

Dérives et effets pervers de l'évaluation quantitative de la recherche : sur les mauvais usages de la bibliométrie

Yves Gingras

Volume 28, Number 2, 2015

URI: <https://id.erudit.org/iderudit/1032829ar>

DOI: <https://doi.org/10.7202/1032829ar>

[See table of contents](#)

Publisher(s)

Editions EMS – In Quarto SARL

ISSN

0776-5436 (print)

1918-9699 (digital)

[Explore this journal](#)

Cite this document

Gingras, Y. (2015). Dérives et effets pervers de l'évaluation quantitative de la recherche : sur les mauvais usages de la bibliométrie. *Revue internationale P.M.E.*, 28(2), 7–14. <https://doi.org/10.7202/1032829ar>

Yves GINGRAS

UQAM

DÉRIVES ET EFFETS PERVERS DE L'ÉVALUATION QUANTITATIVE DE LA RECHERCHE : SUR LES MAUVAIS USAGES DE LA BIBLIOMÉTRIE¹

Avec l'arrivée en milieu universitaire de l'idéologie néolibérale adossée aux techniques du nouveau management public avec ses « tableaux de bord », surtout depuis les années 1990, les chercheurs et les administrateurs utilisent de plus en plus souvent les mots « évaluation », « facteurs d'impact », « indice h ». Le monde de la recherche et de l'enseignement supérieur particulièrement en Europe (voir, notamment, Abélard, 2003 ; Charle et Soulié, 2007 ; Schultheis, Escoda et Cousin, 2008 et Brisset, 2009) est ainsi la proie d'une véritable fièvre de l'évaluation. On veut tout évaluer : les enseignants, les professeurs, les chercheurs, les programmes de formation et les universités. Les indicateurs « d'excellence » et de « qualité » se multiplient sans que l'on sache toujours sur quelles bases ils ont été construits. 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 indicateurs de la production scientifique et de ses usages. Que ce soit pour classer les universités, les laboratoires ou les chercheurs, le calcul du nombre de publications et des citations qu'elles reçoivent sert souvent de mesure « objective » de la valeur des résultats de recherche des uns et des autres.

Il est donc important de rappeler, même brièvement, les nombreux dangers que comporte l'usage simpliste qui tend à se répandre de l'utilisation mécanique d'indicateurs bibliométriques censés mesurer de façon « objective » la productivité et l'impact scientifique des chercheurs. Nous nous limiterons ici à analyser les usages des deux principaux indicateurs amplement utilisés tant par les chercheurs que par les administrateurs de la recherche, à savoir le « facteur d'impact » des revues et « l'indice h » des chercheurs. Nous nous pencherons aussi sur les effets pervers des usages simplistes de mauvais indicateurs sur la dynamique de la recherche scientifique, particulièrement dans le domaine des sciences humaines et sociales.

1 Étant donné, d'une part, la grande pertinence du thème abordé ici et, d'autre part, la parution récente de l'ouvrage d'Yves Gingras sur ce thème, l'auteur et le responsable de cette rubrique se sont entendus pour que le contenu de cette chronique soit tiré en majeure partie de l'ouvrage suivant : Gingras, Y. (2014). *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*. Paris, Éditions Raisons d'Agir. Nous remercions les responsables de la maison d'édition d'avoir accepté que des emprunts significatifs figurent dans la présente chronique, même si la structure de leur présentation est originale.

LES MAUVAIS USAGES DU FACTEUR D'IMPACT

Calculé et publié chaque année depuis 1975 par la compagnie Thomson Reuters dans le *Journal Citation Reports* à partir des données du Science Citation Index (SCI), le facteur d'impact (FI) d'une revue consiste en une simple moyenne arithmétique du nombre de citations obtenues par les articles d'une revue donnée sur une période de deux ans. Le FI caractérise donc des revues, et non pas des articles (Archambault et Larivière, 2009). Ainsi, le FI d'une revue pour 2006 est obtenu en divisant le nombre de citations que les articles publiés en 2004 et 2005 ont obtenues en 2006 par le nombre d'articles publiés au cours de ces deux années (2004 et 2005). Le choix d'une fenêtre courte de deux ans pour évaluer le FI n'est pas anodin. En effet, on a tendance à opposer les FI élevés des revues scientifiques (comme *Science* et *Nature*) aux FI très faibles des revues de sciences sociales. Or, cela est en grande partie un artefact dû au fait que la temporalité des sciences sociales est plus longue que celle des sciences de la nature. En effet, il suffit d'augmenter la fenêtre des citations à dix ans pour que les FI de ces disciplines atteignent des valeurs comparables à celles des revues scientifiques (Glänzel et Moed, 2002). Il est évident que les FI des revues de différents domaines ne sont pas comparables entre eux et que leur valeur numérique n'a pas de sens en elle-même, mais seulement si on les compare à des revues du même domaine.

Bien que, dès le milieu des années 1990, les chercheurs 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 (Seglen, 1997). Ces derniers se sont mis à utiliser le facteur d'impact des revues pour évaluer les chercheurs, confondant ainsi les articles et les revues. Pis encore, des responsables gouvernementaux ou d'organismes de recherche en sont venus, dans certains pays (Pakistan, Chine, Corée du Sud, Japon), à instituer des systèmes de primes fondés directement sur la valeur numérique du facteur d'impact des revues ! Selon la revue *Nature* 441 (p. 792) du 15 juin 2006, le ministère de la Science du Pakistan calculait, en 2006, la somme des facteurs d'impact des articles sur une année pour fixer une prime variant entre 1 000 et 20 000 dollars ! En Chine, l'Institut de biophysique de Beijing avait un système semblable : un FI entre 3 et 5 rapporte 2 000 yuans par point, et 7 000 yuans par point si le FI est au-dessus de 10. Il est pourtant impossible que le FI d'une revue de mathématiques (par exemple) n'ait jamais la valeur de celui d'une revue de médecine ! Or aucune personne sensée ne peut croire que les articles de médecine sont tous supérieurs aux articles de mathématiques et justifient donc d'accorder à leurs auteurs une prime plus importante.

Dernier exemple montrant le genre de corruption intellectuelle engendrée par la course aux classements : certaines universités contactent des chercheurs très cités qui sont employés par d'autres institutions et leur offrent d'ajouter leur adresse dans leurs publications en échange d'une rémunération. Ces affiliations factices, auxquelles aucune tâche d'enseignement ou de recherche n'est attachée, et dont les chercheurs qui y participent sont complices, permettent à des institutions marginales d'améliorer facilement leur position dans les classements des

universités sans avoir à créer de véritables laboratoires². Ces cas extrêmes devraient suffire pour mettre en garde les gestionnaires d'université, ou leurs chargés de communication, contre les usages médiatiques de tels classements douteux. En somme, mieux vaut regarder à l'intérieur de la « boîte noire » des classements plutôt que de l'accepter telle quelle comme si elle contenait un beau cadeau de bienvenue...

Enfin, il n'est pas inutile de noter la fausse précision des facteurs d'impact, qui sont « mesurés » à quatre chiffres significatifs ! En sciences, rares sont les phénomènes naturels que l'on connaît avec une telle précision. Qui veut savoir que la température de demain sera de... 20,233 degrés ? Pourquoi ne pas se limiter au premier chiffre ? Tout simplement parce que les classements perdraient beaucoup, sinon même entièrement, de leur valeur car la plupart des revues seraient *ex æquo*...

L'usage abusif de classements et d'indicateurs faussement précis constitue en somme un comportement qui trahit l'ignorance des propriétés des indicateurs utilisés. Seul l'opportunisme des chercheurs, qui profitent de primes mal calculées, et des revues, qui profitent de l'usage évaluatif des facteurs d'impact, peut les amener à croire, ou à feindre de croire qu'un tel système est juste et rationnel.

L'ÉPIDÉMIE DE « L'INDICE H »

Il est devenu courant de voir des chercheurs indiquer sur leur page Facebook ou dans leur *curriculum vitae* leur « indice h ». Au milieu des années 2000, alors que les milieux scientifiques avaient commencé à concocter des indices bibliométriques pour rendre les évaluations individuelles plus « objectives », le physicien américain Jorge E. Hirsch, de l'Université de Californie à San Diego, y est allé de sa proposition : l'indice h. Cet indice est défini comme étant égal au nombre d'articles n qu'un chercheur a publiés et qui ont obtenu au moins n citations chacun depuis leur publication. Par exemple, un auteur qui a publié 20 articles parmi lesquels 10 sont cités au moins 10 fois chacun aura un indice h de 10. Le caractère improvisé de cet indice se voit déjà au titre même de l'article paru dans une revue pourtant considérée comme « prestigieuse », les *Proceedings* de l'Académie nationale des sciences des États-Unis : « un indice pour quantifier la production (*output*) scientifique d'un chercheur » (Hirsch, 2005). En fait, cet indice n'est ni une mesure de quantité (*output*), ni une mesure de qualité ou d'impact, mais un composite des deux. Il combine de façon arbitraire le nombre d'articles publiés et le nombre de citations obtenues. Cet indice est censé contrer l'usage du seul nombre d'articles, lequel ne tient pas compte de leur « qualité ». Selon un cliché répandu, il serait facile de publier un très grand nombre d'articles de piètre qualité ; en conséquence, le nombre d'articles publiés ne serait pas un bon indicateur de la qualité d'un chercheur. Le problème c'est qu'il a rapidement été démontré que l'indice h est lui-même très fortement corrélé au nombre total d'articles et se révèle ainsi redondant (voir van Leeuwen, 2008) !

² Voir en particulier Bhattacharjee (2011). Voir aussi les nombreuses réactions dans *Science* 335(6072) du 2 mars 2012, 1040-1042. Le lecteur intéressé pourra également prendre connaissance d'une analyse bibliométrique que j'ai publiée récemment (Gingras, 2014) et qui met en évidence ces pratiques.

Pis encore, il n'a aucune des propriétés de base que doit posséder un bon indicateur. Comme l'ont montré Waltman et van Eck (2011), l'indice h est en réalité incohérent dans la manière dont il classe des chercheurs dont le nombre de citations augmente de façon proportionnelle. Ces auteurs en concluent que l'indice h « ne peut être considéré comme un indicateur approprié de l'impact scientifique global d'un chercheur ». Cet indice mal construit est même dangereux lorsqu'il est utilisé comme aide à la prise de décisions car il peut générer des effets pervers. Un exemple simple suffit à le démontrer. Comparons deux cas de figure : un jeune chercheur a publié seulement 5 articles, mais ceux-ci ont été cités 60 fois chacun (pour une période de temps donnée) ; un second chercheur, du même âge, est deux fois plus prolifique et possède à son actif 10 articles, cités 11 fois chacun. Ce second chercheur a donc un indice h de 10, alors que le premier a un indice h de 5 seulement. Peut-on en conclure que le second est deux fois « meilleur » que le premier et devrait donc être embauché ou promu ? Bien sûr que non... On voit ici que l'indice h ne mesure pas vraiment la qualité relative de deux chercheurs et est donc un indicateur techniquement invalide. Malgré ces défauts techniques rédhibitoires, l'usage de l'indice h s'est généralisé dans plusieurs disciplines scientifiques. Il semble taillé sur mesure pour satisfaire d'abord le narcissisme de certains chercheurs. Selon Hirsch, son indice favoriserait « une évaluation plus démocratique de la recherche »³. Tout porte plutôt à croire au contraire que cette « démocratie » non armée de connaissances sur les conditions de validité d'un indicateur se transformera rapidement en populisme évaluatif. Enfin, notons que la phrase « mon indice h est de X », que l'on voit fréquemment dans les demandes de subvention et les CV des chercheurs, n'a strictement aucune signification ! En effet, comme sa valeur dépend de la base de données utilisée, il faudrait toujours dire : « Mon indice h est de X dans la base de données Y ». De plus, même sa valeur dans la base Scopus ou WoS, varie selon que votre institution est abonnée pour la période 1990-2014 ou 1980-2014, car le calcul est fait sur les données auxquelles vous êtes abonnés ! Or, la grande majorité des chercheurs qui utilisent cet indice ignorent totalement ces détails techniques pourtant fondamentaux.

QUANTIFIÉ POUR CONTRÔLER

Les discussions animées entourant l'utilisation d'indicateurs bibliométriques dans l'évaluation des chercheurs laissent le plus souvent dans l'ombre un aspect pourtant fondamental de l'évaluation, à savoir le rôle de l'expertise des chercheurs dans le processus d'évaluation. La volonté de mieux contrôler le système très ancien d'évaluation par les pairs (*peer review*), qui repose sur une connaissance de première main du domaine de recherche du chercheur évalué, fait lentement place à l'idée d'évaluation par des experts (*expert review*), lesquels sont souvent externes au domaine de recherche considéré. L'évaluation quantitative normalisée facilite ce déplacement en fournissant des données soi-disant objectives qui peuvent alors être utilisées par n'importe qui. On est ainsi face à un paradoxe. L'évaluation d'un chercheur exige la constitution d'un comité de pairs qui connaissent bien le domaine. Ces experts savent déjà, par définition, quelles sont les bonnes revues dans leur domaine et n'ont pas besoin d'une liste préétablie par on ne sait quel groupe d'experts les classant en A, B et

3 Voir : <http://pubs.acs.org/cen/science/86/8621sci1.html>.104.

C. Par contre, ces classements permettent à des personnes ignorant tout d'un domaine de prétendre quand même porter un jugement autorisé. Mais alors ils ne devraient justement pas faire partie d'un comité d'évaluation ! La multiplication d'indicateurs mal construits sert donc en fait un processus de contournement de l'évaluation par les pairs, évaluation qui doit prendre en compte des indices de productivité, mais qui doit les interpréter dans le contexte spécifique de l'évaluation⁴.

ÉVALUER SANS CLASSER

On confond souvent évaluation et classement. En réalité, ces deux opérations sont très différentes. Si la seconde suppose la première, la première n'entraîne pas nécessairement la seconde. Contrairement à l'évaluation, qui vise à faire le point à un moment donné sur l'état du travail accompli par un individu ou une organisation et qui s'adresse donc d'abord à la personne ou à l'organisation, le classement est lié à une publication, une publicité. Or, le fait de rendre public n'est pas anodin et ne peut qu'engendrer des effets pervers. Ainsi, annoncer que la revue X est classée C et que la revue Y est classée A produira nécessairement des effets pervers. En effet, de nombreux articles sont publiés dans la revue X et certains sont originaux ou hors du commun et plus importants que ceux parus dans la revue Y, ce que savent reconnaître les chercheurs bien au fait de ce qui se passe dans leur domaine de recherche. Les échanges informels à l'intérieur de la communauté et les différents types d'interactions permettent une évaluation flexible et non stigmatisante. La hiérarchie implicite des revues dans un champ donné n'équivaut donc nullement à un classement officiel.

Par ailleurs, qui a la légitimité de décréter un tel classement ? Il y a une lutte interne au champ scientifique, dont il serait naïf de ne pas tenir compte. La composition des divers comités chargés de classer les revues dans tous les domaines des sciences humaines et sociales suggère que ce ne sont pas toujours les chercheurs les plus reconnus qui, on ne sait trop comment, font partie de ces comités cooptés. Il semble bien qu'on a ici affaire à ce que Pierre Bourdieu (1996, p. 12) a osé appeler la « loi de Jdanov », selon laquelle le dominé dans un champ est plus enclin à participer à une opération de pouvoir qui affectera les dominants. Faire partie de ceux qui déclarent officiellement que telle revue est classée A et telle autre C peut être enivrant...

PSEUDO-INTERNATIONALISATION ET DÉCLIN DES RECHERCHES LOCALES

Un aspect peu discuté de l'importance accordée aux facteurs d'impact et au classement des revues est qu'elle détourne indirectement de l'étude de sujets locaux, marginaux ou peu à la mode. Cela est particulièrement dangereux dans les sciences humaines et sociales, dont les objets sont par nature plus locaux que ceux des sciences de la nature. Il va de soi que certains

4 Pour une évaluation critique de l'usage des indicateurs bibliométriques, voir le rapport du Conseil des académies canadiennes (2012).

sujets sont moins « exportables ». Les revues les plus citées étant anglo-saxonnes (et non pas toujours « internationales ») pour des raisons purement démographiques, les chances d'y accéder dépendent de l'intérêt que ces revues portent aux objets étudiés. Un chercheur qui veut publier dans les revues les plus visibles a intérêt à étudier l'économie des États-Unis plutôt que les spécificités de la Banque du Canada ou l'économie régionale du Québec, sujet de peu d'intérêt pour une revue américaine. Le sociologue dont l'objet est « international », donc délocalisé, ou qui fait de la théorie, a plus de chances d'exporter ses articles que celui qui propose l'étude empirique d'un aspect précis de sa propre société. Une comparaison abstraite entre Habermas et Luhmann, par exemple, circule facilement sur le marché international car elle n'a aucun contenu empirique local qui risque de diminuer son attrait pour une revue anglo-saxonne. Mais, si on souhaite étudier les causes de la déviance dans certains quartiers de Montréal, on risque d'avoir plus de problèmes à « internationaliser » les résultats, même s'ils sont importants localement. Or est-ce vraiment moins important de se pencher sur cet objet que d'étudier les variations du New York Stock Exchange ? Il y a donc un danger réel que les objets locaux mais socialement importants soient dévalorisés et donc, à terme, négligés si les indicateurs de citations sont utilisés mécaniquement sans que l'on tienne compte de l'indexicalité, comme disent les ethnométhodologues, des objets de recherche en sciences humaines et sociales⁵. Les chercheurs vont alors abandonner certains objets de peur que ça ne puisse les amener à publier dans les revues classées A. Assez curieusement d'ailleurs, ces revues sont le plus souvent étrangères et en langue anglaise, les revues nationales tendant à être classées B, et même C. Il est difficile de ne pas déceler dans ces classements une forme de colonialisme intellectuel. En effet, si l'on se fie uniquement au total des citations, il est à peu près inévitable que les revues anglo-saxonnes soient mieux classées, car elles ont généralement un volume supérieur de citations par un simple effet démographique, le nombre potentiel de lecteurs dans une spécialité donnée étant plus élevé en anglais qu'en français, par exemple (Gingras et Mosbah-Natanson, 2010). Par contre, ce nombre ne dit rien sur l'origine internationale ou nationale de ces citations. Cela est particulièrement sensible en sciences humaines et sociales, pour les raisons que nous venons d'indiquer.

CONCLUSION : JUGER PLUTÔT QUE COMPTER

On entend souvent dire que ces classements sont inévitables et qu'il faut « vivre avec ». Cela est tout simplement faux. La résistance des chercheurs est tout à fait capable de bloquer de tels projets malavisés. En Australie, notamment, la vive réaction des chercheurs au classement des revues a réussi à faire plier le gouvernement, qui a abandonné l'usage de ces classements pour l'évaluation de la recherche⁶. En France aussi, les actions de nombreux chercheurs ont forcé les organismes qui faisaient la promotion de tels classements à passer à une simple liste de revues définissant un « périmètre de scientificité » (Saada, 2010 ; Pontille et Torny, 2012). Cet exemple montre bien que l'on peut définir un ensemble de revues considérées comme scientifiques dans un domaine sans tomber dans le panneau des classements qui en fait servent justement à mécaniser les décisions et à contourner l'avis des pairs.

5 Cet effet est déjà visible chez les économistes canadiens (voir Simpson et Herbert Emery, 2012).

6 Voir : <http://theconversation.com/journal-rankings-ditched-the-experts-respond-1598> ; voir aussi Vanclay (2011).

En somme, le monde de la recherche n'a pas à céder devant des exigences qui n'ont rien de scientifiques et appartiennent à des logiques qui lui sont étrangères. D'autant plus que ce sont en fait les revues francophones et les objets de recherche locaux, mais très importants pour la société, qui sortiront perdants de ces dérives de l'évaluation.

RÉFÉRENCES

- ABÉLARD (2003). *Universitas calamitatum. Le Livre noir des réformes universitaires*. Broissieux, Éditions du Croquant.
- ARCHAMBAULT, É. et LARIVIÈRE, V. (2009). History of journal impact factor : contingencies and consequences. *Scientometrics*, 79(3), 639-653.
- BHATTACHARJEE, Y. (2001). Saudi universities offer cash in exchange for academic prestige. *Science*, 334(6061), 1344-1345.
- BOURDIEU, P. (1996). *Sur la télévision*. Paris, Raisons d'Agir.
- BRISSET, C.-A. (dir.) (2009). *L'Université et la recherche en colère. Un mouvement social inédit*. Broissieux, Éditions du Croquant.
- CHARLE, C. et SOULIÉ, C. (dir.) (2007). *Les Ravages de la « modernisation » universitaire en Europe*. Paris, Syllepse.
- CONSEIL DES ACADÉMIES CANADIENNES (2012). *Éclairer les choix en matière de recherche : indicateurs et décisions*. Ottawa, Conseil des académies canadiennes.
- GALLOIS, N. (2013). Les conséquences des nouveaux critères d'évaluation des chercheurs en science économique. *L'Économie politique*, 59(3), 98-112.
- GINGRAS, Y. (2008). La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs. Montréal, Note de recherche du Centre interuniversitaire de recherche sur la science et la technologie.
- GINGRAS, Y. (2014, mise à jour 18 juillet). *How to boost your university up the rankings*. Récupéré le 21 avril 2015 du site : <http://www.universityworldnews.com/article.php?story=20140715142345754>.
- GINGRAS, Y. et MOSBAH-NATANSON, S. (2010). La question de la traduction en sciences sociales : les revues françaises entre visibilité internationale et ancrage national. *Archives européennes de sociologie*, 51(2), 305-321.
- GLÄNZEL, W. et MOED, H.F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, 53(2), 171-193.
- HIRSCH, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.
- PONTILLE, D. et TORNÿ, D. (2012). Rendre publique l'évaluation des SHS : les controverses sur les listes de revues de l'AERES. *Quaderni*, 77(1), 11-24.
- SAADA, A. (2010). L'évaluation et le classement des revues de sciences humaines par l'Agence de l'évaluation de la recherche et de l'enseignement supérieur (AERES). *Connexions*, 1(93), 199-204.
- SCHULTHEIS, F., ESCODA, M.R. et COUSIN, P.-F. (dir.) (2008). *Le Cauchemar de Humboldt. Les réformes de l'enseignement supérieur européen*. Paris, Raisons d'Agir.

SEGLÉN, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 498-502.

SIMPSON, W. et HERBERT EMERY, J.C. (2012). Canadian economics in decline : implications for Canada's economics journals. *Canadian Public Policy*, 38(4), 445-470.

VANCLAY, J.K. (2011). An evaluation of the Australian Research Council's journal ranking. *Journal of Infometrics*, 5(2), 265-274.

VAN LEEUWEN, T.N. (2008). Testing the validity of the Hirsch-index for research assessment purposes. *Research Evaluation*, 17(2), 157-160.

WALTMAN, L. et VAN ECK, N.J. (2011). *The inconsistency of the h-index*. Récupéré le 21 avril 2015 du site : <http://arxiv.org/abs/1108.3901v1>.