative, bien plus que le geste administratif de certification et de reconnaissance par le collège ou l'université, c'est que cette certification s'accompagne d'une note qui situe l'étudiant sur une échelle quelconque, traduite parfois en lettres, parfois en chiffres, sur 20, sur 100 (comme au collégial québécois), ou avec des symboles différents dépendant des cultures ou des conventions. Selon Shepard, ce phénomène d'une note variable pour chaque individu comporte un certain nombre de dangers ou d'effets pervers, qu'elle qualifie dans son texte de *menaces* ou d'*effets délétères* sur la qualité de l'enseignement et de l'apprentissage.

Shepard évoque des recherches, dont celles de Susan Brookhart, démontrant que même si les enseignants cherchent sincèrement à être équitables et transparents à l'égard de leurs étudiants sur ce qui sera évalué et sur le sens de la note, il reste qu'on observe beaucoup de variabilité dans les pratiques d'attribution des notes. Par exemple, certains introduisent dans la note des variables externes à l'apprentissage, dont la participation, la motivation, la présence, l'exécution de devoirs ou d'exercices, l'effort. Ces pratiques de salle de classe ont cours malgré que les experts déconseillent de « polluer » la note avec ce type de variables non pertinentes. Celles-ci compromettraient la validité de la note comme indicateur de l'atteinte des objectifs.

Ce qui est le plus frappant, dans ce chapitre, c'est que, contrairement à la plupart des manuels sur l'évaluation des apprentissages, Shepard y traite d'abord de l'évaluation formative parce que, pour elle, l'évaluation consiste avant tout à aider l'apprentissage. De là, et c'est toute l'originalité de son approche, elle présente l'évaluation sommative en second lieu en mettant en relief ses effets pervers, notamment dans sa fonction d'attribution de notes au bulletin. On y lit les menaces sur le processus pédagogique d'enseignement et d'apprentissage, pourtant bien engagé dans la relation prof-étudiant. On y lit les corruptions et les paradoxes que peut créer la cohabitation de l'évaluation formative avec l'évaluation sommative. Pensons, par exemple, au développement de précieuses compétences transversales grâce au travail en équipe ou à l'approche par projets, qui seraient heurtés durement par un esprit de compétition régnant chez des étudiants soucieux de leur cote R ou autres conséquences des notes individuelles, notamment pour leur admissibilité à un autre programme ou à certains établissements. Pensons aussi au paradoxe de tout notre système scolaire, dont le discours valorise l'évaluation critériée alors que tous les collèges calculent la moyenne de classe dans une évaluation normative culturellement bien ancrée dans les mœurs. Voilà des paradoxes qui se nourrissent des intentions et conséquences très différentes entre l'évaluation formative et l'évaluation sommative.

Shepard démontre aussi que l'évaluation sommative n'a pas que des effets secondaires négatifs et n'est pas qu'un mandat désagréable pour les enseignants. L'évaluation sommative fait étudier. Grâce à elle, les élèves étudient plus, apprennent plus et sont plus motivés. Ils révisent les contenus et se les approprient de façon personnelle, surtout en fonction de la qualité et du niveau des tâches ou questions qui leur sont prescrites en évaluation sommative. Ce qui est évalué est étudié. Mais, plus encore, Shepard affirme que, dans une perspective cognitive, le système pédagogique idéal serait celui dans lequel les évaluations formative et sommative seraient alignées mutuellement sur des objectifs d'apprentissage. Ainsi, l'évaluation sommative sert à baliser les acquis et à confirmer l'atteinte des objectifs à la suite des activités d'apprentissage nourries par l'évaluation formative.

Shepard garde le cap sur la nécessaire cohérence entre ce qui est évalué en évaluation formative et ce qui est évalué en évaluation sommative. Elle conclura en soutenant que ces deux fonctions de l'évaluation, formative et sommative, peuvent et doivent être en concordance et mutuellement pertinentes si elles sont fondées sur l'intention de développer des compétences et non pas sur les normes sociales.

Ce chapitre de Lorrie Shepard est fondamental et mérite à juste titre d'être mis en valeur ici. Les conseillers pédagogiques et les personnes ressources en évaluation des apprentissages se doivent de l'étudier et de s'en inspirer.

3. Les classiques québécois



■ **MORISSETTE, Dominique** (1993). *Les examens de rendement scolaire*, Sainte-Foy, Les Presses de l'Université Laval. (Disponible au CDC, cote [702152 1993](#))

Le livre de Morissette pourrait être qualifié de « classique » en évaluation. Ce manuel traite de tous les sujets afférents à l'évaluation des apprentissages : la planification, depuis le plan de cours jusqu'à la rédaction des objectifs d'apprentissage, la rédaction des outils traditionnels (de type questions d'examen), l'observation en laboratoire, l'évaluation des attitudes, les résultats de l'examen, la validité et la fidélité des évaluations, les politiques institutionnelles d'évaluation. Bien entendu, il s'agit d'un texte antérieur aux mesures de renouveau du collégial et, surtout, à l'évolution des outils d'évaluation des tâches complexes que sont les grilles à échelles descriptives. Toutefois, le manuel reste adéquat et pertinent parce que les fondements qu'il contient ainsi que son approche pédagogique auprès du lecteur sont et restent transposables au discours contemporain sur l'évaluation dans le contexte d'une focalisation sur le développement de compétences complexes.

Ce qui nous intéresse dans ce livre, c'est le chapitre 13, intitulé « L'évaluation pédagogique ». Ce chapitre, en plus de décrire clairement les très classiques distinctions entre évaluation formative, sommative et diagnostique, se penche plus spécifiquement sur la notion fondamentale de jugement. L'auteur cible sa réflexion sur le jugement professionnel et invite les évaluateurs à porter un tel jugement lors de toute évaluation. C'est là qu'il clarifie bellement les nuances à faire entre objectivité et subjectivité. Mais ce jugement professionnel, que signifie-t-il? Quand s'exerce-t-il dans le processus de l'évaluation? Dans ce chapitre 13, le raisonnement de Morissette se lit dans la simplicité, la limpidité et la vision pédagogique qui caractérisent cet auteur pour qui l'aura connu personnellement.

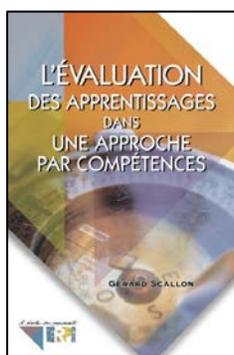
Voilà un texte brillant. Le lire, c'est comprendre les fondements de ce que nous appelons *évaluer* et *juger*. C'est aussi comprendre entre autres, mais surtout, le « pourquoi » de l'évaluation sommative. Le livre de Morissette est parfois oublié, surtout parce qu'il prend de l'âge et que d'autres excellents auteurs ont publié plus récemment. Mais *Les examens de rendement scolaire* reste pertinent et d'actualité pour toute personne qui veut comprendre les fondements de l'acte d'évaluer.

Un fait intéressant : Morissette appelle « examen » toute forme de dispositif qui permet de « mesurer » les apprentissages, incluant ici les travaux longs, les performances de toutes sortes, les prestations écrites ou orales. Cela inclura donc les grilles à échelles descriptives et, bien entendu, les examens traditionnels de type « papier-crayon ». Cette vision généreuse (ou inclusive) du mot *examen* est plausible d'un point de vue étymologique. Lors de l'évaluation, l'étudiant est, en quelque sorte, mis à l'examen lorsque vient le temps de lui demander de montrer qu'il a appris. Pour Morissette, soumettre à l'étudiant un questionnaire, une tâche « authentique » à accomplir, une consigne de travail à remettre dans un objectif de sanction des apprentissages, c'est « examiner » ses apprentissages dans le but d'attester le développement de ses compétences.

Ce chapitre 13 nous offre d'autres belles réflexions sur la nature de l'évaluation, dont d'excellentes distinctions à faire entre :

- jugement et opinion;
- jugement et décision;
- évaluation normative et évaluation critériée.

Enfin, Morissette propose une discussion sur l'importance de bien clarifier, en amont, les objectifs pédagogiques (objectifs spécifiques terminaux ou intermédiaires) sur lesquels le jugement devra porter. Voilà donc un chapitre que je n'hésite pas à qualifier de lumineux, surtout en ce qui a trait à la réflexion sur le jugement.



■ **SCALLON, Gérard** (2004). *L'évaluation des apprentissages dans une approche par compétences*, Saint-Laurent, Éditions du Renouveau Pédagogique. (Disponible au CDC, cote [729607](#))

Gérard Scallon est certes l'auteur québécois de référence sur l'évaluation sommative. On trouve son nom dans plusieurs autres textes, signe de sa crédibilité. Ce manuel, dédié à l'évaluation des compétences, traite l'évaluation sommative sous plusieurs aspects : l'utilisation des échelles descriptives pour guider le jugement de l'enseignant et de l'étudiant, l'évaluation des savoir-faire et des savoir-être, le portfolio, l'autoévaluation.

Dans une pièce maîtresse de ce livre, Scallon discute de la question de l'intégration de l'évaluation à l'apprentissage. En amont de ce manuel, plusieurs autres auteurs ont aussi affirmé que l'évaluation sommative et l'évaluation formative doivent être intégrées et coordonnées aux activités d'apprentissage. Ici, Scallon rappelle le besoin de contextualiser l'apprentissage de l'étudiant dans l'approche par compétences. L'apprentissage et l'évaluation sommative reposent sur une variété de situations qu'on propose aux étudiants. Si ceux-ci ont échoué, en tout ou en partie, il faut qu'il y ait un feedback, individualisé ou de groupe, de la part de l'enseignant, même si on est en situation d'évaluation sommative (et ceci fait écho aux recommandations de Lorrie Shepard, présentées plus haut dans ce Bulletin). Quelles que soient les situations d'apprentissage, on ne peut porter un jugement sur l'apprentissage que si l'on en a suivi la progression, ce qui suppose que plusieurs situations auront servi de balises tout au long du parcours de l'étudiant. C'est plus tard que viendra l'étape du bilan qu'on appelle *évaluation sommative*. La compétence ne peut pas être inférée exclusivement au terme d'une période d'apprentissage. Elle doit avoir été évaluée de façon continue ou cumulative afin de guider le jugement de l'enseignant sur le développement de cette compétence. C'est dans ce contexte de cohérence entre l'enseignement, l'apprentissage et l'évaluation, tant formative que sommative, qu'on pourra inférer que l'apprentissage a eu lieu.



■ **SCALLON, Gérard (1999).** *L'évaluation sommative et ses rôles multiples*, révisé en mai 2000, Université Laval.

Ce fascicule, disponible sur le site de la Faculté des sciences de l'éducation de l'Université Laval, est spécifiquement dédié au concept d'évaluation sommative. Scallon y décrit l'historique du concept et montre les liens fonctionnels de l'évaluation sommative avec sa fonction certificative. L'auteur décrit comment Scriven, en 1967, distinguait l'évaluation sommative de l'évaluation formative dans le contexte qui l'intéressait (l'évaluation d'un produit, par exemple un programme) et comment Bloom et ses collègues, en 1971, dans le *Handbook* décrit ci-haut, ont adapté ces concepts au monde de l'apprentissage scolaire.

Dans la conception de Scriven, l'évaluation sommative consiste en un jugement final, au terme du processus, lorsque tout est terminé. Cette conception n'est pas possible en éducation, et Scallon démontre pourquoi on a besoin de portraits récents de la progression des étudiants. C'est là qu'il discute d'une évaluation sommative continue. Loin de représenter une addition de points, l'évaluation sommative continue, inspirée de De Landsheere (1974) et autres, consiste à faire des bilans au terme d'une période d'apprentissage, quelle que soit la durée de cette période, mais aussi pendant le processus même d'apprentissage, et non pas seulement à la toute fin. Scallon réfère ainsi à Cardinet (1984) qui parle de « boucles d'apprentissage ». Ces bilans, dans leur fonction sommative, servent à attester, à confirmer que tel ou tel apprentissage est fait (ou pas). Dans ce sens, et c'est un des chevaux de bataille de Scallon, on devrait parler d'*évaluation certificative*, quoique ce terme ne soit pas encore entré dans la culture au Québec.

Dans la dynamique de l'enseignement et de l'apprentissage, on voit clairement les passerelles et les liens fonctionnels entre les trois grandes fonctions de l'évaluation, et Scallon les résume bien ici. Mais ce qui caractérise spécifiquement l'évaluation sommative, c'est la certification des compétences. C'est pourquoi, à l'instar de plusieurs auteurs européens, dont Tourneur (1985), Gérard Scallon préconise ouvertement que l'on adopte (un jour, peut-être) l'appellation d'*évaluation certificative* en lieu et place d'*évaluation sommative*.

4. Les guides contemporains

Au Québec, certains auteurs contemporains ont publié dans les dernières années plusieurs textes qui témoignent de leur réflexion et de l'évolution de leur contribution au savoir sur l'évaluation des apprentissages : essais de maîtrise, articles de revue, rapports de recherche. Nous ne présenterons ici que les documents les plus récents et les plus intégrateurs de ces contributions. Le CDC peut diriger le lecteur vers les autres publications de ces auteurs.



■ **LEROUX, J.L., dir., et GROUPE DE RECHERCHE EN ÉVALUATION DES APPRENTISSAGES (GRÉAC) (2015). *Évaluer les compétences au collégial et à l'université : un guide pratique*, Association québécoise de pédagogie collégiale (AQPC) et Chenelière Éducation, 2015, 688 pages. (Disponible au CDC, cote 789012)**

Ce livre, paru en novembre dernier, est le plus récent des ouvrages publiés dans la collection PERFORMA de l'Association québécoise de pédagogie collégiale (AQPC).

Construit en trois grandes sections, le guide propose 20 chapitres qui traitent, sous autant d'angles de vue, des méthodologies pouvant guider les pratiques évaluatives des acteurs des collèges et des universités.

D'entrée de jeu, *Évaluer les compétences au collégial et à l'université* annonce sa posture. Plutôt que de focaliser sur le « quoi » ou le « pourquoi » de l'évaluation, il aborde spécifiquement la question du « comment » évaluer dans une approche par compétences en enseignement supérieur.

Destinés autant aux enseignants chevronnés qu'aux débutants, ces chapitres proposent des ressources pour soutenir des pratiques de qualité en évaluation dans le cadre d'une telle approche.

Il peut être consulté dans n'importe quel ordre, au besoin et selon les questionnements du lecteur, et est organisé en trois grandes familles de textes :

1. Mise en contexte de l'évaluation en enseignement supérieur
2. Ressources pour évaluer les compétences
3. Défis liés à l'évaluation des compétences

On ne peut pas résumer ici le propos de ces 20 chapitres alors que le manuel fait 688 pages. Voici, dans un échantillon de quelques chapitres, un aperçu très sommaire de la riche diversité du questionnement sur lequel ce livre propose des réponses.

Chapitre 2 : « Exercer son jugement professionnel en enseignement supérieur ». Les auteures définissent la notion de jugement professionnel en évaluation et en font une discussion détaillée. Qui exerce ce jugement? Quand? Pourquoi? Quelles sont les caractéristiques d'un jugement professionnel rigoureux? Quels sont les enjeux sociaux, éthiques et pédagogiques du jugement professionnel en évaluation?

Chapitre 6 : « Concevoir des grilles d'évaluation à échelle descriptive ». Quels sont les avantages des grilles à échelle descriptive? Quels sont les avantages ou désavantages des grilles de type analytique ou de type holistique? Quelles sont les règles de mise en place et d'utilisation des grilles à échelle descriptive?

Chapitre 8 : « La rédaction de questions pour évaluer les apprentissages ». Ce chapitre décrit 10 étapes à suivre lors de la construction d'un instrument de mesure. Certaines, mais pas toutes, entrent dans la construction d'un outil d'évaluation artisanal, pour la salle de classe. Si un enseignant n'utilise que des instruments d'évaluation traditionnels de type QCM, quelle conséquence cela pourrait-il avoir sur l'apprentissage des étudiants ainsi que sur ce qui est enseigné? Quels sont les avantages et les inconvénients (ou limites) des questions à réponse dite « longue »?

Chapitre 13 : « Quelques balises pour mieux guider l'évaluation du travail en équipe »³. Dans le cadre de l'évaluation individuelle du travail fait en équipe, qu'est-ce qui caractérise les grands défis de nature psychologique que nous pouvons rencontrer? Dans une annexe, l'auteur décrit des formules pour « repérer » la contribution individuelle dans les travaux en équipe. Lesquelles présentent des problèmes de nature éthique?

Chapitre 15 : « L'éthique en évaluation : quelques repères pour soutenir le personnel enseignant dans l'action ». Dans une situation de dilemme éthique, quelles balises externes ou internes peuvent nous guider? Quelle différence y a-t-il entre « justice » et « équité »? Comment peut-on « éduquer » notre jugement professionnel concernant les problèmes d'éthique?

Chapitre 19 : « Une démarche pour l'évaluation des attitudes ». En vue de choisir des attitudes pertinentes à développer dans un cours ou un programme, quelles questions peut-on poser? Lorsqu'on veut évaluer une attitude, comment fait-on pour en choisir des indicateurs? Les auteures donnent des exemples d'indicateurs significatifs de la présence de certaines attitudes.

Le guide se concentre sur l'évaluation de la « maîtrise » des compétences et fonde ce concept sur la conscience que la notion de compétence est polysémique, renvoyant à des significations très variées. De là, on met le lecteur en appétit dès l'introduction (page 17) en l'avertissant que « le sens véritable accordé à la notion de compétence a d'importantes répercussions sur les stratégies d'évaluation qui seront mises en œuvre pour inférer la compétence ».

Voilà un guide incontournable, fait à l'intention expresse des enseignants de l'enseignement supérieur.



■ **SCALLON, Gérard (2015). *Des savoirs aux compétences : Explorations en évaluation des apprentissages***, Montréal, Éditions du renouveau pédagogique (ERPI Éducation), Pearson. (Disponible au CDC, cote 788895)

Dans ce livre tout récent, Scallon fait l'état des lieux en évaluation, notamment en évaluation sommative. Fort des écrits des années 2000 sur l'évaluation ainsi que des recherches sur les conceptions de l'apprentissage, l'auteur développe le raisonnement sur les méthodologies de l'évaluation. C'est ainsi qu'il consacre son ouvrage à montrer comment on est passés de l'évaluation de savoirs à l'évaluation de capacités et d'habiletés de plus en plus complexes. Les outils et méthodes changent en fonction des objets à évaluer et Scallon caractérise clairement ces méthodes.

L'auteur traite beaucoup de la notion de jugement. Les objectifs d'apprentissage sont de plus en plus diversifiés, afin de donner suite à des mandats de développement de compétences. Si l'on a longtemps mesuré, quantifié, voire additionné les observations sur les réponses des étudiants, à l'époque où les apprentissages portaient en partie sur les savoirs, Scallon discute ici des méthodes qu'impose une approche plus qualitative de l'évaluation. Évaluer, c'est juger.

³ Notons ici que ce chapitre réfère, entre autres, à l'un des Bulletins du CDC : <https://cdc.qc.ca/bulletin/bulletin-7-evaluation-travaux-equipe-sept-2011.pdf>

Le jugement fait partie intégrante du processus d'évaluation. Puisqu'on vise l'apprentissage de capacités de haut niveau, de savoir-agir qui intègrent plusieurs domaines d'apprentissage, il est devenu impératif de prendre en compte le contexte de la performance de l'étudiant, ainsi que ses progrès, son comportement et son engagement. Cela impose d'adopter des procédés de collecte d'informations multiples et d'y appliquer un jugement professionnel.

Scallon décrit ensuite les trois étapes du jugement : on observe, on infère et on évalue. Dans ce contexte, l'inférence consiste à estimer les caractéristiques d'un individu à partir de son comportement observé dans un échantillon de tâches. C'est ainsi que l'évaluation consistera à placer l'individu en situation de performance. On lui présentera trois types d'informations : une mise en situation (un problème, des données, un enjeu); une tâche à accomplir; des directives (consignes et exigences à respecter lors de la performance que nous observerons).

De là, l'auteur démontre que le jugement, dans l'évaluation sommative, qu'il porte sur la démarche de l'étudiant ou sur un produit, doit reposer sur des outils comme les grilles d'évaluation et des échelles de jugement, plus qualitatifs que quantitatifs. Et puisque les compétences de haut niveau sont le fruit d'un long processus d'apprentissage, Scallon discute aussi des méthodes appropriées pour évaluer la progression de l'étudiant. Il nous présente des exemples concrets de progression ainsi que des balises ou des marqueurs de la progression vers le développement de la compétence. Le livre se termine par une présentation commentée de divers procédés d'évaluation, dont des méthodes d'évaluation du professionnalisme.

5. Les articles

Le Centre de documentation collégiale n'a pas trouvé, dans sa collection, d'articles traitant explicitement d'évaluation sommative. Sauf...



■ **HOWE, Robert (2006). La note de l'évaluation finale d'un cours dans l'approche par compétences : quelques enjeux pédagogiques, *Pédagogie collégiale*, vol. 20, n° 1, automne 2006, p. 10-15.**

Cet article est parfois cité au chapitre de la réflexion sur l'évaluation sommative. On y traite du concept de double seuil de réussite, de la valeur de l'évaluation finale d'un cours, de l'impact de l'évaluation sommative sur la motivation des étudiants et sur l'étude, du facteur humain dans l'erreur de mesure.

La discussion qu'on y trouve prend racine dans la recherche de cohérence que nous observons encore récemment entre la conjugaison de l'approche dite par compétences, la notion de compétence minimale et le poids parfois exagéré que certaines PIEA font porter sur l'évaluation sommative de fin de session dans une perspective « intégratrice ».

6. Pour aller plus loin

Le National Council on Measurement in Education (NCME) est une organisation professionnelle dédiée aux personnes concernées par la mesure et l'évaluation en éducation, la recherche et le développement dans le domaine de l'*assessment* ainsi que la psychométrie. Cette organisation, dont le rédacteur du présent Bulletin est membre, est connue notamment par son site web : <http://www.ncme.org/NCME>

Le NCME rend disponible un [glossaire](#) dans lequel l'évaluation sommative est définie du point de vue de son utilisation :

Summative use of assessments: Using assessments at the end of an instructional segment to determine the level of students' achievement of intended learning outcomes or whether learning is complete enough to warrant advancing the student to the next segment in the sequence. This is contrasted with formative use of assessments.

À propos de l'auteur

Diplômé de l'Université de Montréal, Robert Howe est spécialiste en évaluation. Membre du National Council on Measurement in Education (NCME), il agit comme consultant et formateur auprès de plusieurs collègues dans les domaines de l'évaluation des apprentissages, l'évaluation des programmes, l'évaluation de l'enseignement.

Il a été conseiller pédagogique, formateur dans le programme PERFORMA (Université de Sherbrooke), directeur adjoint au cégep régional de Lanaudière à L'Assomption. Il a été chargé de cours à l'Université de Montréal et, depuis 2012, est chargé de cours à l'Université de Sherbrooke, où il enseigne l'évaluation des apprentissages dans le cadre du microprogramme de formation initiale à l'enseignement collégial.

Il a récemment agi comme expert-conseil dans des dossiers d'arbitrage sur des litiges de droit du travail impliquant des aspects d'évaluation des compétences du personnel.

Depuis 2009, il rédige plusieurs *Bulletins de la documentation collégiale* du CDC et publie des billets dans l'infolettre du Portail du réseau collégial du Québec.

La plupart des documents auxquels ce Bulletin réfère sont accessibles en ligne ou disponibles sur demande au Centre de documentation collégiale (CDC).

Pour plus de détails sur nos services ou pour gérer votre abonnement, consultez notre site Web

www.cdc.qc.ca.