

Nouvelles perspectives en sciences sociales



Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie, Yves Gingras, Paris, Raisons d'agir, 2014

Claude Vautier

Volume 12, Number 1, November 2016

URI: <https://id.erudit.org/iderudit/1038380ar>

DOI: <https://doi.org/10.7202/1038380ar>

[See table of contents](#)

Publisher(s)

Prise de parole

ISSN

1712-8307 (print)

1918-7475 (digital)

[Explore this journal](#)

Cite this review

Vautier, C. (2016). Review of [*Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*, Yves Gingras, Paris, Raisons d'agir, 2014]. *Nouvelles perspectives en sciences sociales*, 12(1), 272–279.
<https://doi.org/10.7202/1038380ar>

Tous droits réservés © Prise de parole, 2016

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie

Yves Gingras, , Paris, Raisons d'agir, 2014

PAR CLAUDE VAUTIER

LEREPS, Université de Toulouse, ENFA,
IEP de Toulouse, UT1 Capitole, UT2J

Issu d'une conférence donnée à l'INRA en 2011, cet ouvrage est salubre pour éclairer le débat, vif en France et dans bien d'autres pays, en Europe et hors d'Europe, autour de l'évaluation du travail de recherche : évaluation individuelle (le dossier individuel qui devrait justifier de la carrière et de son évolution) et collective (évaluation HCERES des laboratoires, en France, classement de Shanghai, par exemple, sur le plan international).

« Parmi les outils utilisés pour mettre au point les nombreux “indicateurs d'excellence” qu'une vision gestionnaire de l'enseignement supérieur et de la recherche tente d'imposer à tous comme une évidence, une place de choix est aujourd'hui accordée à la bibliométrie – méthode de recherche qui consiste à utiliser les publications scientifiques et leurs citations comme indicateur de la production scientifique et de ses usages. » *Cette mise à toutes les sauces* de la bibliométrie conduit, selon l'auteur, à poser « l'équation fautive et malheureuse selon laquelle bibliométrie = évaluation » (p. 9 et 10).

Le travail d'Yves Gingras se présente alors comme à la fois une analyse critique de l'évaluation de la recherche et une défense et illustration de la bibliométrie rendue à ses champs véritables de pertinence.

Ainsi, écrit Gingras à la fin de son introduction, les méthodes bibliométriques « ont un champ d'application beaucoup plus vaste que la seule évaluation de la recherche » et peuvent « se révéler un outil indispensable pour analyser la dynamique de la

science », tandis que l'on peut critiquer « divers abus auxquels ces usages donnent souvent lieu » (p. 11).

Le chapitre I traite des origines de la bibliométrie. Après avoir défini la « bibliométrie » comme un sous-ensemble de la « scientométrie », l'auteur nous renvoie aux années 1920-1930 et à l'utilisation de celle-ci comme instrument (manuel) de gestion des revues. L'informatisation dans les années d'après la Seconde guerre mondiale permettra, surtout dans le courant des années 1970, la création d'une véritable « science de la science » selon les termes employés par l'historien des sciences Derek de Solla Price, utilisant une « métrique » et des « indicateurs de la science », mis au service des politiques de recherche. Les années 1970 et 1980 verront le transfert des méthodes bibliométriques aux usages évaluatifs et les années 1990 marqueront l'implantation durable de ces méthodes dans l'évaluation de la recherche et des chercheurs.

Le deuxième chapitre (« Ce que la bibliométrie nous apprend de la dynamique des sciences »), comme l'auteur le dit très clairement dans son titre, montre les intérêts multiples de la bibliométrie, notamment à partir du moment où l'on va disposer de bases de données informatiques. « Outil pour les historiens et les sociologues des sciences », elle permet de mesurer diverses caractéristiques de la recherche, telle l'association entre publications, citations et subventions de recherche (p. 34), ou encore évolution relative des diverses disciplines, thématiques ou champs de recherche dans le temps (p. 37 et 45); on peut, avec son aide, mettre en évidence le fait que les règles définissant ce qu'est un chercheur « publiant » (p. 39-40) conduisent à dévaloriser certains types de publications, par exemple, les contributions à des ouvrages collectifs, « simplement parce qu'on ne peut pas leur attribuer de "facteur d'impact" comme on le fait pour les revues [...] Cela revient à inverser l'ordre des choses : *les pratiques sont alors adaptées aux critères, alors que ce devrait être l'inverse* » (p. 40). Le développement des réseaux de recherche, l'augmentation de la proportion d'articles signés par plus d'un chercheur, l'internationalisation croissante des collaborations

entre auteurs d'articles, la visualisation circulaire (et non linéaire comme le pensait Auguste Comte, par exemple) des chaînes de relations entre disciplines, rien de cela n'eût été facile, voire possible, sans développement des méthodes bibliométriques.

Mais, comme l'indique le titre de l'ouvrage, des dérives se produisent. Et, en premier lieu, la multiplication des évaluations qui fait l'objet du chapitre III.

Si le ton de l'ouvrage est parfois ferme, il n'est jamais stérilement polémique, du moins est-ce ainsi que je l'ai lu. Yves Gingras, dans son troisième chapitre tente de nous convaincre que « le problème est donc moins celui de l'évaluation elle-même que celui de sa prolifération » (p. 55), ce à quoi il parvient, d'autant plus que ses arguments recourent nombre de ceux qui ont circulé dans le cadre des débats critiques autour des systèmes d'évaluation que les autorités de tutelle de la recherche ont voulu imposer, tant en France autour de l'AERES (aujourd'hui, HCERES) que dans des pays comme l'Australie et la Belgique auxquels l'auteur fait référence.

Après avoir rappelé que l'évaluation des chercheurs n'est pas nouvelle (piquante référence à l'évaluation négative d'un texte proposé par Isaac Newton, en 1672 à la Société Royale de Londres... p. 56), l'auteur parcourt les différents domaines où l'évaluation s'est développée : publications savantes, demandes de subventions pour projets de recherche, enseignements des cycles supérieurs, promotions des enseignants-chercheurs, départements et centres de recherche, universités... (p. 55). Dans tous ces domaines (dont l'auteur explore seulement une partie suffisamment significative pour son propos), la systématisation, la « mécanisation » des indices quantitatifs développant une « apparence d'objectivité » (p. 62), conduit à une véritable « taylorisation » de l'évaluation : « La volonté d'organismes comme l'European Research Council ou l'AERES d'utiliser un classement des revues en A, B et C pour faciliter, sinon mécaniser, l'évaluation individuelle constitue *de facto* une forme de taylorisation de l'évaluation » (p. 74-75). Par ailleurs, le caractère arbitraire des indicateurs issus d'une bibliométrie désaxée

(c'est-à-dire, sortie de son champ de pertinence, trop démultipliée, trop systématisée...) ne cesse d'interroger, sans que s'en suivent de véritables réflexions au niveau institutionnel sur leur pertinence, et leur (il)légitimité. Ainsi, nous dit Gingras, de « l'indice H », proposé dans les années 2000 par un physicien américain, Jorge E. Hirsch, que je laisse au lecteur non familier de ces questions le soin de découvrir (p. 62 à 65)¹. Il en va de même du « Facteur d'impact » (FI) des revues qui conduit à des effets pervers tel celui de la multiplication stratégique des auto-citations pour les revues : ainsi, certaines revues demanderaient à leurs auteurs de citer des articles publiés dans leurs pages (p. 66), pour des raisons ayant moins à voir avec la démonstration, le raisonnement scientifique de l'auteur ainsi incité qu'avec l'augmentation du facteur d'impact de la revue...

Par ailleurs, l'utilisation du facteur d'impact pour évaluer des articles individuels, et donc des chercheurs, est absurde, mais persistante : « Bien que, dès le milieu des années 1990, des experts en bibliométrie n'aient cessé d'attirer l'attention sur l'absurdité d'utiliser le FI des revues pour évaluer les chercheurs, cela n'a pas empêché les pires dérives de la part de "décideurs" et, il faut le souligner, de chercheurs supposément rationnels » (p. 70). Dans cette critique précise mais mesurée, Gingras pointe, entre autres, le caractère arbitraire des classements qui sont mis en œuvre, le doute pesant aussi sur la légitimité des chercheurs nommés dans les commissions chargées des classements des revues ou des évaluations individuelles et collectives... « faire partie de ceux qui déclarent officiellement que telle revue est classée A et telle autre C peut être enivrant... » (p. 80).

À aucun moment, dans cet ouvrage, n'apparaît le sentiment que l'auteur conteste le bien-fondé de l'évaluation des chercheurs comme des organismes de recherche. Rappelant que la mise à disposition de crédits par l'État par le biais de ses agences déconcentrées ou décentralisées suppose une gestion claire des

¹ Indice « défini comme étant égal au nombre d'articles n qu'un chercheur a publiés et qui ont obtenus au moins n citations chacun depuis leur publication »... (p. 63).

fonds publics, il assume la nécessité des contrôles et des évaluations, antérieures et postérieures au projet. Ce qu'il met en cause, c'est l'esprit même dans lequel sont menées ces évaluations. Après avoir décrit le système d'évaluation de la recherche mis en place sous Margaret Thatcher et s'intéressant à celui développé en 1990 en Australie et en 2003 en Belgique, il écrit : « Il s'agit alors de produire des évaluations non plus en vue d'améliorer les pratiques et les activités, mais bien afin de sanctionner par des mesures budgétaires les organismes considérés comme les moins "performants" au regard des indicateurs choisis » (p. 73).

De même, ce n'est pas le caractère quantitatif des indicateurs tirés de la bibliométrie qui est contesté, mais leur inadéquation aux objectifs poursuivis. Des indicateurs quantitatifs et qualitatifs ont bien vocation à être complémentaires, mais leur utilité est fonction de leur pertinence face au problème à résoudre (promouvoir un chercheur, améliorer les résultats de la recherche...), de la légitimité de leur mise en œuvre, de ceux qui les mettent en œuvre et, bien entendu, des objectifs qui sont fixés...

Le titre du quatrième et dernier chapitre (« L'évaluation de la recherche ») illustre ce positionnement de l'auteur.

Dans ce chapitre, Yves Gingras nous dit que « ce qui est le plus surprenant dans la multiplication des indicateurs est qu'aucun critère bien défini n'est utilisé systématiquement pour en contrôler la validité » (p. 83). Son chapitre final va donc lui permettre à la fois de définir ce que doivent être, selon lui, les bases de toute évaluation, et les critères de validité des indicateurs.

L'auteur nous rappelle donc très opportunément que l'on ne doit « pas choisir d'abord les indicateurs et ensuite adapter la mission pour les maximiser, mais au contraire établir d'abord la mission et les objectifs pour ensuite tenter de trouver les indicateurs qui les reflètent vraiment » (p. 84). Par ailleurs, que valent les données que l'on pense utiles pour une évaluation si l'on ne s'assure pas « qu'elles sont disponibles, à des coûts abordables et qu'elles sont fiables » (p. 85)? L'exemple de Google Scholar, base bibliométrique gratuite mais susceptible de manipulations et non

fiable, illustre fort bien le propos (et le dilemme, puisque les deux autres bases disponibles, Web of Science (WoS) et Scopus, sont très onéreuses).

Les indicateurs bibliométriques, dit encore Yves Gingras, ont des limites qu'il convient de connaître pour les utiliser, sous peine, dans le cas contraire, de leur ôter toute pertinence. Ainsi, le caractère plus ou moins local de l'objet d'étude modifie-t-il le sens du résultat de l'indicateur : les collaborations internationales vont être plus fortes pour l'étude des trous noirs que pour l'agriculture du Nord de la France...

De fait, un indicateur est toujours incapable de cristalliser l'ensemble des informations nécessaires à une évaluation éclairée. Ou plutôt, en cristallisant plusieurs dimensions en un chiffre unique, il efface une grande partie des informations. Dans ce cas, comme dans celui où l'on choisit les indicateurs d'abord (parce que disponibles, connus, faciles à manipuler...), on cède au « syndrome du lampadaire qui porte à chercher ses clés dans une zone éclairée plutôt qu'à l'endroit précis (mais sombre) où elles ont en fait été égarées » (p. 90)...

Nous sommes donc conviés à « évaluer les indicateurs » et, pour ce faire, à respecter trois propriétés pour qu'un indicateur soit valide : l'adéquation à l'objet, l'homogénéité de la mesure, le respect de l'inertie propre au projet.

Ainsi doit-on se demander si les résultats que produit l'indicateur sont du bon ordre de grandeur (adéquation) : « Ainsi, le niveau d'investissement en R&D est une bonne première indication concernant l'intensité de la recherche dans un pays. Mais supposons que l'on veuille évaluer l'impact scientifique d'un auteur... » (p. 93).

Il faut aussi s'assurer que l'indicateur permet d'effectuer des comparaisons (homogénéité) : « un indicateur homogène (à l'échelle d'un pays) de l'intensité de l'activité de recherche est fourni par le nombre d'articles publiés dans les principales revues scientifiques. Il s'agit là d'une mesure d'extrants qui peut aussi être comparée à celles d'intrants, comme la valeur des investissements en recherche » (p. 94).

Enfin, respecter l'inertie propre au projet signifie, par exemple, que l'évaluation des universités n'a pas à être réalisée chaque année (il est improbable que ce classement change vraiment d'une année sur l'autre), mais que l'évaluation d'un étudiant peut être répétée sur une période bien plus courte, puisque le processus d'apprentissage en cours doit donner des variations sensibles d'un semestre sur l'autre, par exemple.

En fait, selon l'auteur, la seule façon d'expliquer la persistance de l'utilisation des classements du type de celui de Shanghai, dont l'invalidité a été maintes fois démontrée, est de considérer que ces classements (et plus généralement, les évaluations, Gingras stipulant bien la différence et la contigüité des deux notions) sont utilisés à des fins de marketing pour vendre au mieux les formations supérieures et les établissements qui les dispensent (ou les "vendent") : « L'application aux universités des techniques de marketing est problématique, car les universités sont censées incarner la quête de la vérité, et non tenter de vendre à tout prix des "produits" » (p. 99). Manipulation des indices et des classements, « corruption intellectuelle engendrée par la course aux classements » (p. 103), utilisation stratégique pour appuyer ou contrarier les réformes du système d'enseignement et de recherche, toutes ces dérives sont rendues possibles par les nouvelles représentations des universités assimilant celles-ci à des entreprises commerciales devant s'imposer sur un « marché universitaire » en recherche d'une « clientèle étrangère lucrative qui viendrait ainsi combler les revenus déclinants émanant des gouvernements et la baisse des étudiants locaux » (p. 104).

Au-delà des arrangements avec le ciel auxquels se livrent certains chercheurs avides de pouvoir et de prestige, susceptibles, donc, de soutenir l'insoutenable intellectuellement, se profile une perte de sens de l'institution universitaire. Dans une communication au Congrès International « Quelle Université pour demain? Vers une évolution transdisciplinaire de l'Université » (Locarno, Suisse, 30 avril-2 mai 1998), Edgar Morin s'exprimait ainsi :

Alors, l'Université? J'ai déjà dit qu'il nous fallait dépasser l'alternative : l'Université doit-elle s'adapter à la modernité, ou adapter la modernité

à elle. Elle doit faire l'un et l'autre alors qu'elle est violemment entraînée vers le premier pôle. Adapter la modernité à l'Université, c'est rééquilibrer la tendance vers la professionnalisation. La sur-adaptativité est un danger qu'avait bien vu Humbolt puisqu'il disait que l'Université a pour mission de donner les bases de connaissances de la culture et que l'enseignement professionnel doit relever d'écoles spécialisées. L'Université est avant tout le lieu de transmission et de transformation de l'ensemble des savoirs, des idées, des valeurs, de la culture. À partir du moment où l'on pense que l'Université a principalement ce rôle, elle apparaît dans sa dimension trans-séculaire; elle porte en elle un héritage culturel, collectif, qui n'est pas seulement celui de la nation mais de l'humanité...²

Les dérives de l'utilisation de la bibliométrie, outre le préjudice porté à une science utile, décrédibilisent les procédures d'évaluation qui, nous dit Gingras, « sont essentielles à la bonne marche de toute organisation » (p. 107). Et, se référant au conte d'Andersen, "Les habits neufs de l'Empereur", il se demande si « les dirigeants universitaires feront comme l'empereur et continueront encore longtemps à porter chaque année les faux habits taillés par les vendeurs de classements ou s'ils entendront la voix de la raison » (p. 106).

S'ils continuaient à porter leurs faux vêtements taillés par de faux tailleurs, sans se préoccuper d'« ouvrir les "boîtes noires" afin d'interroger la nature et la valeur de chaque indicateur utilisé pour fonder une évaluation » (p. 108), ce n'est pas seulement l'empereur, mais, à terme, tous les habitants de la Cité qui se retrouveraient nus.

² Edgar Morin, « Réforme de pensée, transdisciplinarité, réforme de l'université », CIRET, *Bulletin interactif du Centre International de Recherche et Etudes transdisciplinaires*, n° 12, février 1998.