



gingras.yves@uqam.ca

[**analyse**] Quand on parle d'« évaluation de la recherche » de façon générique, on amalgame souvent différents niveaux d'évaluation qu'il faut pourtant distinguer si l'on veut discuter sérieusement des méthodes appropriées à chacun d'eux. Yves Gingras propose ici un survol de chacun de ces niveaux à peu près dans l'ordre de leur apparition historique. Cette approche permet de rappeler l'ancienneté et l'omniprésence de l'évaluation des chercheurs tout en faisant ressortir la spécificité des modes d'évaluation mis de l'avant au cours des deux dernières décennies.

Yves Gingras est professeur au département d'histoire de l'Université du Québec à Montréal (UQAM) et titulaire de la chaire de recherche du Canada en histoire et sociologie des sciences. Il est aussi directeur scientifique de l'Observatoire des sciences et des technologies (OST) de l'UQAM. Ses travaux portent sur l'histoire et la sociologie des sciences et de l'enseignement supérieur et sur l'évaluation de la recherche. Il est l'auteur de plusieurs ouvrages, le plus récent étant *Propos sur les sciences* (Paris, Raisons d'agir, à paraître en 2010).

Les systèmes d'évaluation de la recherche

Les publications dans les revues savantes

On peut faire remonter le début de l'évaluation de la recherche à la création en mars 1665 des *Philosophical Transactions* de la Société royale de Londres. Publiées sous la direction d'Henry Oldenburg, secrétaire de la Société, cette revue constitue le prototype de la revue savante moderne : son contenu se limite aux domaines scientifiques, et elle est constituée de comptes rendus de livres, de reprises d'articles étrangers et d'articles originaux dont le contenu doit être évalué par d'autres membres de la Société royale¹.

Pendant longtemps, l'« évaluation de la recherche » se limitera à l'évaluation par les pairs des contributions de leurs collègues et compétiteurs dans le champ scientifique en formation. Les revues se spécialiseront au cours du XIX^e siècle pour devenir disciplinaires mais, jusqu'au début du XX^e siècle, ce système demeurera assez informel, le rédacteur de la revue décidant s'il a ou non l'expertise d'évaluer lui-même le texte soumis et, dans le second cas, choisissant dans son entourage immédiat un expert pouvant lui faire rapport sur la valeur du texte soumis. Le taux de refus était généralement très bas.

L'originalité et la qualité des enseignants-chercheurs

Un autre niveau d'évaluation est bien sûr celui des individus eux-mêmes en tant que chercheurs. Depuis la création de l'Université de Berlin en 1810, le professeur d'université se doit d'être aussi un chercheur. Les responsables de l'institution demandent

alors l'avis d'experts extérieurs au département concerné avant d'embaucher de nouveaux enseignants chercheurs ou de leur accorder la titularisation (*tenure*) ou une promotion. Ces experts, qui sont aussi des pairs, rendent un jugement global et qualitatif sur la valeur et l'originalité des travaux du candidat. De façon générale, ce niveau d'évaluation est très local et sa forme précise varie énormément d'un département et d'une université à l'autre. Une fois la titularisation obtenue, l'évaluation est moins fréquente et varie énormément selon les pays et les institutions.

La révision des programmes de formation et des activités des départements disciplinaires

Les professeurs et les chercheurs étant réunis au sein de « départements » ou d'unités de recherche, ces niveaux d'organisation deviennent à leur tour des objets d'évaluation. L'évaluation systématique des départements et autres unités de recherche semble toutefois assez récente et nullement systématique avant les années 1970. Jusqu'à tout récemment, cependant, cet exercice était lui aussi essentiellement fondé sur le jugement de pairs provenant d'autres institutions et appelés à visiter les départements, analysant leurs rapports d'activité et jugeant de la qualité globale de l'enseignement et de la recherche sans vraiment user d'indicateurs quantitatifs bien définis. Bien que le nombre de publications soit généralement pris en compte, il fait partie d'un jugement global qui est, *in fine*, qualitatif.

Tout comme l'évaluation des chercheurs et des professeurs, celle des unités a une fonction essentiellement interne à l'université et vise la mise à jour des programmes de formation environ à chaque dix ans. Les centres de recherche, qui sont pour la plupart créés autour d'un objet donné et qui répondent

souvent à des demandes et des besoins externes à l'université, peuvent avoir une durée de vie moins longue que les départements et, dépendant davantage de ressources externes, sont souvent évalués sur une base plus fréquente que les départements universitaires.

L'évaluation des universités

Le tournant le plus important en matière d'évaluation de départements universitaires correspond à la mise en place du Research Assessment Exercise (RAE) en Angleterre au milieu des années 1980. Le premier exercice a eu lieu en 1986 et ensuite environ tous les trois (1989, 1992) ou quatre ans (1996, 2001)². L'originalité de cette approche est qu'elle relie les résultats de l'évaluation de l'ensemble des unités du pays en créant un classement général qui influe ensuite sur le niveau des ressources consenties par l'État à chaque institution : les « meilleurs » obtiennent davantage et les « moins bons » voient leurs crédits diminués. Formule excessivement coûteuse et lourde à gérer, le RAE a donné lieu depuis le début des années 2000 à un très grand nombre d'analyses et de critiques dont plusieurs concluent que la formule est maintenant dépassée.

On peut imaginer l'ampleur du travail quand on sait que l'exercice de 2001 a porté sur 50.000 chercheurs, 2.600 unités et 173 institutions d'enseignement supérieur³. À la suite d'une évaluation en profondeur du programme par un comité présidé par Sir Gareth Roberts et d'une série de consultations, on constate une longue pause et une reprise de l'exercice en 2008. Celui-ci analyse 67 unités regroupées en 15 grands domaines disciplinaires comparables. Les nombreux comités d'évaluation comprennent chacun quinze membres dont des experts internationaux.

Il faut noter ici que ces évaluations sont le fait de comités de pairs qui fondent leur jugement sur un mélange complexe de données diverses. Il ne s'agit pas d'une évaluation « mécanique » fondée seulement sur des données bibliométriques, comme on le croit parfois. En fait, seul le gouvernement de la Flandre utilise un système d'évaluation des universités qui prend en considération les indicateurs bibliométriques explicites et formalisés, en plus du nombre d'étudiants inscrits, pour distribuer les octrois gouvernementaux à ses universités.

L'évaluation des revues savantes

Un aspect entièrement nouveau de l'évaluation a émergé récemment au niveau européen. Il s'agit de l'évaluation des revues savantes elles-mêmes. Il est en effet important de rappeler que, même s'il a toujours existé dans chaque discipline et spécialité une hiérarchie *implicite* des revues, celle-ci n'a jamais eu un caractère officiel ni explicite et, en fait, variait toujours un peu selon les personnes. Bien sûr, les organismes qui soutiennent financièrement un certain nombre de revues les évaluent sur une base régulière

pour décider de renouveler ou non leur appui, mais cela demeure une évaluation par les pairs qui reste interne à l'organisme et ne se transforme pas en *classement* public explicite.

La nuance est cruciale car l'officialisation d'un classement a des effets pervers évidents qui viennent perturber de façon importante le fonctionnement normal des revues. Aussi, contrairement à ce que l'on entend parfois, il n'existe pas en Amérique du Nord de classement *explicite et officiel* des revues savantes. En effet, l'existence du « facteur d'impact » des revues, publié par la firme Thomson Reuters, ne constitue pas à proprement parler un *classement* officiel. Localement, il peut être utilisé à cette fin, mais il s'agit alors d'un classement indigène utilisé par un comité qui en a décidé ainsi, croyant alors se faciliter la tâche. Il n'a aucun effet global sur la discipline ou même sur les pays.

L'approche européenne est tout autre. Lorsque L'AERES ou l'European Science Foundation créent littéralement un classement des revues en sciences humaines, elles officialisent et figent un état de la structure des relations entre revues, qui est toujours dynamique et changeant. Elles génèrent ainsi des effets en retour qui désavantagent les revues plus jeunes et innovantes ou plus spécialisées. Sans compter que la procédure de classement est opaque et arbitraire.

La réaction vive et immédiate des revues anglophones d'histoire et de sociologie des sciences à cet exercice bureaucratique, malheureusement appuyé, il faut le dire, par des chercheurs peut-être bien intentionnés mais qui n'ont pas vraiment réfléchi aux conséquences d'une telle démarche, montre bien que cette initiative européenne ne s'inscrit nullement dans une tradition d'évaluation établie, par exemple, en Amérique du Nord où l'on laisse plutôt aux comités d'évaluation le soin de tenir compte par eux-mêmes de la valeur des lieux de publication. De plus, cette approche ne tient nullement compte du fait bien établi que les articles ne constituent dans les sciences sociales et humaines qu'une faible partie de la production savante qui laisse encore une place centrale aux livres, beaucoup plus cités que les articles. Enfin, ces classements ne tiennent pas compte du fait que, dans la pratique, les autres niveaux d'évaluation incorporent déjà implicitement une hiérarchie des revues.

Pour conclure ce rapide survol, on peut dire sans trop de risques d'erreur que les chercheurs et les enseignants-chercheurs sont probablement les personnes les plus fréquemment évaluées. Il faudrait d'ailleurs quantifier les ressources matérielles et temporelles consacrées par les chercheurs eux-mêmes à ces nombreux exercices d'évaluation pour constater qu'elles équivalent à plusieurs milliers d'heures de travail perdues pour les activités de recherche... ●

¹ Marie Boas Hall, Henry Oldenburg. *Shaping the Royal Society*, Oxford University Press, 2002, p. 84.

² Paul G. Hare, « L'exercice d'évaluation de la recherche au Royaume-Uni, son impact sur les établissements, les départements et les personnes », *Politiques et gestion de l'enseignement supérieur*, 2003, vol. 15, n° 2, p. 47-69.

³ Voir le site du RAE : www.rae.ac.uk

⁴ Vincent Larivière, Éric Archambault, Yves Gingras, Étienne Vignolas-Gagné, « The Place of Serials in Refereencing Practices : Comparing natural Sciences and Engineering with Social Sciences and Humanities », *Journal of the American Society for Information Science and Technology*, June 2006, vol. 57, n° 8, p. 997-1004.