

La normalisation de la recherche scientifique

limite-elle la controverse et l'innovation scientifique ?



Boris Beaudé - Chôros, Ecole Polytechnique Fédérale de Lausanne

Géopoint 2014 - Université d'Avignon - jeudi 12 juin 2014

La normalisation de la recherche scientifique

limite-elle la controverse et l'innovation scientifique ?

Débat

Mesurer la qualité: vaste problème

En signant la Déclaration de San Francisco sur l'évaluation de la recherche ou Declaration on Research Assessment (DORA), le Fonds national suisse reconnaît que l'analyse quantitative est un indice modérément adéquat pour évaluer l'impact d'un projet de recherche. Que nous dit la bibliométrie sur la qualité de la science?



Valérie Chételat (photo montage)



Je suis d'accord avec la DORA sur un point: appliquée sans le moindre esprit critique, une bibliométrie présente d'importantes faiblesses si elle repose uniquement sur des facteurs d'impact («impact factors»). De nombreuses mesures proposées par la déclaration sont sensées. Mais avec son refus catégorique de toute bibliométrie («eliminate the use of journal-based metrics»), la DORA manque sa cible. Et n'offre, pour ainsi dire, aucune alternative. Dans sa référence 11 consacrée à cette question, six des sept hyperliens n'existent plus.

Les bibliométries - de revues spécialisées ou de chercheurs individuels - partent de l'idée que l'attention dont jouit une contribution (c'est-à-dire la fréquence à laquelle elle est citée) constitue un indice de sa qualité. Or, je préfère une bibliométrie à une évaluation purement subjective par des collègues du même domaine de recherche (pairs) où le risque d'influence des coteries est important. Une évaluation par les pairs est plus transparente, et donc plus raisonnable, si elle est étayée par une bibliométrie quantitative. Ces derniers ont alors toujours la possibilité de la corriger, voire de s'en écarter, s'ils ont des raisons valables de le faire.

Il existe différents types de bibliométries. Elles ont toutes leurs faiblesses. Mais combinées, elles peuvent être exploitées de manière utile pour évaluer la recherche:

1. Les facteurs d'impact des revues spécialisées proviennent de banques de données de firmes privées; ils ne sont sou-

vent guère comparables d'une discipline à l'autre et peuvent être manipulés par les éditeurs moyennant des autocitations. Ils ne conviennent donc que pour une classification grossière des revues, et leur rang («ranking») au sein de la discipline compte davantage que le chiffre

«Je préfère une bibliométrie à une évaluation purement subjective par des pairs.»

Chris Boesch

absolu. Mais le ranking permet une distinction qualitative entre les auteurs qui publient régulièrement dans des revues du tiers supérieur, et ceux qui publient presque exclusivement dans celles de la moitié inférieure.

2. Les analyses individuelles de citations, comme l'indice h, doivent être considérées dans le contexte de la discipline. En raison des différences de cultures, mais aussi parce que certaines disciplines documentent différemment les acquis de connaissance. En sciences de l'ingénieur, par exemple, les brevets et les contributions de plusieurs pages à des congrès comptent comme une publication. Il faut aussi s'intéresser à l'évolution dans le temps du nombre de citations, pour pouvoir prendre en compte la situation

professionnelle du chercheur, comme une réduction de son activité de publication lors de la mise en place d'un groupe de recherche.

3. Enfin, l'analyse des paternités des publications dépend de la discipline. L'évaluation du premier et du dernier auteur a tout son sens en biologie et en médecine, mais n'est d'aucun secours en mathématiques ou en physique des particules. Par ailleurs, si un article du CERN repose sur le travail de centaines d'auteurs, quelle est la contribution d'un médecin parmi les dix auteurs d'un «case report» (étude du cas d'un seul patient)? Les différentes disciplines doivent déterminer elles-mêmes les formes optimales de publication, ainsi que leur pondération. Mais au-delà des diversités, une bibliométrie spécifique à chaque discipline serait souhaitable.

Il est décisif qu'une analyse des publications s'appuie sur une base aussi large que possible et qu'elle implique les points critiques. Malgré toutes les faiblesses de la bibliométrie, je suis convaincu que la qualité de la recherche mérite d'être quantifiée. Les pairs peuvent ensuite se former une opinion sur la base de ces chiffres.

Chris Boesch dirige le Département de spectroscopie et de méthodologie IRM à l'Université de Berne. Il est aussi membre de la division biologie et médecine du Conseil national de la recherche du FNS.

Dans les divers domaines scientifiques couverts par le Fonds national suisse (FNS), une idée prime sur toutes les autres: la qualité des recherches à soutenir. Cette dernière tient à la nature des projets soumis, mais aussi au profil des chercheurs. En ce qui concerne les projets, l'originalité, la pertinence et la faisabilité constituent les principaux critères d'évaluation. Pour ce qui relève du profil, dans les sciences naturelles, la médecine et une partie des sciences sociales, les évaluateurs disposent d'instruments de mesure par lesquels ils estiment le «poids» d'une chercheuse ou d'un chercheur. Ce poids peut correspondre à une fréquence de citations d'un auteur, ou à une mesure de l'impact d'une chercheuse dans sa discipline. Il présente l'avantage de reposer sur des critères identiques pour l'ensemble des personnes.

Mais tout n'est pas si simple! Une application extrême de cette démarche aurait pour conséquence que les fonds de la recherche seraient alloués à des savants en fonction de quelques indications chiffrées. Or, une telle approche présente deux désavantages: en définissant les critères permettant de calculer de tels indices, il est facile d'influencer le processus. Et résumer le travail d'un chercheur à une valeur numérique peut paraître certes commode, mais c'est également une démarche excessivement simplificatrice. Ainsi, une longue liste d'articles n'est pas toujours synonyme de qualité, et le fait d'être cité fréquemment à un moment donné ne signifie pas néces-

sairement que l'on aura un impact durable sur la recherche. C'est pourquoi de nombreuses voix s'élèvent pour que l'évaluation des chercheurs et de leurs projets se fasse sur une base plus large. La Déclaration de San Francisco, qui émane des milieux de la biologie, fait écho à une revendication dans

«Nos chercheuses et chercheurs ne sont pas seulement l'expression d'une valeur numérique.»

Paul Schubert

laquelle se reconnaissent aussi de nombreux chercheurs des sciences humaines.

Il ne s'agit pas d'interdire l'usage des index de référence numériques, mais d'encourager en parallèle l'utilisation d'autres approches reposant notamment sur la lecture critique des publications. De plus, pour des évaluateurs connaissant bien leur discipline, le poids d'un chercheur peut être mesuré par les résultats effectifs qu'il a obtenus, par l'influence de ses recherches dans le monde de l'éducation ou de la politique, ou par sa capacité à faire avancer un nouveau concept dans la perception des autres chercheurs.

Dans la pratique du FNS, c'est bien une approche mixte qui est privilégiée: lorsque les indices de mesure existent dans un domaine, ils sont complétés, à des degrés

variables selon les disciplines, par un examen plus fin couvrant aussi les aspects qualitatifs d'un dossier. Nos chercheuses et chercheurs ne sont pas seulement l'expression d'une valeur numérique. Ils produisent également des résultats dont les effets concrets peuvent être appréciés dans le paysage scientifique ainsi que dans la société. Le travail de sélection réalisé par le Conseil national de la recherche doit donc reposer à la fois sur l'utilisation d'indicateurs et sur une capacité à repérer dans de nouveaux projets ceux qui porteront la science de demain.

Paul Schubert est professeur de grec à l'Université de Genève et préside la division des sciences humaines et sociales au sein du Conseil national de la recherche du FNS.

La normalisation de la recherche scientifique

limite-elle la controverse et l'innovation scientifique ?

Débat

Mesurer la qualité: vaste problème

En signant la Déclaration de San Francisco sur l'évaluation de la recherche ou Declaration on Research Assessment (DORA), le Fonds national suisse reconnaît que l'analyse quantitative est un indice modérément adéquat pour évaluer l'impact d'un projet de recherche. Que nous dit la bibliométrie sur la qualité de la science?



Valérie Chételat (photo-montage)



«Je préfère une bibliométrie à une évaluation purement subjective par des pairs.»

Chris Boesch

Chris Boesch dirige le Département de spectroscopie et de méthodologie IRM à l'Université de Berne. Il est aussi membre de la division biologie et médecine du Conseil national de la recherche du FNS.

chercheur, comme l'activité de publication au lieu d'un groupe

critères des publications. L'évaluation d'un auteur a lieu et en médecine, dans les mathématiques et les particules. Par exemple, au CERN repose sur le nombre d'auteurs, quelle est la médecine parmi les autres? Les différents critères de publication. Mais la bibliométrie dans une discipline serait

analyse des publications aussi large que possible. Les points faibles de la bibliométrie est que la qualité n'est pas quantifiée. Elle se forme en chiffres.

Département de spectroscopie et de méthodologie IRM à l'Université de Berne, la division biologie et médecine du Conseil national de la recherche

Dans les divers domaines couverts par le Fonds national suisse (FNS), un certain nombre de recherches à soutenir. Cette nature des projets soumet le profil des chercheurs. En fait, l'originalité, la faisabilité constituent les critères d'évaluation. Pour ce profil, dans les sciences naturelles et une partie des sciences sociales, les évaluateurs disposent d'une mesure par laquelle ils estiment d'une chercheuse ou d'un chercheur. Le poids peut correspondre à une citation d'un auteur, à l'impact d'une recherche. Il présente l'avantage de critères identiques pour toutes les personnes.

Mais tout n'est pas si simple. L'application extrême de cette méthode pour conséquence que la recherche serait allouée en fonction de quelques chiffres. Or, une telle approche présente des avantages: en définissant un critère permettant de calculer le poids est facile d'influencer le premier travail d'un chercheur. Le numérique peut paraître objectif mais c'est également une simplification excessive. Une liste d'articles n'est pas toujours de qualité, et le fait d'être cité à un moment donné ne signifie pas que l'article est de qualité.

«Nos chercheuses et chercheurs ne sont pas seulement l'expression d'une valeur numérique.»

Paul Schubert

Paul Schubert est professeur de grec à l'Université de Genève et préside la division des sciences humaines et sociales au sein du Conseil national de la recherche du FNS.

L'évaluation / *Entre innovation et conservatisme*

I - De l'intérêt de l'évaluation

L'évaluation / *Entre innovation et conservatisme*

1.

Pourquoi l'évaluation ?

1. *Problématique de marché*

2. *Respect des normes de la pratique scientifique*

L'évaluation / *Entre innovation et conservatisme*

2.

Qu'est-ce que la pratique scientifique ?

1. *La réfutabilité*

2. *Les quatre critères de Merton*

2.1. *Universalisme*

2.2. *Communalisme*

2.3. *Désintéressement*

2.4. *Scepticisme organisé*

L'évaluation / *Entre innovation et conservatisme*

2.

Qu'est-ce que la pratique scientifique ?

2. La critique des critères de Merton

2.1. Anglo-américanisme-contextualisation

2.2. Appropriation

2.3. Intérêt personnel

2.4. Conformisme stratégique

3. « Science normale » et « sciences extraordinaire »

4. « Reconnaissance contre connaissance »

5. Sciences expérimentales, sciences sociales

L'évaluation / *Entre innovation et conservatisme*

3.

Comment la recherche est-elle évaluée ?

1. *Les traces de la recherche*

1.1. *Les communications scientifiques*

1.2. *Les publications scientifiques*

2. *Universalisme contre particularismes*

2.1. *International plutôt que local*

2.2. *Peer review plutôt que mandarinate*

2.3. *Anglais comme médiation*

II - Des limites de l'évaluation

L'évaluation / *Entre innovation et conservatisme*

1.

Les limites de l'universalisme

1. L'anglais est inégalement pratiqué

1.1. Comités des revues les plus prestigieuses.

1.2. Auteurs des revues les plus prestigieuses.

2. Spécificité des sciences des sujets

2.1. La traduction du subjectif et du symbolique.

2.2. La problématique des universaux.

3. Les modes

3.1.1. Les sujets à la mode et la publication.

3.1.2. Les sujets à la mode et la citation.

L'évaluation / *Entre innovation et conservatisme*

2.

Les limites de la quantité

1. *L'évaluation des traces est quantitative*

1.1. *Elle compte le nombre de publications*

1.2. *Elle compte le nombre de citations*

2. *La «qualité» des revues est évaluée sur la quantité*

2.1. *Elle compte le nombre de publications*

2.2. *Elle compte le nombre de citations*

L'évaluation / *Entre innovation et conservatisme*

3.

Qualité et Peer Review

1. *La qualité dépend donc du Peer Review*

1.1. *Sélection de la norme*

1.2. *Sélection de l'innovation*

2. *La faiblesse dépend donc du Peer Review*

2.1. *Sélection de la norme*

2.2. *Sélection de l'innovation*

4.

Les limites du Peer Review

1. *L'anonymat*
2. *La reconnaissance*
3. *La concurrence*
4. *Le conservatisme*
5. *La représentativité*
6. *La difficulté*
7. *La fiabilité (2 à 3 personnes)*

5.

Les limites de la bibliométrie

1. *La quantité nuit à la visibilité*
2. *La qualité n'est pas valorisée*
3. *Les modes sont accentuées*
4. *La fiabilité est limitée*
5. *L'innovation n'est pas valorisée*
6. *Le conservatisme est favorisé*

5.

Les limites de la bibliométrie

7. *L'accès aux publications coûte cher (publication ou lecture)*

8. *Cloisonnement disciplinaire*

9. *L'enseignement et les taches institutionnelles sont dévalorisés et délaissés (avenir de la recherche)*

10. *La recherche est délaissée (pas ou fausses expériences)*

11. *ET l'évaluation est délaissée...*

6.

Les limites de la bibliométrie propres aux SHS

- 1. Les temporalités ne sont pas adaptées*
- 2. Les livres ne sont pas valorisés*
- 3. Les modes sont plus localisées*
- 4. La langue pose des problèmes*
- 5. La double subjectivité est sensible*
- 6. La grande pluralité des paradigmes est entretenue*

III - Quelles pistes pour l'évaluation

L'évaluation / *Entre innovation et conservatisme*

1.

Valoriser la pratique scientifique dans son ensemble

- 1. Ne pas accorder trop d'importance aux indicateurs.*
- 2. S'intéresser vraiment aux personnes, aux projets et à l'innovation (lire les contenus).*
- 3. Valoriser l'enseignement et la participation aux institutions (avenir de la recherche).*
- 4. Valoriser la participation à l'évaluation.*
- 5. Valoriser la direction de thèses et les thèses.*
- 6. Valoriser la participation à des séminaires.*
- 7. Considérer des périodes plus longues.*
- 8. Accepter les échecs.*

L'évaluation / *Entre innovation et conservatisme*

2.

Répenser la publication scientifique

1. *Publier moins*
2. *Publier mieux*
3. *Méthodes plus transparentes (disponibilité des sources)*
4. *Evaluation transparente*
5. *Evaluation ouverte*
6. *Ne pas juger à priori de sa valeur*
7. *Responsabiliser les auteurs et les critiques*

L'évaluation / *Entre innovation et conservatisme*

2.

Répenser la publication scientifique

8. Penser les publications comme des coproductions

9. Ne pas figer un texte

10. S'inspirer plus du développement open source (GitHub, Wikipédia...)

11. S'inspirer des autres disciplines (PLoS, ArXiv, PeerJ, BMJ, ACP...)

L'évaluation / *Entre innovation et conservatisme*

3.

Quelles options

1. *Prépublication*
2. *Open Access*
3. *Open Peer Review*
4. *Évaluations non anonymes*
5. *Open comments*
6. *Commentaires non anonymes*
7. *Publication des évaluations*
8. *Publication des commentaires*
9. *Historique des textes*

L'évaluation / *Entre innovation et conservatisme*

4.

Quels enjeux

- 1. Lire des textes plus riches*
- 2. Ecrire des textes plus riches*
- 3. Avoir le temps du recul*
- 4. Avoir le temps de la recherche*
- 5. Evaluer le contenu des recherches*
- 6. Pouvoir ne pas être d'accord*
- 7. Pouvoir prendre le risque d'innover*

L'évaluation / *Entre innovation et conservatisme*

« Ce qui compte ne peut pas toujours être compté, et ce qui peut être compté ne compte pas forcément ».

Albert Einstein

L'évaluation / *Entre innovation et conservatisme*

Textes

Jean-Michel Berthelot. 2008. «L'emprise du vrai». PUF.

Grégoire Chamayou, 2009, «Petits conseils aux enseignants-chercheurs qui voudront réussir leur évaluation.», Contretemps.eu.

Swiss Informatics Research Association (Collectif), 2008, «Lettre à la CRUS concernant une évaluation bibliométrique».

Levelt Committee, Noort Committee, Drenth Committee, 28 novembre 2012, «Flawed science: The fraudulent research practices of social psychologist Diederik Stapel».

Javier Gutiérrez and Pedro López-Nieva , 2001, « Are international journals of human geography really international?», *Progress in Human Geography, Vol. 25, n°1*, pp. 53–69.

Michael Nielsen, 2011, « Reinventing Discovery: The New Era of Networked Science », Princeton University Press.

Alain Desrosières, 2008, « Gouverner par les nombres », Mines ParisTech.

Vincent de Gaulejac, 2012, « La recherche malade du management », Éditions Quæ.

Editorial (2005). Not so deep impact. *Nature* 435, 1003–1004.

Sites

Research Analytics - Thomson Reuters - evidence.co.uk.

Publish or Perish - harzing.com.

Academic productivity - academicproductivity.com.

Evaluation de la recherche en SHS - evaluation.hypotheses.org.

DORA (San Francisco Declaration on Research Assessment) - <http://am.ascb.org/dora/>