

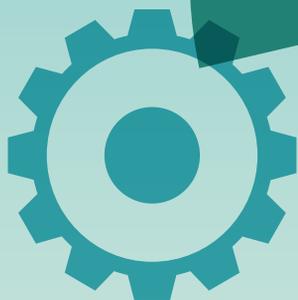
PENSER LA SCIENCE

Edwin ZACCAI • Benoît TIMMERMANS

Marek HUDON • Barbara CLERBAUX

Bruno LECLERCQ • Hugues BERSINI

L'évaluation de la recherche en question(s)





L'ÉVALUATION DE LA RECHERCHE EN QUESTION(S)

PENSER LA SCIENCE

Edwin ZACCAI · Benoît TIMMERMANS
Marek HUDON · Barbara CLERBAUX
Bruno LECLERCQ · Hugues BERSINI

L'évaluation de la recherche en question(s)



Classe des Sciences
ACADÉMIE ROYALE DE BELGIQUE

Académie royale des Sciences,
des Lettres et des Beaux-Arts de Belgique
rue Ducale, 1
1000 Bruxelles, Belgique
editions@academieroyale.be
www.academie-editions.be

Mémoire de la Classe des Sciences
Collection in-8°, IV^e série, tome VII
n° 2113

© 2016, Académie royale de Belgique

Crédits

© Les auteurs, pour leur texte respectif.

Suivi et conception graphique
Grégory Van Aelbrouck et Loredana Buscemi,
Académie royale de Belgique

ISBN 978-2-8031-0561-8
ISSN 0365-0936
Dépôt légal : 2016/0092/24

Table des matières

<i>Introduction</i>	9
EDWIN ZACCAI	
<i>Évolutions récentes de l'évaluation de la recherche</i>	13
BRUNO LECLERCQ	
<i>Quels types d'évaluation pour quelle recherche et quelle société ?</i>	23
VÉRONIQUE HALLOIN	
<i>Évaluations, audits, classements: comment faire ? qu'en faire ?</i>	33
PIERRE MARAGE	
<i>Mise en perspective historique des « évaluations » de la recherche publique</i>	51
DOMINIQUE PESTRE	
<i>Évaluer la recherche ?</i>	57
ALAIN ERALY	
<i>Du Bon usage des indicateurs</i>	67
YVES GINGRAS	
<i>Le H-index à l'index ?</i>	75
HUGUES BERSINI	
<i>Réflexion sur l'évaluation de la recherche en physique et gestion</i>	85
BARBARA CLERBAUX ET MAREK HUDON	

<i>L'impact des rankings sur l'existence et la diversité des revues scientifiques et sur les pratiques disciplinaires. Le cas de la géographie</i>	91
CHRISTIAN VANDERMOTTEN	
<i>Faire une carrière scientifique aujourd'hui</i>	101
BERNARD FUSULIER	
<i>La méthode BVLM</i>	111
L'ATELIER DES CHERCHEURS POUR UNE DÉSEXCELLENCE DES UNIVERSITÉS	
<i>Comment nous désirons être évalué-e-s ?</i>	129
ISABELLE STENGERS	
<i>Éléments pour construire une évaluation alternative</i>	141
MARYSE SALLES ET GABRIEL COLLETIS	
<i>Ateliers d'élaboration de propositions</i>	157

Introduction

L'évaluation de la recherche scientifique constitue un sujet d'une grande importance. D'abord pour les chercheurs et professeurs dans le monde académique. Ensuite pour les institutions qui les emploient et les financent. Enfin pour la société qui entretient des attentes envers les connaissances scientifiques et leurs applications.

Mais de quelle évaluation, de quels chercheurs, de quelles institutions et société parle-t-on ?

À l'examen il existe une grande diversité de chercheurs et de disciplines, ou encore de pratiques et d'objectifs de recherche. La question est alors de savoir dans quelle mesure les modes d'évaluation sont pertinents et appropriés à cet ensemble quelque peu insaisissable. Question qui se pose avec plus d'acuité dans le cadre de bouleversements induits par l'informatisation, l'internationalisation, les contraintes financières des pouvoirs publics, et plus largement par les évolutions contemporaines des sociétés.

Sujet d'une grande importance, disions-nous, et pourtant sa discussion explicite demeure souvent cantonnée à de petits cercles spécialisés, sans parler de discussions de couloirs informelles autant que récurrentes sur les ressentis de ces processus.

La structure Penser la Science – Séminaires Ilya Prigogine a été créée à l'Université libre de Bruxelles pour promouvoir échanges et débats appelés à nourrir une culture des sciences dans leur diversité. Chaque année un thème de réflexion est choisi. En 2015, une petite équipe, constituée des trois nouveaux permanents de Penser la Science, Marek Hudon, Benoit Timmermans et Edwin Zaccai, entourés de Hughes Bersini, Barbara Clerboux et Bruno Leclercq, a décidé de s'emparer de ces questions sur l'évaluation de la recherche.

Deux journées ont été organisées, le 9 mai et le 16 octobre. Elles ont réuni près de deux cents inscrits, dont une vingtaine d'orateurs belges et étrangers,

issus du monde académique ou d'institutions. Pour certains chercheurs l'évaluation de la recherche est leur sujet de... recherche, d'autres au contraire ont profité de cette occasion pour élaborer une réflexion sur cet aspect de leur métier et sur ses conséquences. Les exposés ont été enregistrés et restent disponibles sur le site de Penser la Science¹. Signalons que le FNRS a consacré un dossier à ce même thème en interviewant plusieurs orateurs de ces séminaires².

La première des journées de ce projet visait à effectuer un état des lieux, tandis que la seconde cherchait plus particulièrement à formuler des propositions alternatives. Toutefois la plupart des intervenants se prononcent à la fois sur des constats et des propositions, comme on le verra dans ce volume.

Un premier ensemble de contributions réfléchit à l'ensemble de la problématique posée, à travers des regards complémentaires. Tout d'abord Bruno Leclercq, professeur de philosophie à l'Université de Liège, se propose de poser le décor de ces évaluations. En articulant une série de références, il analyse plusieurs concepts pertinents qui se retrouveront mobilisés dans ce qui suit. Véronique Halloin, secrétaire générale du F.R.S.-FNRS, analyse successivement le contexte, les types d'évaluation, et une série de constats et de questions actuelles en concluant sur les contraintes de financement de la recherche. Pierre Marage, physicien et ancien vice-recteur à la recherche de l'ULB, détaille un certain nombre de situations d'évaluations en distinguant notamment entre des évaluations d'équipes, de projets et d'individus. Il formule des propositions dans différents domaines, et notamment quant aux biais du genre dans la carrière académique. Dominique Pestre, directeur d'études à l'École des Hautes Études en Sciences Sociales (Paris), replace l'évaluation de la recherche dans le cadre plus large de la gouvernementalité libérale des sociétés, en évoquant la grande vogue des classements dans différents domaines et en suggérant des éléments utiles à une pensée critique à ce sujet. La contribution d'Alain Eraly, professeur à l'ULB de sociologie et de gestion et membre de l'Académie royale de Belgique, prolonge cette réflexion. Il aborde dans cette perspective nombre d'éléments qui reviendront dans d'autres contributions, comme les biais des indicateurs chiffrés ou le ressenti psychologique des chercheurs.

Les quatre contributions qui suivent entrent davantage dans la technicité d'outils d'évaluation, et particulièrement d'indicateurs chiffrés ainsi que de

¹ <http://penserlascience.ulb.ac.be/>

² « L'évaluation de la recherche en question(s) », *FNRS News*, n° 102, septembre 2015, <http://www.fnrs.be/docs/Lettre/lettre102.pdf#page=10>

leurs conséquences. Yves Gingras, professeur en histoire et sociopolitique des sciences à l'Université du Québec et auteur de plusieurs travaux de portée internationale sur l'évaluation de la recherche, détaille, exemples à l'appui, un répertoire de mésusages d'indicateurs, et conclut sur une série de propositions à cet égard. Hugues Bersini, directeur de l'IRIDIA (Institut de recherches interdisciplinaires et de développement en intelligence artificielle) à l'ULB et membre de l'Académie, se penche plus particulièrement sur le H-Index. Il se livre à une défense raisonnée de son emploi, en énonçant des conclusions en forme de recommandations d'usage. L'un des biais du H-Index est le fait que ses valeurs types diffèrent sensiblement d'une discipline à l'autre. La contribution suivante est issue justement de deux professeurs de l'ULB issus de disciplines différentes, Barbara Clerbaux, professeur au département de physique, et Marek Hudon, professeur à la Solvay Brussels School. Parmi les difficultés pointées dans ce texte, remarquons notamment la faible comptabilisation dans les systèmes utilisés des transferts de savoirs entre université et sociétés, avec par exemple des cas de pondération... nulle pour des livres influents. Enfin Christian Vandermotten, géographe à l'ULB et membre de l'Académie, étudie de manière quantifiée certaines conséquences pour les revues scientifiques des stratégies de publication des chercheurs, eux-mêmes influencés par les facteurs d'impacts normés.

Les deux contributions suivantes ont en commun de se placer plus particulièrement du point de vue des chercheurs soumis aux modes d'évaluation actuels. Bernard Fusulier, professeur de sociologie à l'Université catholique de Louvain et maître de recherches du FNRS, envisage la condition des chercheurs et chercheuses dans leurs différentes dimensions de vécu, y compris familiales. Son analyse décrit les modalités du régime « comptable-productiviste » de l'évaluation de l'excellence, et se clôt sur l'évocation d'une « alter-excellence » qui fait parfaitement le lien avec le chapitre suivant. Celui-ci, produit par un collectif de chercheurs, le LAC (L'Atelier des Chercheurs pour une désexcellence des universités), expose avec provocation et humour la méthode BVLM, pour « Beaucoup, Vite, Loin, Mal ». Dans le contexte d'évolutions plus larges décrites aussi par d'autres auteurs de ce livre, ce chapitre va plus loin dans l'évocation de malaises ressentis à travers la mise en scène de cas vécus.

Enfin, les trois contributions qui clôturent ce volume s'efforcent plus particulièrement de formuler des propositions de changement, même si répétons-le, on trouve de celles-ci dans pratiquement tous les chapitres. Isabelle

Stengers, professeur de philosophie à l'ULB, décrit les transformations d'un modèle plus ancien de production scientifique vers celui d'une économie de la connaissance. Elle plaide pour une évaluation qui se soucie du devenir des productions scientifiques dans la société en envisageant les nombreuses questions que cela suppose. Gabriel Colletis et Maryse Salles, qui enseignent à l'Université de Toulouse, proposent des alternatives à l'évaluation actuelle de la recherche en se basant notamment sur un rapport français ayant étudié les normes et les pratiques en vigueur en économie. Plutôt que de concentrer les critiques et propositions sur des éléments concrets comme les indicateurs, ils plaident pour un réexamen des principes mêmes guidant l'évaluation. Le dernier texte présenté reprend la synthèse de quatre ateliers organisés lors du séminaire à l'ULB, où les participants étaient invités à élaborer des propositions portant sur quatre thèmes : les indicateurs, les dialogues entre évaluateurs et évalués, les évalués dits en situation défavorable (moins bien placés par rapport aux critères les mieux valorisés), l'évaluation de l'interdisciplinarité³.

La richesse et le foisonnement de ces nombreuses analyses confirment à nos yeux l'importance actuelle des questions abordées dans l'ensemble de ce livre et le désir de pouvoir dialoguer à leur sujet. Relevons, dans ce contexte, la présence tout au long des journées de personnalités telles que des doyens de facultés, responsables de recherche et responsables d'institutions. À cet égard, outre les contributeurs à ce volume, nous voudrions aussi remercier pour leur participation à ces journées Jean-Pierre Bourguignon, président du Conseil Européen de la Recherche (ERC) et Serge Schiffmann, vice-recteur à la recherche et au développement à l'ULB, dont on trouvera d'ailleurs les enregistrements des interventions sur le site de Penser la Science.

Nous tenons aussi à remercier deux doctorants qui se sont impliqués avec générosité dans l'organisation de ces journées, Marine Lugen et Quentin Hiernaux, ainsi que Frédéric Schmitz pour ses talents de webmaster. Enfin, nous sommes heureux d'avoir pu voir la présente publication parrainée par trois académiciens, contributeurs de cet ouvrage, Hugues Bersini (également membre du comité scientifique du projet), Alain Eraly et Christian Vandermotten.

Edwin Zaccai, professeur à l'ULB
et directeur du Centre d'Études
du Développement Durable (CEDD)

³ Signalons d'ailleurs que l'édition 2016 de Penser la Science sera consacrée à ce thème.

Évolutions récentes de l'évaluation de la recherche

Quelques concepts pertinents

BRUNO LECLERCQ

Professeur de philosophie à l'Université de Liège

Tant sur le plan scientifique que sur le plan académique, les activités (des) universitaires font désormais l'objet d'évaluations nombreuses et à divers niveaux, dont le trait le plus saillant est l'articulation autour d'*indicateurs de performance*, qui entendent d'une manière ou d'une autre « quantifier la qualité » de manière à « objectiver » l'évaluation et permettre des comparaisons (*benchmarking*) voire des classements (*ranking*).

1. PRÉGNANCE DES INDICATEURS DE PERFORMANCE

Sur le plan des activités *scientifiques*, sont désormais extrêmement prégnants — pour l'évaluation des candidats à des postes universitaires, mandats de recherches, bourses, etc., mais aussi pour l'évaluation des équipes qui prétendent au financement de certains projets de recherches — des indicateurs tels que le nombre de publications (éventuellement pondéré par le *ranking* des revues qui ont accueilli ces publications), les facteurs d'impact (qui sont fonction du nombre de citations dont ces publications ont fait l'objet dans d'autres articles), mais aussi le nombre de conférences prononcées à l'étranger, le nombre de mois passés dans des séjours de recherches à l'étranger, le nombre de projets de recherches réalisés en collaboration internationale (pondéré par l'importance de ces projets mesurée d'après la hauteur

des financements publics ou privés), la taille de l'équipe de recherches dans lequel le chercheur ou la chercheuse est inséré(e), son attractivité mesurée d'après le nombre de doctorants et de postdocs étrangers qu'elle accueille...

Sur le plan des activités *académiques*, les rankings internationaux des institutions d'enseignement supérieur¹ font prévaloir, outre des indicateurs de performance de recherche, des indicateurs d'efficience en matière d'enseignement : taux d'encadrement (ratio staff/étudiants), taux de satisfaction des étudiants, taux de succès des étudiants dans les études ultérieures, taux de pénétration des diplômés sur le marché de l'emploi, taux de satisfaction de leurs employeurs, degré d'internationalisation (nombre d'étudiants de programmes d'échange, nombre de nationalités représentées parmi les étudiants, nombre et vigueur des partenariats internationaux...)...

Ces indicateurs, qui sont au fondement des très médiatiques « rankings » d'universités, interviennent aussi, à des degrés divers, dans des processus d'évaluation des universités ou départements par les agences Qualité nationales ou internationales à des fins d'accréditation et/ou financement des programmes de formation², ainsi que par les universités elles-mêmes pour décider en interne de leurs offres de formation et de l'affectation de leurs moyens financiers pour soutenir ces offres³, mais aussi de plus en plus dans les processus d'évaluation individuelle des enseignants pour prolongation de leur contrat ou pour leurs promotions.

Loin d'être accessoires, ces diverses évaluations académiques et scientifiques jouent un rôle absolument déterminant dans la sélection des enseignants et des chercheurs, ainsi que dans le financement (et l'existence même) des programmes de formation et des projets de recherche, soit qu'ils orientent directement les politiques publiques, soit qu'ils pèsent sur le choix d'acteurs privés tels que les étudiants au moment de décider du lieu de leur formation,

¹ Cf. méthodologie de rankings de l'Université de Shanghai (<http://www.shanghairanking.com/fr/aboutarwu.html>), Quacquarelli Symonds (<http://www.topuniversities.com/qs-stars/qs-stars-methodology>) ou du Times Higher Education Supplement (<http://www.timeshighereducation.co.uk/world-university-rankings/2014-15/world-ranking/methodology>).

² Pour un regard critique sur les rôles et méthodologies, très variables d'un endroit du monde à l'autre, des agences Qualité dans l'enseignement supérieur, voir FALLON, C., LECLERCQ, B. (eds.), 2014. *Leurres de la qualité dans l'enseignement supérieur ? Variations internationales sur un thème ambigu*, Bruxelles, Academia.

³ La chose était notamment patente dans le guide d'auto-évaluation qui guidait les premières démarches qualité pilotées par le SMAQ à l'Université de Liège.

les employeurs lorsqu'ils recrutent des diplômés ou les entreprises quand il s'agit de conclure des partenariats de recherche.

2. DES EXIGENCES D'OBJECTIVATION, DE REDDITION DE COMPTES ET DE TRANSPARENCE GESTIONNAIRE

À travers ces indicateurs, ce qui se joue, c'est évidemment une démarche de *quantification* de la qualité, laquelle vise à satisfaire des exigences d'*objectivation* et de *transparence* : pour pouvoir opérer rationnellement leurs choix quant aux candidats à embaucher, quant aux programmes de formation ou aux projets de recherches à financer, quant aux institutions dans lesquelles effectuer sa scolarité, etc., les acteurs du « marché » de l'enseignement supérieur et de la recherche doivent disposer d'informations claires sur les performances des dispositifs dans lesquels ils investissent, et en outre d'informations *standardisées*, c'est-à-dire permettant les comparaisons entre produits concurrents. Et cette logique de marché ne vaut pas seulement pour les opérateurs privés ; qu'elle doive aussi s'étendre aux opérateurs publics, c'est en effet le principal *credo* de la Nouvelle Gestion Publique (*New Public Management*) : pour tirer le meilleur bénéfice des investissements publics, les pouvoirs publics doivent eux aussi disposer d'informations claires et standardisées quant aux performances des services qu'ils financent, de manière notamment à pouvoir comparer les politiques plus ou moins efficaces, les « pratiques plus ou moins bonnes ». D'où l'exigence d'imputabilité et de reddition de comptes (*accountability*) pesant sur les prestataires de services publics, « reddition de comptes » devant désormais clairement être entendu dans le sens de comptes chiffrés quant aux performances réalisées et éventuellement quant à leur plus ou moins grande adéquation à des objectifs chiffrés préalablement fixés⁴.

Dans le domaine public, cette exigence de reddition de comptes et de transparence est d'autant plus importante que les décideurs doivent eux-mêmes de plus en plus rendre des comptes et motiver leurs décisions : face

⁴ Voir à cet égard la section « Le savoir et l'évaluation de sa qualité » (et les références bibliographiques mentionnées là) de LECLERCO, B., « La valeur du savoir », in *Contrôler la qualité dans l'enseignement supérieur*, Bruxelles, Academia, 2012, 155-160. Que cette imputabilité (*accountability*) soit tout autre chose que la responsabilisation à l'égard de certains idéaux partiellement indéterminés, c'est ce qu'il importe de remarquer et que nous avons essayé de souligner dans la section « L'utilité indéterminée du savoir » du même texte.

à l'opinion publique, mais aussi face à des instances de contrôle nationales (Cour des comptes...) ou internationales (Communauté européenne, OCDE...), les ministres eux-mêmes doivent expliquer le choix de leurs politiques ; en outre, sous peine de recours au Conseil d'État, telle université publique doit pouvoir justifier qu'elle a engagé cet enseignant ou ce chercheur parmi l'ensemble des candidats, tel fonds de recherches public doit pouvoir justifier qu'il a retenu tel projet plutôt que tel ou tel autre. La transparence s'impose dans les deux sens : si les décideurs exigent de voir clair dans les performances des acteurs qui sollicitent leur soutien financier, ces derniers exigent en retour de voir clair dans les principes qui président aux décisions d'affectation des moyens. Tout arbitraire doit être banni ; chaque décision publique doit pouvoir faire valoir des éléments objectifs et standardisés, des indices de performances. C'est ce qu'avec Catherine Fallon nous avons appelé le paradigme de la transparence gestionnaire et de sa logique extensionnelle, lequel entend s'opposer à une certaine « opacité » de contextes intensionnels tels que celui de la donation de sens politique⁵. Les chiffres parlent d'eux-mêmes et imposent les décisions : parce qu'il est plus performant que ses concurrents, tel candidat *doit* être engagé, tel programme de formations ou tel projet de recherches *doit* être soutenu, telle politique *doit* être menée. La prise de décision proprement politique, avec ses positionnements axiologiques et ses choix de société, cède ainsi le pas à la « gestion par les faits » et en particulier « par les nombres »⁶.

⁵ FALLON, C., LECLERCQ, L., 2015. « Le miroir aux alouettes : de l'opacité politique à la transparence gestionnaire ? », à paraître dans *Carrefours de l'éducation*.

⁶ Voir sur ce sujet : BADET, F., JANY-CATRICE, F., 2010. « Les politiques de quantification », in *Revue Française de Socio-Économie*, 5, p. 9-17 ; BRUNO, I., 2008. « La recherche scientifique au crible du benchmarking. Petite histoire d'une technologie de gouvernement », in *Revue d'histoire moderne et contemporaine*, 55, 4bis, p. 28-45 ; BRUNO, I., 2009. « Le temps des 'chercheurs-entrepreneurs' : sens et pouvoir du benchmarking dans l'«espace européen de la connaissance' », in *Quaderni* [En ligne], 69, mis en ligne le 05 avril 2012 ; BUREAU, M.-C., 2010. « Du travail à l'action publique : quand les dispositifs d'évaluation prennent le pouvoir », in *Cahiers internationaux de sociologie*, 128-129, p. 161-175 ; DEBACKERE, K., GLÄNZEL, W., 2004. « Using a bibliometric approach to support research policy making : the case of the Flemish BOF-key », in *Scientometrics*, 59, 2, p. 253-276 ; DESROSIÈRES, A., 2008. *Gouverner par les nombres*, Paris, Presses des Mines ParisTech ; DIDIER, E., TASSET, C., 2013. « Pour un statactivisme. La quantification comme instrument d'ouverture du possible », in *Tracés*, 24, p. 123-140 ; FALLON, C., FERON, P., 2014. « Indicateurs de performance et gestion de la qualité comme outils de gouvernement », in *Leurres de la qualité dans l'enseignement supérieur ?*, *op. cit.*, p. 115-139 ; FALLON C., LECLERCQ B., 2015. « Le miroir aux alouettes : De l'opacité politique à la transparence gestionnaire », *op. cit.* ; LASCOUMES, P., LE GALÈS, P. (dir.), 2004. *Gouverner par les instruments*, Paris, Presses de Sciences Po ; LASCOUMES, P., SIMARD, L., 2011. « L'action publique au prisme de ses instruments », in *Revue Française de Science Politique*, 61, 1, p. 5-22 ; LASCOUMES, P., 2004. « La Gouvernamentalité : de la critique de l'État aux

Cela veut-il dire qu'il n'y a plus aucun choix, plus aucune décision politique ? Si, bien sûr ; mais ils sont désormais cachés dans la sélection et le paramétrage des indicateurs. Car c'est là que se joue l'appréciation de ce qui compte, c'est-à-dire très concrètement de ce qu'il convient de compter et de comment le compter.

3. LOGIQUE DE LA RÉPUTATION CHIFFRÉE

Le fait que tous — membres de la commission (de compétences diverses) et experts spécialisés — procèdent à l'évaluation d'après les mêmes indicateurs a bien sûr pour avantage de permettre généralement une assez grande convergence entre les avis de tous, et par là même de faciliter la décision. Mais cela, qui conforte le système d'évaluation en le parant d'effets de consensus et d'évidence, risque bien en fait d'être encore le reflet de ce qu'on ne procède pas à une évaluation au sens propre, mais seulement à une application mécanique de critères d'évaluation (opérationnalisés dans des indicateurs) qui ne sont pas vraiment interrogés.

Outre le confort qu'offre à chaque évaluateur la disponibilité d'indicateurs tout faits qui simplifient grandement son travail et garantissent en outre à ses avis une large convergence avec ceux d'autres évaluateurs mais aussi une certaine incontestabilité (appuyée sur des chiffres), un élément qui explique l'usage (assez peu critique) des mêmes indicateurs par tous repose également dans le fait que les évaluations s'effectuent en permanence à des niveaux multiples de sorte que tous doivent aussi, au moment où ils évaluent, tenir compte de leurs propres scores à l'égard de ces indicateurs à l'aune desquels ils seront eux-mêmes évalués. Ainsi, une université *doit* recruter l'enseignant-chercheur qui a les meilleures performances sur les indicateurs standard, car c'est cela qui améliorera son propre score dans les rankings

technologies du pouvoir », in *Le Portique*, p. 13-14, consulté sur <http://leportique.revues.org/index625.html> ; OGIEN, A., 2007. « La gouvernance, ou le mépris du politique », in *Cités*, 4, 32, p. 137-155 ; OGIEN, A., 2009. « Métamorphose de l'autonomie : l'université, de la rationalisation à la concurrence », in *Quaderni* [En ligne], 69, p. 11-26 ; OGIEN, A., 2013. *Désacraliser le chiffre dans l'évaluation du secteur public*, Versailles, Quae éditions, coll. « Sciences en questions » ; ROSE, N., 1991. « Governing by numbers : Figuring out democracy », in *Accounting, Organizations and Society*, 16, 7, p. 673-692 ; ROUVROY, A., 2011. « Pour une défense de l'éprouvante inopérationalité du droit face à l'opérationnalité sans épreuve du comportementalisme numérique », in *Dissensus. Revue de philosophie politique de l'ULg*, 4, p. 127-149 ; SALAIS, R., 2010. « Usages et mésusages de l'argument statistique : le pilotage des politiques publiques par la performance », in *Revue française des affaires sociales*. 1-2, p. 129-147.

et consécutivement son recrutement d'étudiants, mais aussi sa capacité à obtenir des financements de recherches, tous éléments qui peuvent à leur tour, dans un cercle vertueux, améliorer le ranking... Un fonds de recherches *doit* privilégier l'usage de certains indicateurs dans ses évaluations de chercheurs, d'équipes et de projets, car il doit lui-même rendre des comptes à ses bailleurs de fonds (*accountability*) ou défendre sa réputation d'après ses propres performances à l'égard de ces indicateurs.

D'une manière générale, comme l'a bien montré Roger Dale⁷, le système universitaire fonctionne selon le principe d'une logique de la réputation, où chacun peut s'attirer d'autant plus de collaborations prestigieuses, c'est-à-dire avec des partenaires eux-mêmes réputés (équipes, chercheurs, étudiants, revues...), et d'autant plus de financements, qu'il jouit d'une bonne réputation. Cela, bien sûr, est typique du monde universitaire et n'est pas nouveau. Ce qui est relativement nouveau, c'est que cette logique de la réputation qui était typiquement intensionnelle (c'est-à-dire qu'elle reposait sur une donation de sens et des prises d'attitude) est désormais extensionnelle (et dépend d'équivalences préétablies) : le monde scientifique étant devenu tellement large que chacun ne peut connaître de première main les travaux de tout le monde, la réputation s'est formalisée dans quelques indicateurs de performance standardisés qui permettent la comparaison et le classement.

Et tous font allégeance à ces indicateurs pour ne pas prendre eux-mêmes trop de risques de réputation. Non seulement, en effet, chacun doit, pour rester crédible, veiller à ses propres performances à l'égard de ces indicateurs, mais en outre il doit, pour être un évaluateur crédible, veiller à guider son évaluation du travail des autres d'après ces mêmes indicateurs. Les évaluateurs sont d'ailleurs le plus souvent choisis — par les comités de recrutement, les commissions scientifiques de fonds de recherche, les jurys de bourses... — pour leur réputation, c'est-à-dire pour leurs bonnes performances sur ces indicateurs, et ils sont donc d'autant moins susceptibles de les contester que ces indicateurs valorisent leur propre travail ; à l'inverse, il est évidemment très difficile pour ceux dont la réputation, en termes de performances chiffrées, est plus modeste de mettre en questions les indicateurs qui les disqualifient...

⁷ DALE, R., 2014. « Constructing risk management of Higher Education sector through reputational risk management of institutions ; causes, mechanism and consequences », in *Leurres de la qualité dans l'enseignement supérieur ?*, *op. cit.*, p. 65-84.

Le modèle, donc, se renforce constamment ; une fois en place, les indicateurs jouent un rôle extrêmement normatif en ce qu'ils orientent le travail de chacun, qui s'efforce de préparer ses prochaines évaluations en performant sur ses indicateurs, mais aussi en ce qu'ils tendent à s'imposer à tous en disqualifiant rapidement ceux qui les négligent et en valorisant ceux qui les prennent pour repères de leur propre travail et de leurs évaluations du travail des autres.

Conséquence de ce système, une forte standardisation, non seulement de l'*ethos* universitaire — en termes de stratégies de publication, de mobilité (étudiante et scientifique)... — mais aussi éventuellement des travaux et résultats scientifiques eux-mêmes, dans la mesure où la manière la plus sûre de collaborer avec des équipes réputées, de publier dans des revues bien cotées, d'obtenir des bourses ou des prix prestigieux, c'est de répondre aux attentes de ces équipes, revues ou fondations, et de conformer son travail à leur manière de faire science, à leur paradigme. Même si cela paie parfois, et si c'est sans doute la condition de véritables percées scientifiques, sortir des sentiers battus constitue toujours un énorme risque de réputation. Et ce qui vaut pour la recherche vaut aussi sur le plan de l'enseignement. Hardies seraient les universités qui décideraient de faire entièrement abstraction des indicateurs qui sont au fondement des rankings ; parce que ces rankings existent, toutes les universités doivent désormais se ranger dans un même moule et progresser suivant les mêmes lignes.

4. LES INDICATEURS, OUTILS OU OBSTACLES (ÉPISTÉMOLOGIQUES) DE L'ÉVALUATION ?

Certes, tout ceci, qui s'efforçait d'analyser une certaine logique de l'évaluation articulée autour des indicateurs de performance, ne constitue pas encore nécessairement une critique du système actuel. On pourrait éventuellement dire que, les indicateurs ayant été bien choisis, il est en fait *souhaitable* qu'ils guident mécaniquement tous les processus de sélection d'enseignants, de chercheurs, de projets à financer... et qu'ils orientent uniformément la progression de toutes les universités.

Des questions se posent, cependant, quant à savoir 1) si ces indicateurs couvrent bien toutes les dimensions de la qualité ; 2) si, pour les dimensions qu'ils prennent en compte, ils permettent vraiment de s'en faire une idée plus

précise ou seulement une idée très partielle et partielle. Ces questions font aujourd'hui l'objet d'une importante littérature critique et nous nous permettrons, à cet égard, de renvoyer en particulier aux travaux d'Yves Gingras⁸. Une troisième question sur laquelle nous voudrions pour notre part insister et par laquelle nous voudrions conclure est celle de savoir si, même globalement bien choisis, les indicateurs peuvent systématiquement et quasi mécaniquement guider toutes les évaluations.

Pour éclairer cette question, nous nous appuyerons sur la distinction qu'opère Kant entre jugements déterminants et jugements réfléchissants. Dans la *Critique de la faculté de juger*, Kant affirme que, contrairement aux jugements de connaissance, qui sont « déterminants » en ce qu'ils consistent à appliquer à des données empiriques nouvelles des concepts et catégories théoriques dont on disposait préalablement, mais aussi aux jugements moraux, qui sont « déterminants » en ce qu'ils consistent à appliquer à des situations nouvelles des catégories morales dont on disposait préalablement, le jugement esthétique est un jugement « réfléchissant » au sens où il ne se borne pas à appliquer des concepts déjà tout constitués, mais s'interroge aussi, face à la singularité de ce qu'il doit juger, quant aux concepts mêmes qui devraient guider son jugement⁹. Pour le dire vite, le critique d'art qui appliquerait systématiquement les mêmes cadres théoriques pour former ses jugements de goût sans jamais laisser la possibilité aux œuvres singulières qui sont soumises à son appréciation de l'affecter au point de l'obliger à

⁸ GINGRAS, Y., 2003. « Idées d'universités. Enseignement, recherche et innovation », in *Actes de la recherche en sciences sociales*, 148, p. 3-7 ; GINGRAS, Y., 2008. « Du mauvais usage de faux indicateurs », in *Revue d'histoire moderne et contemporaine (la fièvre de l'évaluation)*, vol. 5, n° 55-4bis, p. 67-79 ; GINGRAS, Y., 2014. *Les dérives de l'évaluation de la recherche*, Paris, Éditions Raisons d'agir ; Voir aussi : ABELHAUSER, A., GORI, R., 2008. « L'imposture bibliométrique », in *Mensuel du SNESup*, 569, p. 16, consulté sur [http://snesupvm.free.fr/Ecrits/BULLETTIN/08-1115_Imposture.Biblio metricque_AA.RG.pdf](http://snesupvm.free.fr/Ecrits/BULLETTIN/08-1115_Imposture.Biblio%20metricque_AA.RG.pdf) ; COLLETIS, G., LAMARCHE T., 2011. « Pour une nouvelle démarche d'évaluation des laboratoires de recherche, des enseignants-chercheurs et des chercheurs en économie », rapport de la Commission d'évaluation de l'Association française d'économie politique, disponible à l'adresse <http://obouba.over-blog.com/article-pour-une-nouvelle-demarche-d-evaluation-83794867.html> ; DUJARIER, M.-A., 2010. « L'automatisation du jugement sur le travail. Mesurer n'est pas évaluer », in *Cahiers internationaux de sociologie*, 128-129, p. 135-159 ; KARPIK, L., 2012. « 'Performance', 'excellence' et 'création scientifique' », in *Revue Française de Socio-Économie*, 10, p. 113-135 ; PRAS, B. (pdt.), 2010. *L'évaluation des Enseignants-Chercheurs en Sciences de Gestion*, Rapport de la Commission de réflexion sur l'Évaluation des Enseignants-Chercheurs en Sciences de Gestion mise en place par la FNEGE ; WEINGART, P., 2005. « Impact of bibliometrics upon the science system : Inadvertent consequences ? », in *Scientometrics*, 62, 1, p. 117-131.

⁹ KANT, E., 1985 [1^{re} éd. 1790]. *Critique de la Faculté de Juger*, Première introduction, § V, Paris, Gallimard, p. 37-44.

interroger la pertinence de ses cadres théoriques, serait en permanence dans le jugement déterminant et jamais dans le jugement réfléchissant ; et il serait assurément un mauvais critique d'art...

La question qui se pose alors est celle de savoir si l'évaluation des activités universitaires peut être entièrement *déterminante* et appliquer systématiquement les mêmes catégories évaluatives ou si elle doit être *réfléchissante* et s'interroger sans cesse sur les critères mêmes de son évaluation. Y a-t-il, demandions-nous plus haut, proprement évaluation là où l'évaluateur applique mécaniquement des critères opérationnalisés dans des indicateurs sans s'interroger vraiment sur ce que devraient être les critères pour juger dans le cas présent ? À en juger d'après le nombre de publications ou celui des mois passés à l'étranger, ce chercheur n'est pas le plus performant de sa génération, mais, à la lecture de ses travaux, je m'aperçois que, plus important encore que tout cela, il y a ici des idées originales, une manière de faire singulière, qui le recommandent à notre soutien prioritaire... La prévalence des indicateurs aujourd'hui laisse-t-elle place à ce type de jugement réfléchissant, à ce type d'évaluation au sens propre ?

Quels types d'évaluation pour quelle recherche et quelle société ?

VÉRONIQUE HALLOIN

Secrétaire générale du F.R.S. - FNRS

1. LE CONTEXTE

La recherche et l'innovation au cœur des stratégies nationales

Depuis une quinzaine d'années, la plupart des pays industrialisés ont décidé d'investir dans la recherche et l'innovation pour améliorer leur compétitivité et assurer croissance économique et emploi, en ligne avec la stratégie de Lisbonne 2000 et la nouvelle stratégie 2020 de l'Union Européenne.

La recherche et l'innovation (et donc aussi les universités) sont désormais considérées comme facteurs d'attractivité et de compétitivité d'un pays et se retrouvent au cœur des politiques nationales ; il en résulte une évolution et une intensification des pratiques d'évaluation de ces dispositifs et de leurs acteurs.

Les activités de recherche sont souvent classées en catégories, pour lesquelles des définitions ont été proposées dans le manuel de Frascati de l'OCDE¹, afin de satisfaire ses propres besoins statistiques : recherche fondamentale, recherche appliquée, mais aussi plus récemment recherche orientée, ou stratégique.

¹ http://www.uis.unesco.org/Library/Documents/OECDFrascatiManual02_fr.pdf.

Nous nous limiterons ici à la recherche fondamentale², qui « est entreprise principalement (mais pas exclusivement) pour produire des connaissances nouvelles sur les fondements des phénomènes et des faits observables, sans nécessairement envisager une application ou utilisation particulière au moment où elle est entamée (ce qui n’implique pas qu’elle n’a pas d’application par la suite) ; ces activités de recherche ne s’exercent pas dans un cadre contractuel fixé et leur financement n’est pas lié directement à un accord préalable sur le contenu. » On la décrit aussi comme recherche libre, ou encore « *basic research* » ou « *curiosity driven research* ».

Cette recherche, essentiellement le fait des universités en Belgique, est surtout financée par les pouvoirs publics.

Un environnement en pleine évolution et de plus en plus compétitif

Si l’on ne peut que se réjouir de l’importance croissante portée par quasi toutes les parties prenantes de nos sociétés aux activités de recherche, celles-ci s’exercent dans un environnement en évolution rapide, marqué par des avancées technologiques qui ont une répercussion tant sur les activités de recherche que sur les pratiques d’évaluation.

L’accessibilité à des bases de données et l’offre d’outils informatiques favorisent le développement d’indicateurs bibliométriques variés et facilitent l’établissement au niveau national ou international de classements visant à comparer les performances de chercheurs ou d’institutions, même si les méthodologies mises en œuvre sont largement discutables.

Le financement de la recherche est depuis longtemps essentiellement basé sur un système d’allocation compétitive des ressources³ ; on assiste par contre un peu partout dans le monde à une augmentation très importante des demandes de financement au cours de ces dix dernières années, qui se traduisent par des réductions continues des taux de succès. Si la compétition est en soi un élément moteur au service de la qualité et de l’excellence et un prix à payer pour la liberté de la recherche, elle est arrivée à un point tel

² Définition inspirée de celles de l’OCDE ainsi que du rapport du groupe de travail de l’Académie royale sur le financement de l’enseignement supérieur en Communauté française (http://www.cref.be/Rapport_fin_Acad_royale.pdf) ; étude réalisée pour le CREF, Conseil des Recteurs des Universités francophones de Belgique, en juillet 2011).

³ Aurélie TRICORE, « Le financement compétitif de la recherche scientifique en Europe et aux États-Unis », in *Lieux communs*, n° 10, 2007, p. 69-91, <halshs-00344808> — <https://halshs.archives-ouvertes.fr/halshs-00344808/document>.

qu'elle peut engendrer des effets négatifs variés : orientation des activités de recherche vers des sujets plus « porteurs », plus à la mode, risque de concentration des moyens sur les équipes les plus performantes, découragement d'équipes de recherche, pression à la publication, fraudes scientifiques...

Cette situation amène à questionner l'équilibre à atteindre entre financement structurel de base, qui devrait être assuré par les institutions dans lesquelles les chercheurs exercent leurs activités, et les financements compétitifs sur projets, plus spécifiques aux agences de financements nationales (par exemple le FNRS en Belgique francophone) et internationales (par exemple l'ERC au niveau européen).

Enfin, l'importance portée à l'innovation amène aussi dans beaucoup de pays une pression nouvelle de justification *a priori* des impacts et retombées attendues d'un projet de recherche⁴, mettant en danger les projets originaux et risqués, voire certaines thématiques de recherche.

De manière plus large en ce qui concerne l'évaluation ex-post, il n'y a actuellement pas de consensus sur la manière d'évaluer les retombées et impacts sociétaux de la recherche⁵ et de la diffusion de la connaissance vers l'ensemble de la société. En ce qui concerne la recherche fondamentale plus particulièrement, il n'existe ainsi pas d'indicateur simple et robuste permettant de décrire son impact sur la qualité de l'enseignement universitaire et la formation d'esprits brillants et la formation d'esprits brillants, ni son importance sur les découvertes et l'innovation de manière générale.

2. UNE INQUIÉTUDE POUR LES PERSPECTIVES DE CARRIÈRE

Le nombre de docteurs diplômés annuellement a plus que doublé au cours des 25 dernières années en Belgique⁶. Cette augmentation très significative pose la question des débouchés pour ces docteurs parfois insuffisamment considérés par les entreprises et la fonction publique.

Une étude de l'OCDE⁷ réalisée sur des cohortes de doctorants diplômés entre 1996 et 2006 montre toutefois que les taux d'emploi étaient très bons,

⁴ *Nature*, vol. 502, 17 octobre 2013, p. 397.

⁵ <http://www.idrc.ca/EN/Documents/Brief-Final-French.pdf>

⁶ <https://www.belspo.be/belspo/ScienceConnection/012/CareersDoctorateHolders.pdf>

⁷ <http://www.uis.unesco.org/ScienceTechnology/Documents/44893058.pdf>

sauf dans une moindre mesure dans le domaine des humanités, et même s'il faut en moyenne 5 à 6 ans avant d'obtenir un contrat à durée indéterminée. Par ailleurs, quelque 20 % d'entre eux, en Belgique, semblent avoir obtenu un emploi qui n'était pas relié à leur niveau de qualification.

La question des débouchés est sans doute plus préoccupante en ce qui concerne les post-docs, dont le nombre croît dans tous les pays ; ils sont un moteur essentiel de la recherche, mais le milieu académique n'offre que peu de possibilités d'engagement à durée indéterminée⁸.

Partageant ces préoccupations et ces légitimes questions qui peuvent être posées quant à l'insertion professionnelle des jeunes docteurs, et, *in fine*, sur l'attractivité de cette filière, le FNRS met actuellement en œuvre des études sur le devenir de ses anciens boursiers et mandataires, via des enquêtes auprès de ceux-ci.

Les résultats de l'enquête menée par l'ESF⁹ auprès de 499 détenteurs d'un diplôme de doctorat (*career tracking pilot project*) permettent d'alimenter ces réflexions.

Les types d'évaluation

L'évaluation est devenue une activité incontournable dans nos sociétés et un prérequis à beaucoup de prises de décision, les indicateurs variés se multiplient donc dans tous les domaines.

En ce qui concerne la recherche et les chercheurs, s'agissant d'argent public, il est assez logique de devoir rendre des comptes sur son utilisation, et donc de mettre en œuvre des procédures d'évaluation. Afin d'éviter des dérives ou effets pervers, il est toutefois essentiel que celle-ci soit adaptée :

- à l'objet évalué (un chercheur, un article, un groupe de recherche, une institution, un projet de recherche...)
- aux objectifs de l'évaluation (une décision de financement, une cartographie, une promotion, un classement, points forts-points faibles...)
- à l'utilisateur du résultat de l'évaluation (le chercheur, une agence de financement, les responsables d'une université, le pouvoir politique...).

⁸ *Nature*, vol. 520, 9 April 2015, p. 144-147.

⁹ <http://www.esf.org/publications/corporate-publications.html>

Ainsi, certains indicateurs bibliométriques tant décriés peuvent être parfaitement adaptés pour réaliser par exemple une cartographie comparative de l'intensité des activités de recherche dans un domaine scientifique dans des pays différents. Même superficielle, une telle analyse simple à réaliser peut fournir des indications intéressantes.

La question du choix des poids à donner aux différents indicateurs d'évaluation retenus est tout aussi délicate et peut refléter un choix politique¹⁰.

Ainsi par exemple pour la sélection d'un projet de recherche qui serait basée sur deux critères d'évaluation (originalité et faisabilité), le tableau qui suit montre que le choix d'un score final calculé à partir d'une moyenne arithmétique pondérée pourrait entraîner l'élimination systématique de projets qui pourraient pourtant être considérés comme intéressants (le projet C n'est jamais classé 1er quel que soit la pondération utilisée).

PROJET	CRITÈRE		PONDÉRATIONS				
	Originalité	Faisabilité	1/2-1/2	1/4-3/4	1/3-2/3	2/3-1/3	3/4-1/4
A	6	19	12,5	15,75	12,66	10,33	9,25
B	20	5	12,5	8,75	10	15	16,25
C	12	12	12	12	12	12	12

En ce qui concerne les chercheurs exerçant des activités de recherche fondamentale, leurs principales missions sont la production et la diffusion de connaissances.

Le chercheur est ainsi confronté au minimum à l'évaluation de ses projets de recherche lors de demandes de financement, laquelle évaluation fait souvent intervenir également une appréciation des activités passées ; dans ces cas, il est évident que des indicateurs bibliométriques sont à manipuler avec beaucoup plus de précautions, et ne remplacent en aucune manière le système de *peer review*, avec ses qualités et imperfections. C'est ce dernier système qui reste d'application pour l'évaluation des articles soumis à publications.

Quelle que soit la méthodologie d'évaluation mise en œuvre, il est courant qu'elle ait ensuite une influence sur les activités futures de l'objet évalué, l'espoir étant évidemment que cette influence soit positive.

¹⁰ http://www.ulb.ac.be/documents/rp/docs/Discours_rentree_2008.pdf

Des constats et des questions

La qualité de ces procédures d'évaluation est une priorité pour un organisme de financement de la recherche fondamentale tel que le F.R.S.-FNRS, qui se doit également d'être attentif aux évolutions des besoins de la recherche et des conditions dans lesquelles celle-ci s'exerce.

Les questions et commentaires qui suivent font partie des sujets de réflexion partagés par la majorité des organismes de financement de la recherche dans le monde.

3. LES IMPACTS DU PUBLISH OR PERISH SUR L'ACTIVITÉ DE RECHERCHE

La pression à la publication, supportée par l'approche des indicateurs bibliométriques, tend à privilégier la quantité au dépend de la qualité et peut donc avoir des implications néfastes sur la qualité de la recherche et sa diffusion, dont : saucissonnage des articles, augmentation parfois artificielle du nombre de co-auteurs, désintérêt pour les projets risqués ou trop originaux, abandon de thématiques étroites, fragilisation des groupes de recherche de petite taille, production excessive de travaux de qualité modérée, fraudes scientifiques...

Certains organismes de financement, tel le DFG (German Research Foundation) en Allemagne, tentent d'enrayer ce phénomène en limitant par exemple à 5 ou 10 le nombre d'articles dont le chercheur peut se prévaloir dans ses demandes de financement.

4. LE CARACTÈRE ENDOGÈNE DE L'USAGE DE LA BIBLIOMÉTRIE

On observera que cette course aux publications et aux indicateurs bibliométriques, de plus en plus dénoncée par les chercheurs, revêt néanmoins un caractère endogène. Ainsi, en ce qui concerne le FNRS, c'est pour répondre au souhait des membres des commissions scientifiques internationales, mises en place pour évaluer les demandes de financement, que le chercheur est invité à fournir plusieurs indicateurs bibliométriques (nombre d'articles,

somme des citations, h-index) dans ses demandes de financement, en utilisant la base de données de son choix.

Les niveaux attendus de ces indicateurs varient évidemment selon les domaines scientifiques, mais sont connus des pairs de ce domaine qui constituent les commissions scientifiques. Ainsi, les experts des humanités et sciences sociales sont parfaitement au courant de ce que les bases de données des journaux scientifiques ne couvrent que partiellement l'ensemble de l'activité scientifique dans leurs domaines.

Ces indicateurs ne constituent qu'un élément parmi d'autres dans les travaux d'une commission, et, même s'ils sont réducteurs et imparfaits, ils n'en sont pour autant pas moins révélateurs de certains aspects des activités du chercheur. La dérive consisterait à leur donner trop d'importance.

5. LES INDICATEURS DE LA QUALITÉ DE LA RECHERCHE

Les output, c'est-à-dire les résultats de l'activité de recherche, sont généralement évalués par le biais d'indicateurs bibliométriques.

Les indicateurs bibliométriques développés, dont le très controversé h-index, ont des points forts et des faiblesses qui font l'objet de larges discussions¹¹. Celles-ci ne convergent d'ailleurs pas selon le domaine scientifique. Ainsi, leur usage en sciences de la vie et de la santé, voire dans certains secteurs des sciences exactes et naturelles, soulève beaucoup moins de débats que dans les sciences humaines et sociales.

S'interroger sur l'impact réel d'une publication est légitime. Le recours à des indicateurs numériques est à la base une idée intéressante vu l'explosion de production d'articles scientifiques. Mais que l'excellence d'un chercheur et l'impact de ses travaux puissent se calculer sur la seule base du nombre de publications et de citations, est évidemment réducteur et discutable.

Un piste alternative ou complémentaire pourrait être, par exemple, de demander au chercheur de joindre à ses demandes de financement 3 articles de son choix, représentatifs de ses activités de recherche, et de commenter pour chacun d'eux ses implications et apports personnels.

¹¹ *UIS Bulletin on science and technology statistics*, Issue n° 2, September 2005. http://www.csiic.ca/PDF/UIS_bulletin_sept2005_EN.pdf – Yves GINGRAS, *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*. Paris, Éditions Raisons d'agir, 2014.

Mais il ne faudrait pas pour autant perdre de vue que les jugements humains des experts peuvent également être subjectifs ou biaisés.

6. LA DIVERSITÉ DES DISCIPLINES

En matière de recherche fondamentale, il est essentiel d'en exercer la politique et la gestion en respectant la liberté de chercher comme fil conducteur. Cela implique une liberté dans les thématiques de recherche et le suivi d'une politique essentiellement *bottom-up*, c'est-à-dire de réponse aux initiatives des chercheurs pour repousser les frontières du savoir dans tous les domaines de la science.

Le problème surgit lorsqu'il s'agit de répartir les moyens financiers entre les différentes disciplines, car il est extrêmement compliqué d'effectuer des comparaisons de qualité de travaux ou de projets entre disciplines différentes.

Par ailleurs, il est également très difficile d'évaluer des projets transdisciplinaires, que ce soit au sein d'une commission disciplinaire ou d'une commission transdisciplinaire *ad hoc* généralement elle-même composée en grande partie d'experts disciplinaires.

7. L'ÉVALUATION PAR LES PAIRS (« PEER REVIEW »)

Le système d'évaluation par les pairs est pratiqué depuis longtemps dans l'évaluation scientifique, surtout pour les articles soumis à publication et les évaluations des demandes de financement.

Généralement considéré comme le système d'évaluation le mieux accepté de la communauté des chercheurs, il fonctionne de manière plutôt satisfaisante, mais est pourtant confronté à une série de problèmes qui ont sans doute pour la plupart d'entre eux existé de tout temps (l'évaluation par les pairs, comme tout processus humain, peut faillir), mais ont pris aujourd'hui une ampleur sans précédent en raison des évolutions évoquées précédemment :

- Le *peer review* dépend d'experts scientifiques de plus en plus sollicités, notamment en raison de l'accroissement considérable du nombre de journaux scientifiques et de publications, mais aussi de la multiplication et l'internationalisation des évaluations institutionnelles ; il en résulte un nombre croissant de désistements d'experts sollicités, et de

nécessaires réflexions, par exemple, sur des mesures incitatives reconnaissant et valorisant ces missions dans les carrières scientifiques.

- Les surcharges évoquées ci-dessus semblent également amener un nombre croissant de publications erronées, les experts n'ayant plus assez de temps à consacrer à la vérification approfondie de découvertes faites de plus en plus rapidement.
- L'anonymat des experts pointus sollicités pour évaluer des articles ou appuyer les travaux des commissions scientifiques (dont les membres sont connus), qui reste à mes yeux une condition indispensable pour une expression libre et une prise de distance, amène toutefois des questionnements sur la confidentialité, surtout dans un environnement de plus en plus compétitif. La signature d'un accord de confidentialité avant l'acceptation d'une mission de *peer review* d'un projet ou d'un article ne constitue en effet pas une garantie suffisante de protection des idées ou résultats qu'ils contiennent.

Le problème est à ce point d'actualité qu'une action COST européenne (PEERE¹²) a été mise en place, pour la période 2014-2018, incluant 26 pays et 60 participants. Elle vise à améliorer l'efficacité, la transparence et la responsabilisation de revue par les pairs.

Le FNS (Fonds national suisse de la recherche scientifique), institution de référence en matière de financement de la recherche en Suisse, a récemment fait l'objet d'un dossier dans sa revue *Horizon*¹³.

8. LES INSTRUMENTS DE FINANCEMENT

En recherche fondamentale, les critères d'évaluation et les décisions d'octroi ne se basent que sur l'excellence scientifique des candidats et projets, garant de la qualité de la recherche.

Il est important de concentrer des moyens significatifs sur les chercheurs les plus brillants, mais il est tout aussi important d'assurer aussi le financement d'une bonne recherche de base et de qualité à travers tous les thèmes,

¹² <http://www.peere.org>

¹³ <http://www.snf.ch/fr/pointrecherche/magazine-de-recherche-horizons/edition-actuelle/>

de manière à permettre à de nouveaux domaines de recherche et de jeunes équipes d'émerger.

C'est au niveau des instruments de financement, de leurs règlements, et de la répartition des budgets entre ceux-ci, qu'un tel objectif peut être atteint. Ceux-ci doivent être régulièrement révisés de manière à répondre à l'évolution des besoins de la recherche et des chercheurs.

9. LE CONTEXTE ÉCONOMIQUE

Une réflexion critique sur les systèmes d'évaluation de la recherche et leur évolution relève d'une démarche scientifique qui ne peut qu'être intéressante et bénéfique.

Cette réflexion doit toutefois être menée en tenant compte du contexte économique ainsi que des conditions matérielles dans lesquelles s'exercent les activités de recherche.

Il ne faut en effet pas perdre de vue que ce sont surtout les moyens financiers insuffisants, et non une faiblesse du système d'évaluation, qui font qu'aujourd'hui tous les très bons (voire excellents) projets de recherche ne peuvent pas être financés. Les systèmes d'évaluation actuels, même si perfectibles, permettent de sélectionner des projets et des chercheurs de grande qualité, qui contribuent à l'excellence reconnue de la recherche scientifique en Belgique.

Évaluations, audits, classements : comment faire ? qu'en faire ?

PIERRE MARAGE

Physicien, ancien doyen de la Faculté des Sciences, ancien vice-recteur à la recherche de l'ULB

1. LE CONTEXTE

Des évaluations inéquitables, nocives, inefficaces, oppressives ? Guidées mécaniquement par des indicateurs quantitatifs (le trop fameux *h-index*) ?

La question de l'évaluation est devenue critique dans le monde universitaire, et un réel malaise se manifeste parmi les chercheurs, ainsi qu'en témoignent diverses interventions lors de la première journée consacrée à ce thème par *Penser la Science*.

Indépendamment des classements internationaux et de la compétition entre institutions, l'évaluation de la recherche affecte directement le vécu des chercheurs à deux niveaux :

- d'une part, l'attribution des financements, où se posent des problèmes d'équité et de répartition, mais aussi de pertinence, d'efficacité et, de manière générale, de réponse à la « demande sociale » ;
- d'autre part, l'évaluation individuelle des carrières des chercheurs, avec également de grandes difficultés d'équité et d'objectivisation, mais aussi de forts enjeux de culpabilisation, notamment pour les « laissés pour compte », et de mal-être au travail.

Cette focalisation du questionnement et du mal-être sur les questions d'évaluation me semble fondamentalement refléter — et aussi partiellement l'occulter en l'individualisant — la place de la recherche dans la société contemporaine et son organisation.

En effet, le savoir scientifique et technique est devenu un enjeu économique majeur, d'où l'importance des retours de financement pour les institutions et les équipes de recherche les plus performantes, et la concurrence exacerbée entre elles.

On assiste également à la marchandisation accélérée de l'enseignement et de la formation, avec une concurrence redoublée entre établissements, illustrée par les fameux classements internationaux.

Enfin, l'organisation de la recherche et le statut des chercheurs sont directement marqués par l'évolution libérale de la société :

- sous-financement du secteur public, dont relève l'essentiel de la recherche fondamentale (mais pas seulement celle-ci) ;
- idéologie libérale de défiance envers les services publics, se traduisant (paradoxalement ?) par des contrôles bureaucratiques redoublés, impliquant de lourdes charges administratives, le tout au nom de la sacro-sainte « transparence » ;
- dérégulation sociale des métiers de la recherche : multiplication des contrats à durée déterminée, mobilité internationale, flexibilité horaire ; le monde de la recherche constitue un secteur d'« avant-garde » de la dérégulation sociale, celle-ci étant facilitée par la conjonction du culte de l'excellence des individus et de l'« idéalisme » des motivations.

La question est donc de savoir comment distribuer les ressources pour la recherche et sélectionner les chercheurs, efficacement et équitablement, en assurant

- la défense de la valeur humaniste de la production de savoirs, dans la reconnaissance de leur diversité ;
- la prise en compte de la « demande sociale » —, tout en reconnaissant qu'elle émane de groupes sociaux variés, aux intérêts divers et souvent contradictoires ;
- le respect fondamental de l'évaluation par les pairs ;

- la nécessaire indépendance de la recherche publique, lieu de critique et de parole libre ;
- et un statut digne pour les enseignants-chercheurs, en tant que *travailleurs*.

Dans ce document, j'essaierai de dégager, comme me l'ont demandé les organisateurs, quelques propositions concernant l'évaluation, sur base de mon expérience.

Mais je dois encore souligner ici toute l'acuité de la problématique de la recherche contractuelle.

En raison du caractère compétitif, aléatoire et instable du financement de la recherche, et en raison de l'accroissement et de l'élargissement constants des missions des enseignants-chercheurs engagés au cadre, les chercheurs sur contrat, notamment les post-docs expérimentés, jouent dans les institutions un rôle croissant, tant pour la continuité de la recherche que pour leur contribution aux missions d'encadrement, d'enseignement et de soutien logistique.

Or la carrière de chercheur sur contrat est particulièrement dérégulée : très grande mobilité internationale, contrats temporaires de courte durée et à répétition, peu de perspectives de stabilisation, peu de perspectives de promotion, problème lancinant du financement et de la fin de carrière des chercheurs de plus de 40-50 ans, etc.

Aucun doute que la situation particulièrement précaire des chercheurs sur contrat est un facteur majeur de malaise au sein des institutions, qui s'en trouvent affaiblies et déstabilisées dans l'ensemble de leurs composantes. Je n'y reviendrai cependant pas ici, car elle ne relève pas directement de la question de l'évaluation qui est l'objet de ces journées de *Penser la Science*.

2. LE FINANCEMENT DE LA RECHERCHE : EN AMONT DE L'ÉVALUATION

La complexité du système de financement

Comme le soulignent dans leurs interventions J.-P. Bourguignon et V. Halloin, le financement de la recherche constitue un « système » aux composantes multiples, qui varient largement dans leur ampleur et dans leurs objectifs (du

reste souvent insuffisamment explicités, voir ci-dessous) : pour nos universités de la FWB, on relève les financements universitaires, dont les Actions de Recherche concertées (ARC), les divers types de financement par le FNRS, les Pôles d'Attraction Interuniversitaires (PAI), les fonds européens dont les financements de l'ERC, les financements régionaux, etc.

L'existence de ce « système » est une source utile de diversification. Cependant, elle conduit à une complexité parfois mal exploitée et souvent mal vécue par les chercheurs. D'où leur recours fréquent à des sources « routinières », qu'ils ont déjà pratiquées, ou à celles qui leur apparaissent comme les plus proches (particulièrement les ressources universitaires ou facultaires), même si celles-ci ne sont pas les plus adaptées, en raison de leur exigüité et de l'hyper-compétitivité qui y règne.

C'est la responsabilité des autorités universitaires d'informer les chercheurs quant à la calibration de leurs demandes selon les sources et de les orienter de façon à optimiser l'exploitation des ressources extra-universitaires, tout particulièrement européennes et régionales, et à optimiser l'énergie dépensée à déposer des projets, en évitant une hyper-concurrence interne improductive et source éminente de frustrations.

Pour les sources de financement les plus importantes, en particulier internationales (ERC etc.), nos universités devraient sans doute bien davantage aider les promoteurs à monter les projets et à répondre aux appels. Et, pourquoi pas ?, collaborer à cet effet...

Contre l'opacité, l'explicitation des objectifs stratégiques

Aux yeux de nombreux chercheurs, une réelle opacité affecte souvent les priorités institutionnelles dans l'attribution des crédits.

Or l'explicitation des objectifs stratégiques visés par le financement est indispensable, tant pour les candidats que pour les évaluateurs.

Pour prendre un cas simple : au FNRS, les Crédits de Recherche visent le soutien à une recherche de base de qualité, alors que les Projets de Recherche se veulent nettement plus sélectifs. Les choses devraient être claires pour tous à ce propos.

Mais, dans le cas des ARC, quel poids faut-il donner à la dimension « concertée », et dans quel contexte faut-il l'interpréter ?

Il est donc indispensable que les instances de financement définissent clairement leurs priorités. Par exemple, s'agit-il de soutenir en priorité des

jeunes chercheurs, des projets interdisciplinaires, des équipes confirmées ? Attend-on de financements issus de l'institution un effet d'entraînement pour d'autres financements, notamment régionaux ou européens, ou plutôt le soutien à la recherche de base ? Vise-t-on à favoriser la création de plateformes, ou chaque projet doit-il être évalué en soi ? Convient-il de procéder à un certain équilibrage entre disciplines ?

De tels choix stratégiques ne peuvent être laissés à l'appréciation et à la responsabilité des évaluateurs. En l'absence d'indications claires, ceux-ci sont extrêmement embarrassés, et les décisions peuvent prendre un aspect aléatoire. Quant aux chercheurs, l'absence ou l'insuffisante précision de la définition des objectifs peut entraîner méfiances, aigreurs, et stratégies défensives.

Dans un certain nombre de cas, des règles claires peuvent être prises (même si chaque procédure peut évidemment introduire des biais). Pour la répartition entre disciplines, par exemple, on observe différentes pratiques. Au FNRS, la répartition des financements entre commissions est proportionnelle aux montants des demandes. Pour les ARC, l'ULB a décidé par contre une répartition équivalente entre les trois grands secteurs de la recherche, visant délibérément à favoriser les sciences humaines. Pour l'attribution des PAI en cours, certaines universités ont suivi strictement les rapports d'évaluation scientifique, alors que d'autres ont fait des choix stratégiques — mais ceux-ci ont-ils été explicités auprès des chercheurs, et comment ont-ils été définis ? Quant aux financements régionaux, le soutien à des secteurs stratégiques est évidemment privilégié (Plan Marshal).

Une problématique sensible est celle des équipes et des personnalités « vedettes », risquant d'accaparer les financements. C'est ainsi que le FNRS a défini des règles de non cumul des financements, ou que l'ULB a décidé que les ARC ne peuvent être cumulés avec des bourses de l'ERC ni concerner deux appels successifs. Mais par ailleurs, pour le positionnement international des équipes les plus compétitives, il serait évidemment absurde de les écarter systématiquement des financements locaux...

L'explicitation des objectifs de la part des institutions devrait être la règle, même si elle requiert non seulement un réel effort intellectuel, mais aussi un certain courage politique...

3. ÉVALUATION ET CLASSEMENT DES PROJETS DE RECHERCHE

Disons d'abord que les procédures d'évaluation ne doivent bien entendu pas s'appliquer de manière identique à tous les types de financement. Il convient clairement de ménager des procédures de candidature et d'évaluation légères pour les financements de base, souvent relativement modestes, pour lesquels il est essentiel de privilégier la confiance aux chercheurs et d'éviter une compétition exagérée. Par contre, des procédures plus lourdes se justifient pour les programmes sélectifs dotés de financements importants.

Comme on le sait, les évaluations et les classements sont souvent opérés en deux étapes (au FNRS, pour les ARC, à l'ERC, etc.) :

- l'évaluation des projets individuels par des experts de la (sous-) discipline ;
- le classement opéré par des commissions interdisciplinaires (ou tout au moins chapeautant plusieurs sous-disciplines).

Le malaise de nombreux chercheurs face aux évaluations est souvent alimenté par un manque de confiance dans l'objectivité des évaluations de premier rang, et par des doutes concernant le fonctionnement des commissions de classement.

Il est donc important que règne la transparence à ces égards, et aussi que soit assuré *a posteriori* un retour adéquat des évaluations vers les chercheurs.

Mais, outre la confiance, se situe évidemment ici également un enjeu d'efficacité pour les institutions.

Les évaluations de premier rang

Les experts de premier rang sont invités à remettre un avis sur chaque projet séparément, sans qu'ils aient connaissance des autres projets, et donc sans qu'ils puissent formuler leurs évaluations en termes relatifs.

Les experts doivent donc pouvoir situer leur travail en fonction des objectifs stratégiques du financement concerné, compte tenu du contexte institutionnel et de ses priorités (cf. ci-dessus).

Les avis sont demandés sous forme de réponses à des questionnaires portant en général sur l'évaluation du projet, le CV des promoteurs, l'environnement de recherche.

Pour être pertinents, ces questionnaires doivent à mon avis respecter les principes suivants :

- les questions doivent être en nombre suffisant pour couvrir de façon suffisamment fine l'ensemble des critères à prendre ultérieurement en considération par les commissions de classement ; des questions trop générales et en petit nombre sont souvent difficiles à interpréter, car recouvrant des dimensions différentes (exemple : « impact du projet » : à quel point de vue ? pour la recherche fondamentale ? pour le développement industriel ? pour le développement de collaborations nationales ou internationales ? pour le développement de l'équipe ? — de telles dimensions devraient être séparées dans le questionnaire) ;
- pour chaque question, il faut absolument demander aux experts des commentaires qualitatifs suffisamment détaillés, afin d'éclairer les discussions en commission ;
- afin de pondérer les différentes dimensions de l'évaluation, il est très utile de demander aux experts un jugement récapitulatif global sur le projet ;
- en tout cas, les évaluations quantitatives (notations chiffrées) devraient être évitées, et même tout simplement proscrites, car éminemment susceptibles d'une variabilité incontrôlée.

Ce dernier point est important. En effet, les experts n'examinant qu'un seul projet et l'échelle relative ne leur étant pas connue, ils ont tendance à attribuer beaucoup de notes élevées, avec parfois l'une ou l'autre note plus basse, visant selon eux à assurer la « crédibilité » de leur évaluation. Il en résulte une faible dispersion des notes, se jouant sur des fractions de point, mais aussi une part d'aléatoire dépendant de la psychologie de l'un ou l'autre expert. Dès lors, même si la quantification offre une méthodologie commode (et paresseuse...) pour les organes de décision, on peut douter de la comparabilité des évaluations et de la stabilité des résultats.

Enfin, même si l'identification et le recrutement d'experts de premier rang sont souvent ardues et sont compliqués encore par des règles de « conflit d'intérêt » souvent beaucoup trop contraignantes (et d'ailleurs illusoire !),

il devrait être de règle que, pour chaque projet évalué, trois évaluations au moins soient recueillies, et si possible quatre : ceci doit permettre de limiter les fluctuations aléatoires, mais aussi permettre à la commission de classement d'identifier les avis marginaux éventuellement à écarter (voir ci-dessous).

Comme cette exigence en particulier implique un effort administratif important pour assurer le suivi du collationnement des réponses, il faut réserver ces procédures assez lourdes aux financements les plus sélectifs — et il convient que les institutions osent affronter d'éventuelles critiques superficielles des médias ou du monde politique concernant la soi-disant « transparence ».

Commissions de classement

Avant toute chose, il faut insister encore sur le fait que c'est le devoir des institutions que de dégager clairement les objectifs stratégiques du financement et la finalité de la sélection, et d'en informer les membres des commissions de classement.

La présence lors de la réunion de personnalités représentant l'institution, pouvant ou non prendre part aux débats, est importante à cet effet. Ces personnes doivent également pouvoir, le cas échéant, fournir aux membres de la commission (notamment étrangers) des informations factuelles concernant le fonctionnement de la recherche en Belgique ou dans l'institution visée.

Le travail des commissions est de procéder au classement de projets émanant de disciplines (ou sous-disciplines) parfois fort variées.

Pour cela, il n'est pas nécessaire (ni possible !) que tous les domaines (sous-)disciplinaires soient représentés au sein des commissions. Mais il est indispensable que les commissions soient composées de personnalités ayant une bonne expérience de l'évaluation, car il leur revient d'interpréter les évaluations et les commentaires des experts de premier rang à la lumière des objectifs stratégiques du financement.

En effet, quelle que soit leur appartenance disciplinaire, les membres des commissions de classement n'ont pas pour tâche principale d'examiner directement chaque projet de recherche, ce qui serait d'ailleurs inévitablement la source de biais disciplinaires, mais de traiter les rapports des experts de premier rang. Cependant, il faut certainement veiller à ce qu'existe au

sein de la commission une expertise suffisante concernant les différentes « cultures » qui peuvent exister entre (sous-)disciplines, lesquelles peuvent affecter non seulement les projets et les CV des promoteurs, mais aussi marquer la méthodologie d'évaluation des experts.

En vue de l'établissement d'un classement, les membres des commissions doivent procéder à la discussion critique des rapports des experts — et notamment prendre éventuellement une distance critique particulière à l'égard d'avis extrêmes. Pour cela, ils doivent s'appuyer sur les commentaires qualitatifs des experts.

En tout état de cause, si des indicateurs quantitatifs sont produits (notamment concernant les CV des promoteurs), ils ne peuvent constituer qu'un élément d'information secondaire et ne doivent en aucun cas être utilisés mécaniquement.

Transparence de la procédure et retour des évaluations

La transparence de la procédure et le retour vers les équipes constituent des dimensions très importantes devant accompagner l'évaluation des projets. Il en va d'une part de la confiance des équipes envers la procédure d'évaluation, d'autre part de leur permettre de bénéficier au mieux des avis des évaluateurs.

Pour assurer la transparence, les objectifs du financement doivent évidemment avoir été bien définis et connus des chercheurs.

Les principes de fonctionnement des évaluations de premier rang, en particulier le contenu des questionnaires d'évaluation, et ceux des commissions d'évaluation devraient être connus des chercheurs, avant même le dépôt des candidatures. Ceci doit également permettre aux chercheurs de mieux comprendre les objectifs de l'évaluation, et de mettre en évidence dans les dossiers les aspects les plus pertinents (indépendamment de leurs mérites scientifiques, tous les chercheurs ne sont pas de grands stratèges ni de grands communicateurs...). La composition des commissions de classement devrait également être connue.

En ce qui concerne le retour vers les équipes, une règle évidente doit être la communication des rapports des experts de premier rang, qui constituent une ressource très importante pour les équipes.

Pour ce qui concerne le retour des commissions de classement, la question est plus compliquée. Il faut en effet, pour assurer le bon fonctionnement des commissions, que soit assurée la confidentialité des débats. D'autre

part, s'agissant d'un classement, ce sont les mérites relatifs des différentes demandes qui interviennent. Idéalement, un rapport approuvé par les membres de la commission (ou le président et le secrétaire) devrait informer tous les candidats sur les principaux critères et leur hiérarchisation, sans pour autant entrer dans le détail des débats.

En outre, c'est la responsabilité des autorités académiques d'organiser une rencontre avec chaque promoteur, afin de discuter avec lui du dossier déposé, à la lumière des évaluations de premier rang mais aussi de certains traits généraux apparus lors des débats en commission.

4. ÉVALUATIONS INDIVIDUELLES

Le milieu académique est le lieu d'une très grande compétition entre chercheurs pour les recrutements à titre définitif, mais aussi pour les promotions et le développement de la carrière.

Comme souligné par plusieurs intervenants, cette compétition est parfois vécue de façon très douloureuse par les personnels. Outre les contraintes exercées par un milieu de travail socialement dérégulé, le moment de l'évaluation, en vue d'un recrutement ou d'une promotion, est un moment particulièrement aigu du « mal-vivre » des chercheurs et des enseignants-chercheurs.

C'est ce que souligne notamment B. Fuselier, en relevant chez les chercheurs mais aussi dans les documents académiques officiels les expressions fréquentes d'investissement « total » ou « entier », par opposition à ce que serait un « travail normal ». Et A. Eraly insiste sur le « syndrome de l'imposteur », et souligne que « le système produit du sentiment d'infériorité en quantité industrielle ».

On retrouvera, sur la question des évaluations individuelles, beaucoup des remarques formulées à propos de l'évaluation des projets de recherche.

Même si c'est certainement difficile, les institutions doivent impérativement définir leurs objectifs, faire connaître le fonctionnement du système d'évaluation et assurer le retour d'évaluation vers les chercheurs. En outre, se posent avec une particulière acuité les problématiques liées à la bibliométrie et à la question du genre.

Recrutements à titre définitif

Qu'il s'agisse de postes profilés (chaires d'enseignement et de recherche) ou non profilés (notamment les postes permanents au FNRS), il faut que les institutions, comme pour le financement de la recherche, définissent clairement et fassent connaître leurs objectifs stratégiques.

Elles doivent en particulier préciser leurs attentes, de manière cohérente, en ce qui concerne les missions respectives d'enseignement, de recherche et de service à la collectivité, et dire selon quels schémas, éventuellement variés, ces missions peuvent se combiner. Le manque de clarté à ce propos est source de beaucoup d'incompréhensions.

Les procédures de recrutement comportent plusieurs niveaux :

- dans le cas de concours (par exemple pour les postes de chercheur qualifié au FNRS ou pour d'autres postes non profilés), une évaluation par des experts extérieurs de premier rang ;
- souvent, une audition, sous forme d'une rencontre avec une commission de classement et/ou la présentation d'un séminaire et/ou d'une leçon ;
- la production d'un classement par une commission *ad hoc*.

Pour les évaluations par des experts extérieurs, les remarques formulées à propos de l'évaluation de la recherche restent d'application : questionnaires suffisamment détaillés pour couvrir tous les aspects du dossier ; importance décisive des commentaires qualitatifs ; jugement global ; éviter les notations chiffrées.

De même pour le fonctionnement des commissions et l'évaluation des CV : valoriser l'expérience des membres dans l'évaluation plutôt que viser la représentation des (sous-)disciplines, mais veiller cependant à une bonne connaissance des cultures de celles-ci ; rôle décisif des commentaires qualitatifs des experts ; en même temps, distance critique par rapport à ceux-ci ; et très grande prudence par rapport aux traductions quantitatives et à la bibliométrie (voir ci-dessous).

Pour ce qui est des auditions, elles fournissent souvent aux membres des commissions des informations importantes, notamment sur la personnalité des candidats, leur vision de leurs missions, l'ampleur de leur connaissance de leur domaine de recherche, et leurs perspectives stratégiques. Pour sa part, le FNRS ne procède pas à des auditions pour les postes hautement

compétitifs de Chercheur qualifié (au contraire, paradoxalement, des bourses de doctorat du FRIA et du FRESH), la raison avancée étant le grand nombre de candidats ; on pourrait pourtant sans doute trouver des modalités de présélection permettant de contourner cette difficulté. Il faut cependant noter que, généralement, les universités procèdent elles-mêmes à des auditions parmi les candidats CQ sélectionnés par le FNRS.

Promotions

Bien que moins aiguë et douloureuse que dans le cas des recrutements, la problématique des promotions est importante pour les individus et pour les institutions. Outre l'aspect matériel, la dimension symbolique est ici cruciale : il est question de reconnaissance par les pairs et par les autorités académiques, d'image de soi, d'insertion dans le milieu professionnel, et donc aussi, pour les institutions, de motivation des chercheurs.

Les problématiques générales sont semblables à celles liées au recrutement, avec une acuité particulière de la nécessaire clarification des attentes des institutions. Mon expérience est que, parmi les académiques, règne souvent une grande confusion concernant le poids relatif des différentes missions universitaires. Les institutions devraient dire le poids qu'elles accordent, en vue des promotions, à la qualité de l'activité dans chacune d'entre elles, et/ou à une excellence particulière dans l'une d'elles. Elles devraient aussi clarifier ce qu'elles entendent par un rôle de « direction » (en particulier pour les Directeurs de recherche du FNRS).

Une clarification hautement souhaitable de ces problématiques passe par l'affichage par les institutions (universités, FNRS) des perspectives de carrière : taux de promotions aux différents niveaux, perspectives en début de carrière, durées d'attente, distribution des niveaux atteints en fin de carrière.

La problématique de la transparence est ici cruciale, ainsi que celle du retour vers les candidats.

Du « bon usage » de la bibliométrie dans l'évaluation des CV

Le bon usage / ou le meilleur usage possible / ou le mauvais usage supposé / ou encore le mauvais usage intrinsèque / de la bibliométrie dans l'évaluation des CV par les commissions de classement focalise beaucoup d'interrogations, de mal-vivre, de fantasmes, et d'espérances déçues de la part des chercheurs.

Lors des journées de *Penser la Science*, cette problématique a été abordée par de nombreux intervenants, en particulier Y. Gingras, H. Bersini, V. Halloin.

Cette problématique est particulièrement sensible dans le cas des procédures de recrutement à titre définitif, mais aussi pour les promotions et pour l'attribution des crédits de recherche, où le CV des promoteurs est l'une des dimensions prises en compte.

La complexité de cette problématique réside dans la tension entre, d'une part, l'évidente utilité d'évaluer la productivité scientifique, basée sur l'évaluation par les pairs (*peer review*), et d'autre part la reconnaissance de l'existence des biais très réels qui peuvent affecter cette évaluation.

Pour résumer et pour mémoire (voir notamment les interventions mentionnées), ces biais peuvent notamment concerner les dimensions suivantes :

- biais affectant l'évaluation quantitative de l'activité : nombre d'articles publiés // stratégies de détournement : « saucissonnage » d'articles, multiplication de publications ne différant que sur des détails, etc. ;
- biais affectant les tentatives de quantification de la valeur des publications : nombre de citations, facteur d'impact des revues // biais dus aux stratégies commerciales des éditeurs, aux autoréférences et aux « renvois d'ascenseur », etc. ; il faut cependant insister sur le fait que, malgré des cas aberrants bien connus (articles erronés d'auteurs prestigieux), le nombre de citations constitue sans aucun doute une manière efficace d'estimer la valeur d'un article ;
- biais (indubitablement très forts) entre disciplines, notamment (mais pas seulement) au détriment des sciences humaines : indexation des publications et des citations dans des bases de données contrôlées par les principaux éditeurs, domination des revues anglo-saxonnes, biais liés à la langue de publication, etc. ;
- biais liés aux stratégies de publication des équipes.

Il n'existe évidemment pas de solution simple et universelle à cette problématique.

Mais, à mon avis, les autorités académiques et les institutions devraient donner des directives très claires aux commissions de classement, agissant à quelque niveau que ce soit : en aucun cas, la bibliométrie basée sur le recours

aux bases de données de publications, de citations et de facteurs d'impact ne peut constituer le facteur unique ni déterminant d'un classement.

La question du genre

L'un des biais principaux affectant le recrutement et les promotions dans le cadre académique est le biais de genre.

Combattre ce biais relève non seulement de l'équité entre individus, mais comporte aussi un enjeu majeur d'efficience, pour les institutions et pour la société dans son ensemble, en termes de mobilisation des ressources et des capacités.

Une question complexe est la suivante. Il est bien connu que le biais dans la représentation des femmes dans la carrière académique intervient dès le niveau des candidatures (plus faible représentation des femmes parmi les candidats aux postes permanents ; plus faible taux de candidature pour les promotions). Cette observation ne peut cependant servir d'« excuse » aux biais de sélection, qui renverrait de fait à une culpabilisation des femmes. En effet, le biais de candidature reflète le biais dans les perspectives de carrière des femmes dans la société dans son ensemble (se traduisant notamment par l'inégale répartition des tâches au sein des couples), et au sein des institutions académiques en particulier.

Sur ce dernier point, il faut rappeler une fois de plus les biais puissants qui affectent les processus de promotion. Le milieu académique reste massivement dominé par une image masculine des fonctions dirigeantes et des personnalités prestigieuses (recteurs et présidents d'universités, prix Nobel, docteurs *honoris causa*, etc.). À moins d'une politique déterminée des autorités académiques, cette image domine les commissions de sélection, qui à leur tour reproduisent la hiérarchie de genre.

Les autorités académiques devraient donc mener une action systématique et de longue haleine, comportant notamment les axes suivants :

- assurer à tous les niveaux au sein des commissions d'évaluation une présence des femmes visant à la parité ;
- fournir / exiger, à tous les niveaux, des statistiques incluant la dimension de genre (effets pédagogique et vérificatif) ;
- évaluer systématiquement les biais de genre potentiels induits par toutes les mesures académiques, quelles que soient par ailleurs leurs motivations. Par exemple : biais de genre induits par l'exigence d'une

mobilité post-doctorale à l'étranger ; biais de genre (liés aux maternités) affectant les nominations tardives ; biais de genre dans l'évaluation de la production scientifique (prise en compte des maternités/paternités), etc. ;

- l'expérience démontrant que la bonne volonté ne suffit pas, les institutions académiques doivent assurer la formation systématique d'experts dans les questions de genre, appelés à œuvrer au sein des commissions de classement ;
- de même, les institutions doivent assurer une formation de base sur les questions de genre des membres appelés à siéger au sein des commissions, et en particulier des présidents et secrétaires ;
- enfin, et surtout, l'enjeu essentiel est celui du développement au sein des institutions de contextes de travail satisfaisants pour les femmes... et pour les hommes.

Le retour vers les candidats

Plus encore que dans le cas du financement des projets de recherche, le retour vers les candidats est, dans le cas de l'évaluation des personnes, une question de respect, et aussi d'efficacité pour l'institution.

Les mêmes règles de base doivent s'appliquer que dans le cas des demandes de financement :

- communication automatique aux candidats de l'intégralité des rapports des experts de premier rang ;
- rédaction et communication aux candidats d'un rapport de la commission de classement ;
- proposition aux candidats d'une rencontre-entretien de débriefing.

Cette rencontre doit permettre d'analyser avec le candidat malheureux les critères qui ont guidé le choix de la commission, de discuter les forces et les faiblesses de son dossier, d'évaluer avec lui les chances d'une nouvelle candidature éventuelle. Elle doit donc l'aider à améliorer son dossier, ou à envisager en toute lucidité une réorientation de ses perspectives de carrière.

La proposition de rencontre est particulièrement importante pour les candidats internes, au moins ceux qui ont été présélectionnés. En outre, il devrait être entendu que tous les candidats peuvent en tout cas solliciter un entretien afin d'être éclairés sur l'évaluation de leur dossier.

L'organisation de telles rencontres de débriefing constitue évidemment une charge pour les institutions, mais c'est un facteur d'amélioration des dossiers des candidats et donc de la recherche institutionnelle, et c'est aussi un facteur d'amélioration de l'ambiance de travail et de cohésion de l'institution.

5. AUDITS ET ÉVALUATIONS DE LA RECHERCHE

Pour terminer cette note, quelques mots sur les expériences d'évaluation de la recherche et d'audit.

Les institutions académiques prennent des initiatives de divers ordres en vue d'évaluer, avec le soutien d'experts extérieurs, la qualité de leur recherche.

Ces initiatives peuvent sans aucun doute être utiles. Mais l'expérience montre qu'elles peuvent également être ressenties par les chercheurs comme agressives, répressives ou culpabilisantes — ou encore lourdes, contraignantes et peu utiles.

Ici encore, une partie du malaise peut être dû à un manque de clarté dans la définition des objectifs de l'exercice. En gros, deux options sont ouvertes :

- s'agit-il d'une aide à la décision des autorités ? Il s'agit alors à proprement parler d'« évaluations » ;
- s'agit-il principalement d'aider les équipes existantes à améliorer leurs performances ? Auquel cas il s'agit plutôt d'« audits ».

Dans le cas d'évaluations à proprement parler, il faut que soient clairement précisées les décisions des autorités visées par le processus, ainsi que leurs champs d'application concrets. Par exemple : attribution de financements tels que PAI, ARC ou autres financements relevant d'arbitrages internes ; recrutements ; soutiens spécifiques de la part de l'administration, etc. En outre, il faut que soit d'emblée énoncé l'impact attendu des décisions des autorités, et la manière d'évaluer leurs effets.

Dans le cas d'audits, la dynamique est toute autre. Il s'agit pour les équipes de formuler leurs objectifs, y compris les indicateurs qui leur semblent pertinents, et de définir leur stratégie en vue de les atteindre. L'institution leur offre alors une ressource extérieure, afin d'examiner avec elles la pertinence

de ces objectifs et la stratégie proposée pour les atteindre, à la lumière des ressources disponibles et du contexte compétitif local et général.

En ce cas, la phase d'auto-évaluation constitue l'étape la plus importante du processus : comme le dit D. Pestre dans son intervention : « Le but premier de l'évaluation est l'auto-évaluation ».

Enfin, l'exercice demandé aux équipes de recherche n'est crédible qu'à la condition que l'institution en tant que telle s'y livre elle-même et accepte de jouer le jeu, car les audits locaux révèlent souvent des problèmes relevant de l'institution dans son ensemble.

Mise en perspective historique des « évaluations » de la recherche publique

Remarques sur la gouvernementalité libérale et
le gouvernement par le nombre

DOMINIQUE PESTRE

Directeur d'études à l'École des Hautes Études
en Sciences Sociales — EHESS (Paris)

L'évaluation est une forme intéressante de la gouvernementalité libérale qu'il convient de décrypter. Elle répond à une logique dont le cœur est de réinstaller l'individu « en charge de lui-même » et de mettre en place des dispositifs qui lui permettent de se gérer de façon « optimale ». Du moins est-ce ainsi que parlent ses promoteurs.

La méthode consiste à fixer des *objectifs* qui sont à réaliser, des objectifs *de bon sens, évidents, incontestables* — par exemple le devoir qu'a chacun de s'améliorer, le fait que les universités doivent viser l'excellence, le fait que la médecine doit reposer sur des preuves, etc. Il s'agit d'en déduire une *métrique* par rapport à laquelle les individus et les organisations se jaugeront et compareront leurs performances ; ceci afin de repérer les meilleures pratiques, celles qui pourront être recommandées à tous. Au-delà ou en-deçà de tout débat d'idées interminable sur les valeurs et les objectifs, il s'agit d'inventer un *modus operandi* pragmatique qui vise à faire que chacun ait

l'information « objective » dont il a besoin pour se « mesurer aux autres » et « progresser ».

Les éléments d'une telle gouvernementalité ne sont évidemment pas neutres. Ne pas perdre son temps avec le débat d'idées — et même se méfier des intellectuels critiques dont la pensée complexe va à l'encontre du bon sens —, c'est supposer une société consensuelle dans laquelle la confrontation démocratique des positions et choix n'a plus à se poser. Quant aux principes supérieurs indiscutables qui sont donnés au départ, ils sont le plus souvent puisés au vocabulaire du *New Public Management* : se prendre en main, se faire meilleur — ce qui signifie en pratique être plus « actif », « entreprenant », « efficient ».

L'originalité est qu'au lieu de décider la norme d'en haut, par des instances législatives ou juridiques par exemple, le dispositif « laisse » la norme se construire « par le bas » (ainsi va la doxa) à travers la confrontation des « acteurs de la société civile » — des organisations internationales, des entreprises, des associations, des think tanks.... Cette « confrontation » opère sur « un marché des idées » en théorie ouvert mais de fait très structuré et asymétrique, duquel « émerge » un « plus petit dénominateur commun » qui devient la norme directrice (norme d'emploi, indicateurs d'excellence, ...). Se crée ainsi une métrique décrétée « transparente » (puisque fondée sur des indicateurs visibles par tous), mais aussi provisoire (elle est volontiers reconnue comme imparfaite, mais on se promet de l'améliorer sans cesse), et qui permet à tous de (ou *oblige chacun* à) se mesurer et se laisser guider par ce *benchmark*. Le travail de normalisation des individus commence alors, avec des primes aux meilleurs, des coupes pour les récalcitrants, mais aussi et surtout une conformation volontaire de chacun aux normes présentées comme un idéal à atteindre par tous. S'engendrent ainsi des effets systémiques, intentionnels et pervers, notamment par adaptation aux logiques du système.

Le paradoxe de ce « libéralisme » est qu'il est très interventionniste puisqu'il vise à « ré-initialiser » les individus et les institutions dont il dit pourtant qu'elles doivent fonctionner librement. Il lui faut fabriquer un monde très instrumenté (pour construire les informations nécessaires à la comparaison par exemple) afin de permettre aux individus de « bien choisir » leur voie et d'aller dans la bonne direction, vers le bien.

1. L'EXEMPLE DU BENCHMARKING DES HÔPITAUX EN FRANCE (SELON FRÉDÉRIC PIERRU¹)

Science et Avenir, journal de popularisation scientifique, est en 1997 à l'origine du premier palmarès des hôpitaux français. Intitulé la « liste noire des hôpitaux » (ce qui est soi très symptomatique d'une intention de stigmatisation), l'article initial suscite immédiatement de vives critiques (sur les critères choisis, la fiabilité des données...) et une demande d'arrêt de la publication de ce type de classements. Toutefois, dans la mesure où les palmarès font vendre, *Science et Avenir* continue à publier année après année son palmarès, en veillant toutefois à améliorer la fiabilité de son information. Et, pour ce faire, le journal s'appuie sur les données officielles du ministère, à un moment où celui-ci souhaite lui-même activer une phase de transformation de la gestion des hôpitaux vers davantage de professionnalisation et de rentabilité. De plus en plus sophistiquées en termes de données, les éditions suivantes du palmarès sont aussi de plus en plus cadrées en termes de critères de qualité et d'objectifs d'amélioration... le tout allant globalement dans le sens de la réforme des hôpitaux voulue par le gouvernement.

Du fait de sa simple existence (et de sa répétition année après année), le palmarès produit progressivement ses effets. Le public s'est fait à l'idée qu'il y a de bons et de mauvais hôpitaux et il n'est plus possible pour les acteurs de santé de ne pas se soucier de leur place dans le palmarès. Les hôpitaux doivent se soumettre à la logique du palmarès ; leur action est désormais tout entière orientée vers les performances à l'égard des critères retenus dans le palmarès. La réforme des hôpitaux se met ainsi en place tout « naturellement » du fait que chaque acteur agit conformément à la métrique et met lui-même en œuvre ce qu'on souhaite obtenir de lui. C'est l'auto-soumission qui fait la force d'une telle gouvernementalité.

2. QUE RETENIR DE CET EXEMPLE ?

La première leçon de cet exemple est la dimension « irrésistible » des palmarès dans le monde libéral — dans la mesure où il y a toujours des acteurs intéressés à les produire et les vendre, et des acteurs puissants (ici

¹ B. MAS, F. PIERRU, N. SMOLSKI & R. TORRIELLI, *L'Hôpital en réanimation*, Paris, Éditions du Croquant, 2011.

l'administration de l'Assistance publique) qui poursuivent des objectifs et sont en mesure de peser sur les choix finaux.

La seconde leçon réside dans la dynamique propre à l'espace public, qui est centrale au succès opératoire du palmarès-benchmark : ce qui fait l'efficacité normative des indicateurs et des classements, c'est la diffusion et la circulation libres de l'information (personne n'impose ses choix à *Science et Avenir*, et c'est en cela que c'est libéral).

La troisième leçon tient à la difficulté pour les acteurs individuels et institutionnels à échapper au *benchmark* — car il est souvent suicidaire de ne pas en tenir compte. S'impose ainsi à tous une tyrannie de la (pseudo-) transparence ; chacun doit rendre des comptes (chiffrés) sur ce qu'il fait et dès lors s'efforcer de pouvoir présenter un bilan compétitif à l'égard de ses concurrents.

La quatrième et dernière leçon, essentielle, est l'effacement progressif des protections qui, d'après Benjamin Constant, font la « liberté » des modernes et la démocratie. À savoir l'autonomie partielle du politique par rapport à l'économique, de l'espace personnel par rapport à la culture d'entreprise ; à savoir encore la séparation entre devoir public et exigence du collectif, et les droits à l'autonomie individuelle.

3. COMMENT MENER LE TRAVAIL CRITIQUE ?

En réaction à ces phénomènes, deux pistes parmi d'autres sont ici suggérées.

La première consiste à reconnaître l'importance de la langue dans la gouvernementalité et à mettre en évidence le nouvel univers non seulement rhétorique, mais aussi sémantique et grammatical, qui régit les discours actuels — bref à souligner le « poids des mots » qui disent l'évidence de notre monde : « évaluation », « excellence », mais aussi « gouvernance », « société civile »... tous termes qui sont à la fois descriptifs et normatifs. Il convient d'être attentif aux connotations que les expressions et formules employées véhiculent, aux effets de sens induits, à leurs biais systématiques². Dans la mesure où, contrairement à un système autoritaire, la

² Voir l'étude récente de Franco Moretti et Dominique Pestre sur la langue de la Banque Mondiale, « Bankspeak, The Language of World Bank Reports », in *New Left Review*, 92, Mar/Apr 2015, p. 75-99. Il est par exemple frappant de voir comment le rapport au temps se modifie dans les rapports de la Banque Mondiale. À partir des années 1990, les rapports font de moins en moins usage des adverbes de temps et de verbes au temps passés, qui est le temps de la reprise et du

nouvelle gouvernementalité adopte une stratégie plus *soft* de « conduite des conduites » (Foucault), son influence passe par la mobilisation de la subjectivité des acteurs (leurs désirs, leur sens du devoir...) et cela suppose toute une manière verbale de dire les choses.

Une autre stratégie consiste à faire la généalogie des dispositifs (au sens foucauldien), c'est-à-dire à montrer comment les outils, les concepts, les normes finalement retenues se sont progressivement mis en place et, devenus finalement incontournables, se sont mis à structurer nos actions. Il s'agit ainsi de montrer ce qui a rendu ces notions évidentes (au point d'effacer la contingence même de leur construction) mais aussi ce qui les a rendu suffisamment ambiguës et contraignantes pour être reprises par une multitude d'institutions et individus (ce sont à cet égard des « mots frontières »).

Ce travail de généalogie, qui peut être mené pour le cas des termes et outils de l'évaluation, sera ici esquissé, au titre de simple exemple, pour le cas de la catégorie de « société civile ».

L'évidence actuelle de cette catégorie — la société civile existe — contraste avec la manière dont on parlait il y a encore 40 ans. Le concept de « société civile », qu'utilisait Tocqueville en 1820 mais qui avait disparu depuis les années 1870, a été ré-exhumé dans la période récente, dans des sens et usages divers et pour partie contradictoires.

Une première origine du retour en grâce de cette catégorie se trouve dans l'Amérique latine et dans l'Europe de l'Est des années 1970 et 1980. Elle indique une dissidence sociale à l'égard des pouvoirs militaires et/ou totalitaires, la dissidence de mouvements sociaux qui ne revendiquent pas tant le pouvoir pour eux-mêmes que le respect d'un certain nombre de droits et libertés. L'idée est celle d'affirmer des droits civils contre la violence. Dans ces deux cas, il vaut de noter que les Églises jouent un rôle de ferment essentiel dans la (re-)naissance de la notion de « société civile ».

Une seconde origine, qui date également des années 1970, réside dans la défense de l'individu contre l'État (et notamment l'État-providence) par certains *think tanks* libéraux, mais aussi, dans la décennie suivante, dans certains combats d'ONG pour des causes sociétales (environnement, égalité des genres, lutte contre le sida, associations...) à l'égard desquelles l'État

bilan, et de plus en plus de la forme grammaticale du présent continu (forme en « -ing »), qui marque la projection performative dans le futur sans que jamais on ne soit en mesure de saisir comment on tire des leçons du passé.

paraît trop passif. L'idée est ici celle d'une prise en main par les acteurs sociaux eux-mêmes de problématiques délaissées par l'État.

Une troisième origine, plus tardive (au tournant des années 1980/90), tient à la reconnaissance par le monde des affaires qu'il convient qu'il se saisisse lui-même des questions d'environnement (après une opposition plutôt directe dans les années 80), parce que le monde a changé et qu'il est devenu « *green and participative* » (ce qu'affirment explicitement les consultants des entreprises autant que les hommes politiques en campagne). Le monde des entreprises qui, avec l'effondrement du communisme, se sent investi de la responsabilité de diriger le monde, va promouvoir la « bonne gouvernance », en lien avec la « société civile » (les associations, les organismes internationaux comme la Banque Mondiale...) sur une base libérale (économiquement et politiquement). Les États et leurs normes strictes sont marginalisés au profit d'un *soft power* plus globalisé (avec son pôle finance et son pôle gouvernance, et entre les deux le rôle clé du management — voir l'article de Moretti et Pestre pour une démonstration) fondé sur la responsabilité et l'auto-régulation (des entreprises). C'est d'ailleurs de cette époque que date l'apparition des termes de « gouvernance » et de « civil » dans le vocabulaire de la Banque Mondiale.

Aujourd'hui, l'évidence que la « société civile » existe tient notamment au fait que les sens de cette expression varient et incluent notamment tout à la fois le nouvel ordre (économique) du monde *et* sa contestation. On pourrait dire que c'est sur ce même type d'ambiguïté que reposent les notions de développement durable, de modernisation écologique, de technologies vertes, mais aussi de participation, de transparence et bien sûr, comme nous l'avons montré, d'évaluation ou d'excellence.

En montrant ce que ces constructions sociales effacent (notamment les différences sociales) et ce qu'elles produisent (du consentement inévitable si l'on peut dire), en soulignant aussi qui utilise systématiquement ces notions et qui ne les utilise pas ou peu, ce type d'analyse permet de mettre leur évidence à distance et de s'en défier.

Évaluer la recherche ?

ALAIN ERALY

Professeur à l'Université libre de Bruxelles
Membre de l'Académie royale de Belgique

Le discours de l'évaluation pénètre désormais toutes les sphères de la vie professionnelle. Évaluation des managers, des fonctionnaires, des enseignants, des élèves, des stratégies d'entreprise, des politiques publiques, des hôtels, des restaurants, des villes, de la qualité des produits, de la qualité des soins dans les hôpitaux, de la performance des universités, de l'école, de la justice, de la police : partout la même incitation à rendre des comptes et s'exposer à des sanctions. L'affaire est entendue : l'évaluation n'est pas un simple dispositif de gestion des ressources, elle sert à motiver, canaliser les efforts, susciter la réflexivité, favoriser le développement personnel et collectif ; elle n'est pas seulement un instrument de contrôle, mais une exigence proprement morale et démocratique de transparence, d'équité et de responsabilité.

Dans le champ de la recherche scientifique, l'évaluation relève aujourd'hui de l'évidence naturelle. Comment concevoir la science sans la transparence ? Comment imaginer de soustraire des chercheurs et des équipes à toute forme d'*accountability* ? Et comment allouer les crédits de recherche, les postes de chercheurs, les promotions sinon en évaluant les projets et les candidats ? Il n'y a pas le choix : plus rares les ressources publiques au regard des candidats en lice, plus âpre la compétition entre eux et plus impérieuse la nécessité de disposer de critères objectifs pour les départager. Au demeurant, l'évaluation de la recherche a toujours existé, elle appartient en propre à la logique universitaire. De tous temps, les chercheurs ont fait l'objet d'un triple examen : examen de la qualité de leurs travaux (livres, articles, conférences, thèses de doctorat, etc.), examen des projets de recherche qu'ils soumettent

aux organismes de financement de la recherche, examen de leur contribution plus générale aux trois missions de l'université lors des nominations et des promotions. Ce qui change aujourd'hui, assurent les prosélytes, c'est juste que l'analyse *quantitative* de la recherche scientifique (bibliométrie, scientométrie, infométrie), jointe au recours systématique à des jurys et des experts étrangers, confère à l'évaluation un caractère beaucoup plus rigoureux. Le talent et le mérite sont devenus plus faciles à mesurer, la médiocrité et la flemmardise plus difficiles à cacher. Il est très difficile de s'opposer publiquement à l'évaluation : toute résistance est en soi un aveu, elle ne manque pas de se retourner contre son auteur immanquablement soupçonné de chercher à dissimuler ses propres carences : « Notre dispositif est sans doute imparfait, mais il a le mérite d'exister. Si vous n'avez rien à cacher, que craignez-vous ? » Et chacun d'évoquer l'image de tel chercheur improductif qui s'est glissé entre les mailles du filet méritocratique en profitant du laxisme d'une commission, du soutien d'un collègue influent ou d'un rapport de force momentané. L'évaluation est un *opérateur de légitimité* : elle est prise tout entière dans un imaginaire normatif qui lui donne son sens et confère au chercheur aussi bien qu'à l'institution universitaire leur justification sociale.

Il n'est pas question de remettre en cause les fonctions *théoriques* de l'évaluation de la recherche : allocation des ressources rares, contrôle de leur bon usage, arbitrage au sein des commissions, équité, motivation et responsabilisation des chercheurs et des équipes. Toutes ces fonctions sont *a priori* légitimes et nécessaires. La question reste toutefois de savoir dans quelle mesure l'évaluation à base d'indicateurs bibliométriques telle qu'elle est pratiquée remplit effectivement ces fonctions et si, à côté des conséquences *attendues*, on n'observe pas des conséquences *inattendues* qui finissent parfois par obérer l'utilité des dispositifs. Je passe sur les limitations techniques et méthodologiques de la bibliométrie : tendance à restreindre l'évaluation à ce qui est quantifiable, tendance corrélative à remplacer la qualité par la quantité, emprise du marché de l'édition (un article de vingt pages dans une revue classée pèse plus lourd qu'une thèse de cinq cents pages assortie d'un rapport de thèse élogieux), stratégies de réseaux et de citations, primauté de la méthodologie sur l'innovation théorique, privilège exorbitant accordé à l'anglais et à la culture anglo-saxonne, hyperspécialisation des articles, dévalorisation du livre, mise au ban de la « grande théorie », pénalisation des travaux d'érudition et des analyses qualitatives, etc. Ces déficiences sont

souvent mentionnées par les experts¹, plus rarement suivies d'effets dans les commissions scientifiques. J'aimerais quant à moi évoquer d'autres types de difficultés qui sont moins souvent analysées – pour ne pas dire carrément occultées.

LA PRODUCTIVITÉ MESURE-T-ELLE LA CONTRIBUTION À LA CONNAISSANCE ?

Toute évaluation est supposée mesurer une contribution objective à des finalités plus générales. L'évaluation de la qualité des soins dans un hôpital mesure en principe sa contribution à la santé publique, l'évaluation d'un département ministériel sa contribution aux objectifs d'efficacité et d'efficience publiques, l'évaluation d'une voiture son utilité pour l'acheteur, et ainsi de suite. Qu'en est-il de l'évaluation de la recherche ? À quel genre de finalités se réfère-t-elle ? La réponse officielle, celle qui fait globalement consensus, consiste à dire : l'évaluation mesure la contribution du chercheur à la connaissance. Or, l'observation la plus superficielle du fonctionnement des commissions d'évaluation montre que cette finalité est somme toute assez peu prise en compte : les indicateurs comparatifs de *productivité* — nombre d'articles dans des revues classées, nombre de citations, nombre de communications et de conférences, séjours à l'étranger, etc. — pèsent bien plus lourd que la contribution intrinsèque du chercheur à la connaissance. On dira certes que ces indicateurs quantitatifs mesurent *indirectement* cette contribution, mais outre que cet optimisme est largement remis en cause par les études sur l'usage des indicateurs bibliométriques, le fait n'en demeure pas moins : les commissions d'évaluation évaluent la productivité des chercheurs plus souvent que leur contribution à la connaissance et elles tendent à déléguer cette responsabilité dernière au marché de l'édition scientifique en partant

¹ Voir par exemple Y. GINGRAS, *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*, Paris, Raisons d'Agir, 2014 ; A.-W. HARZING & N. ADLER, « When Knowledge Wins : Transcending the sense and nonsense of academic rankings », in *The Academy of Management Learning & Education*, vol. 8, no. 1, 2009 ; L. SEGALAT, *La science à bout de souffle ?* Paris, Seuil, 2009 ; J. GLÄSER & G. LAUDEL, « Advantages and Dangers of "Remote" Peer Evaluation », in *Research Evaluation*, 14(3), dec. 2005 ; M. OSTERLOH & Br. S. FREY, « Research Governance in Academia : Are there Alternatives to Academic Rankings ? », in *CES IFO Working Paper n° 2797*, sept. 2009 ; A. ERALY, « Les enjeux de l'évaluation. Du discours aux pratiques », in *L'évaluation de la recherche en sciences humaines et sociales. Regards de chercheurs*, Louvain-la-Neuve, Bruylant-Academia, 2011, p. 15-36.

de l'axiome que ce qui est publié dans une bonne revue ou dans une maison d'édition reconnue est valorisable et que ce qui n'est pas publié ne vaut rien. Deux petits articles redondants rédigés à la hâte dans le confort de son bureau pèsent ainsi plus lourd qu'une vaste étude ethnographique, fruit de plusieurs années d'immersion dans un village indonésien, mais qui peine à trouver un éditeur. Très nombreux sont aujourd'hui les académiques qui ont l'impression de n'être pas évalués sur la *réalité* de leur production scientifique, en particulier sur la valeur de leur travail empirique et de leur contribution théorique.

Plus généralement, il est frappant d'observer la facilité avec laquelle les universités ont investi ce critère de la productivité scientifique en faisant l'économie d'un débat critique sur les finalités de l'université, et par exemple sur le rôle de la recherche appliquée et la contribution de la recherche au développement social et économique. À quoi sert l'université ? Quel genre de biens publics les universités doivent-elles produire en contrepartie de leur financement ? La compétition bibliométrique, les rankings, l'inflation de revues et d'articles au contenu similaire : ces évolutions sont-elles nécessairement conformes au bien public ? Sont-ce là ce qu'une communauté démocratique est en droit d'attendre d'une institution universitaire ? Ces questions appellent assurément des réponses nuancées ; il est malheureux qu'elles soient si rarement mises en débat. Comme si la position concurrentielle tendait à devenir la justification nécessaire et suffisante de toute politique universitaire.

LES CHIFFRES NOURRISSENT-ILS LE DÉBAT OU TENDENT-ILS À S'Y SUBSTITUER ?

Une discipline n'est pas l'autre, certaines demeurent ancrées dans la tradition du livre lorsque d'autres priorisent les articles ; certaines reposent sur des méthodes quantitatives, d'autres sur des méthodes qualitatives ; certaines recherches requièrent un travail d'équipe et admettent, même pour des articles mineurs, la signature de très nombreux auteurs lorsque d'autres se mènent plus solitairement et débouchent sur la publication d'ouvrages à titre de seul auteur ; dans certains domaines, les bases de données existent déjà et l'activité de publication peut commencer rapidement tandis qu'ailleurs, le matériau empirique doit être rassemblé au fil d'un processus d'enquête qui peut s'étendre sur des années. Devant l'évidence de cette hétérogénéité du

champ scientifique, les experts se plaisent à souligner la prudence qui doit présider au maniement des critères d'évaluation de la productivité scientifique ; il n'est pas question, insistent-ils, de réduire la performance d'un chercheur à une poignée de statistiques ; les indicateurs chiffrés apportent des éléments d'information utiles, et même nécessaires, pour autant qu'ils soient contextualisés.

Or, cette prudence est loin d'être toujours de mise au sein des commissions. On doit déplorer, à cet égard, la rareté des études empiriques sur la psychologie des évaluateurs et le fonctionnement *concret* des commissions. En pratique, l'indicateur chiffré est un formidable absorbeur de tensions, un réducteur de polémique et de suspicion, mais avant tout une source d'économie d'attention, de réflexion et de débat. Les commissions couvrant généralement plusieurs disciplines, il est fréquent que les évaluateurs se retrouvent en situation d'incompétence — par manque de connaissance du domaine ou par manque de temps. Incapables, ou non désireux, d'évaluer la qualité intrinsèque des articles et des ouvrages mentionnés dans les rapports des candidats, ils préfèrent s'en tenir à des critères chiffrés supposément plus objectifs, donc moins polémiques, tout juste comme on avait tendance, dans les bureaucraties, à privilégier des critères formels indiscutables comme l'ancienneté pour octroyer les augmentations barémiques.

Dans les faits, les indicateurs chiffrés semblent d'abord servir à faire l'économie de longs débats conflictuels : ils ont pour fonction tacite, dirait Bruno Latour, de *vider l'agora*. Les chiffres devraient nourrir le débat rationnel et fonder l'argumentation : trop souvent, ils se substituent au débat et tiennent lieu d'arguments. De fait, comment justifier que les critères soient si rarement mis en débat au sein des disciplines et des départements ? Comment justifier que dans certaines commissions, les candidats ne reçoivent même pas la possibilité d'être auditionné afin de défendre leur projet et leur bilan ? Comment justifier que des carrières entières puissent se jouer sans donner la parole aux intéressés ? Comment justifier que certaines commissions s'abstiennent de justifier leurs décisions ? On trouve normal qu'un enseignant organise une visite des copies et qu'il consacre plusieurs heures à un étudiant dans le cadre de la supervision de son travail de fin d'études — et l'on trouve inutile d'auditionner un candidat pour l'octroi d'une bourse de recherche de plusieurs années et de lui donner un feed-back systématique sur son projet de recherche... C'est une chose de discuter de la validité formelle des indicateurs,

c'en est une autre d'examiner comment ils sont utilisés *in situ* au sein des commissions.

LES CHIFFRES SONT-ILS SOURCE D'ÉQUITÉ OU D'INIQUITÉ ?

Les dispositifs d'évaluation sont souvent mis en place au nom d'un principe général d'équité. Il n'est pas juste, trouve-t-on, de traiter semblablement un chercheur très productif et un autre qui reste à la traîne. L'évaluation vise précisément à instaurer un juste rapport entre les contributions et les rétributions. Sans du tout contester cette valeur d'équité, non plus que le raisonnement qui la sous-tend, force est de constater qu'en pratique, encore une fois, l'évaluation chiffrée de la recherche produit de l'iniquité aussi bien que de l'équité, et cela pour une raison simple : l'inégalité des conditions de départ. Si l'évaluation portait sur la totalité du travail des chercheurs, c'est-à-dire sur leur contribution à l'ensemble des missions de l'université (recherche, enseignement, gestion de l'université, formation continue, coopération, etc.), il serait possible sans doute de prétendre à une équité générale. Mais cette prise en compte de la globalité des activités prestées est l'exception plutôt que la règle. Dans beaucoup de commissions scientifiques, les experts privilégient la seule production scientifique et ils statuent sans trop se soucier de la réalité des charges assumées par ailleurs par les candidats — pour ne rien dire des contraintes financières et familiales qui peuvent handicaper un chercheur et l'empêcher par exemple de séjourner dans des universités étrangères. Les experts sont invités à se prononcer sur des dossiers de chercheurs ou d'équipes de recherche en faisant abstraction des conditions réelles de recherche, c'est-à-dire indépendamment des charges effectivement assumées par les candidats, de leurs mérites propres, des contraintes exercées par leur institution.

Cette décontextualisation de l'évaluation génère un fréquent sentiment d'irrespect et d'injustice parmi les chercheurs et les académiques. Beaucoup d'entre eux vivent, jour après jour, cette expérience singulière de se trouver submergé de tâches pédagogiques et administratives en sachant qu'ils seront *in fine* évalués sur une performance scientifique dont la bureaucratie universitaire ne cesse de les détourner et qu'ils finissent par devoir loger dans leurs moments de loisir au détriment de leur vie familiale. La situation est encore plus frustrante lorsqu'ils se retrouvent en compétition avec des candidats

nationaux ou internationaux qui ont bénéficié de conditions de recherche (temps disponible, encadrement scientifique, accès à des financements) beaucoup plus favorables.

Cette décontextualisation engendre par suite une tension croissante entre les missions de l'université et cela d'autant plus que l'intégration des différentes missions n'est assumée clairement ni au niveau central ni au niveau des facultés ou des départements. Je veux dire que les contraintes de productivité scientifique, de productivité pédagogique et de productivité administrative émanent d'instances relativement autonomes les unes des autres, sans véritable coordination entre elles, ce qui peut conduire à des contradictions manifestes : le même académique est invité à accroître le nombre et la qualité de ses publications, à attirer plus de doctorants, à décrocher davantage de contrats de recherche, à renouveler ses cours et ses méthodes d'enseignement, à participer aux commissions pédagogiques, à lancer de nouvelles filières et des MOOC, à prendre la direction d'un programme ou d'un département, s'engager dans un projet de développement, à assumer la direction d'un maximum de travaux de fin d'études, et tout cela en multipliant les séjours et les congrès internationaux. Il en résulte une tension croissante entre les trois missions, tension vécue parfois douloureusement par les acteurs.

Le choix d'une université de privilégier, donc d'évaluer prioritairement la recherche est bien sûr légitime, et même enthousiasmant ; ce qui l'est moins, c'est l'art de l'ambiguïté qui consiste à éviter toute espèce d'arbitrage explicite entre les trois missions en donnant à croire qu'il est possible, à niveau constant de ressources, de viser l'excellence dans tous les domaines, et en particulier d'accroître substantiellement la production scientifique sans du tout affecter les deux autres missions. Un peu partout, on observe que la pression en faveur de la recherche induit une perte relative de légitimité de la mission d'enseignement alors même que le nombre d'étudiants grandit et que les lacunes de formation des étudiants en première année exigent un encadrement renforcé. Est-ce vraiment une évolution souhaitable ?

ÉVALUE-T-ON DES CHERCHEURS OU LES PROPRIÉTÉS D'UN RÉSEAU ?

Toute évaluation est un jugement de valeur et induit un faisceau d'attentes normatives. La simple présence de rubriques à remplir (séjours à l'étranger,

articles dans des revues classées, communications dans des colloques internationaux, etc.) génère une représentation implicite de ce qu'est un « bon » et un « mauvais » chercheur. Or, on vient de le voir, chaque chercheur est désormais aux prises avec un faisceau grandissant de telles injonctions et ces injonctions sont au moins partiellement contradictoires. Puisqu'il est impossible de satisfaire à l'ensemble des exigences, l'académique se trouve en quelque sorte confronté à la fatalité de sa propre déficience, il est condamné par l'institution à décevoir cette même institution. Le système génère un déficit mécanique de reconnaissance qui ne compte pas pour peu dans l'acrimonie ou le cynisme de certains collègues. On ne parle guère de cette perte d'estime de soi, de ce *syndrome de l'imposteur* que traînent discrètement un nombre croissant de collègues et qui les conduit à vivre en permanence dans la perception d'un décalage entre les paillettes de l'imaginaire académique et le concret prosaïque de leur activité.

En vain les décideurs prennent-ils soin de répéter qu'il n'est pas question de juger les personnes elles-mêmes, mais seulement leur productivité : il est impossible en pratique d'échapper aux enjeux de reconnaissance. L'usage des indicateurs chiffrés conduit à une forme de réification de soi, ou pour mieux dire de « fétichisme », qui mériterait une analyse approfondie. Marx appelle « fétichisme de la marchandise » le processus par lequel un produit issu du travail concret, sitôt qu'il reçoit un prix sur le marché, c'est-à-dire qu'il devient marchandise, acquiert une valeur d'échange. Le prix, qui n'est rien que la résultante d'un rapport social déterminé, en vient à se fixer dans l'objet même à la façon d'une propriété objective. Voici que la marchandise semble posséder une « valeur » comme elle possède une couleur, un poids, une texture : telle armoire *vaut* deux cents euros, c'est là sa valeur propre. En réalité, insiste Marx, la valeur n'est rien qu'une abstraction dont l'apparente objectivité dissimule l'objectivité des rapports sociaux de production et d'échange. Le fétichisme de la marchandise signale une perte du lien généalogique : une propriété émerge d'un ensemble complexe d'interactions et se trouve fictivement attribuée à la chose elle-même.

Ne peut-on en dire autant des indicateurs chiffrés de productivité scientifique ? En dernière analyse, un indicateur chiffré tel qu'un *impact factor* n'est rien qu'une *position dans un réseau scientifique* ; il ne qualifie pas une personne, mais seulement le référencement d'un ensemble de travaux au sein d'une collectivité ; en toute rigueur, il est donc une propriété du réseau et non

une qualité intrinsèque de la personne — même s'il est vrai que le travail de cette personne n'est pas sans influence sur le référencement. Or, le chiffre tend à produire sa propre fiction : l'indicateur chiffré en vient vite à mesurer la valeur du chercheur lui-même tout juste comme la moyenne des notes attribuée à un étudiant est tenue pour un indicateur de sa valeur intrinsèque et non pour un indicateur des pratiques d'évaluation en usage dans l'institution. À l'arrivée, l'évaluation produit une hiérarchie de *distinction* académique qui alimente fierté chez les uns, sentiment d'aliénation et d'échec et retrait défensif chez d'autres.

DE LA COMPARAISON À LA COMPÉTITION : LE GLISSEMENT EST-IL INÉLUCTABLE ?

Mais l'essentiel n'est sans doute pas là. L'usage des indicateurs chiffrés offre une échelle de comparaison entre chercheurs et entre institutions. Or, de la comparaison au classement et du classement à la compétition, la pente est naturelle. La pratique des indicateurs chiffrés s'inscrit dans un cadre plus général, celui de la légitimation de la concurrence nationale et internationale dès lors qu'elle est associée à la construction européenne, à la mobilité des étudiants et des chercheurs et à l'excellence académique. La compétition dont il est ici question n'a pas pour premier effet de projeter les universités dans l'univers de l'économie marchande, mais d'abord de les river les unes aux autres dans l'hypnose d'une compétition artificielle, largement fallacieuse, fondée sur le nombre d'articles dans des revues classées, le nombre de doctorants, le nombre de citations, le nombre de prix Nobel obtenus, etc. Telle est la source première de l'instrumentalisation du savoir : l'inscription de la recherche dans une logique de compétition académique étrangère à cette recherche même — et bien souvent étrangère à la qualité intrinsèque de cette recherche.

Ensuite, il faut souligner les boucles de rétroaction nombreuses qui opèrent entre les différents processus. Par exemple, la formalisation des indicateurs bibliométriques est à la fois une condition de possibilité et une conséquence de la compétition : la formalisation autorise la comparaison internationale tandis qu'en retour celle-ci favorise un surcroît de formalisation. Le fait de tenir les étudiants pour des clients finit par les amener à se regarder comme tels et à se comporter en conséquence. Le fait, pour les écoles de gestion, de se soumettre à un ranking aussi contestable — sur un plan intellectuel aussi

bien qu'idéologique — que celui du *Financial Time* ou celui de *Business Week* et de fournir à ces journaux les statistiques demandées, la liste des adresses mail des anciens diplômés, etc. contribue à étendre la portée de ces rankings, à confirmer la croyance sociale en leur validité, donc la probabilité que les étudiants y fassent référence pour faire leur choix ; cela renforce du même coup la réputation des rankings et plus généralement leur pouvoir de créer un marché de l'enseignement supérieur².

L'évaluation de la recherche n'est pas neutre ; en aucun cas, elle ne saurait se comprendre comme une simple représentation du volume de connaissances générées par chaque université. Elle contribue tout au contraire à déterminer, structurer, formater la recherche en fonction des critères retenus. Ces critères, il faut le remarquer, échappent très largement aux universités, ils définissent les règles de la compétition internationale qui s'impose désormais à ces dernières. Avec l'extension du champ de la comparaison quantifiée se développent des organismes extérieurs aux universités — généralement étrangers à toute espèce de contrôle démocratique — qui tendent à dicter leurs critères et classements. Et comme dans toute compétition, le système tend à avantager les acteurs dominants, il fascine les médias et le grand public au point de faire oublier tous les laissés-pour-compte, tous ceux qui n'ont pas les moyens — en particulier les moyens financiers — de se hisser dans les classements, et d'abord les universités du Sud.

L'idée d'une concurrence académique s'impose donc en même temps que l'idée d'un marché de l'enseignement et de la recherche. En privé, les décideurs admettent volontiers que les rankings sont une forme de scandale intellectuel, que la compétition entre universités est largement fictive et même dénuée de sens, que l'on compare en permanence des institutions incomparables en termes de cadre réglementaire, de contraintes financières, de politiques publiques, d'autonomie, etc. Mais aucun décideur ne se hasarderait à retirer son institution du jeu au risque de la priver d'une partie de son attractivité. Dans l'univers de la concurrence universitaire, le critère de la productivité scientifique tend ainsi à supplanter l'idéal de la connaissance.

² Ce qui était par exemple l'objectif explicite de John Byrne, directeur de *Business Week*, lorsqu'il a lancé le ranking des MBA. Voir à ce sujet : D. C. Bok, *Universities in the Marketplace: The Commercialization of Higher Education*, Princeton, Princeton University Press, 2003.

Du Bon usage des indicateurs

Critères de validité, échelles et disciplines

YVES GINGRAS

Professeur titulaire de la Chaire de recherche
du Canada en histoire et sociologie des sciences
à l'Université du Québec à Montréal (UQAM)

Le monde académique connaît aujourd'hui, en Europe plus encore qu'en Amérique du nord, une « fièvre de l'évaluation », marquée par une course aux indicateurs et aux classements. Or, si l'évaluation de la recherche par les pairs est un phénomène bien accepté et d'ailleurs assez ancien (dès le XVII^e siècle, les publications sont systématiquement évaluées avant parution dans des organes comme les *Transactions of the Royal Society of London* et, au début du XX^e siècle, les organismes de financement de la recherche font déjà systématiquement évaluer les projets de recherches qui leur sont soumis), elle est aujourd'hui non seulement multipliée à de nombreux niveaux (publications, projets, individus, équipes, institutions, programmes...) mais fait en outre place à une évaluation essentiellement quantitative ainsi qu'à des classements. Or, entre évaluation d'article, évaluation de chercheur et évaluation d'université, mais aussi entre évaluations qualitative et quantitative, ou encore entre évaluation et classement, il y a des différences sensibles d'objectifs et de méthodes, qu'il convient absolument de mettre en évidence et de garder à l'esprit sous peine de commettre des erreurs monumentales.

1. GÉNÉALOGIE DES INDICATEURS

À cet égard, une première part du travail doit consister à retracer la généalogie des instruments qui servent aujourd'hui à l'évaluation de la recherche. C'est ainsi, par exemple, qu'on verra que le Facteur d'impact fut initialement créé (en 1975, par Eugene Garfield) pour éclairer les bibliothèques sur les choix à opérer en matière d'acquisition de revues. Il ne s'agissait aucunement d'évaluer les *articles* et encore moins les *chercheurs*, mais bien les *revues*. Dès lors qu'il s'agit de se servir de cet indicateur pour évaluer l'influence scientifique d'un article, on se confronte à d'importants problèmes méthodologiques. D'une part, le Facteur d'impact moyen d'une revue n'est pas transposable à chacun des textes qui y sont publiés. D'autre part, la temporalité de la recherche varie fortement d'une discipline à l'autre ; dans les sciences sociales, la demi-vie moyenne d'un article (temps nécessaire à ce qu'il recueille la moitié des citations qu'il suscitera *in fine*) est sensiblement plus grande qu'en sciences de la nature, de sorte qu'il faudrait en principe calculer les Facteurs d'impact sur des périodes nettement plus longues que les deux ans qui sont aujourd'hui retenus pour des raisons pratiques (simplicité, rapidité, coût...).

En soi, le travail généalogique permet déjà de montrer que bon nombre d'instruments sont utilisés aujourd'hui à des tâches très différentes de celles pour lesquelles ils avaient été conçus et auxquelles ils étaient adaptés.

Dans cette perspective généalogique, il convient aussi de repérer certains usages stratégiques particuliers qui ont permis à certains instruments de gagner leurs fonctions actuelles. Ainsi, l'importance accordée à un classement d'universités comme celui de Shanghai, initialement mis au point dans une université chinoise inconnue en dehors du mur de Chine, ne peut se comprendre que si on tient compte du fait que sa publication a servi l'agenda politique d'un certain nombre de dirigeants européens qui entendaient précisément réformer le monde universitaire et se sont servis du classement pour justifier l'urgence de leurs réformes (par exemple en France pour mettre en place la loi LRU sur l'autonomie des universités). De même, l'imposition de certains indicateurs bibliométriques dans des comités d'évaluation est souvent le fait de chercheurs particuliers, qui, comme ce fut le cas en économie, trouvent là une opportunité de faire triompher certaines manières de faire science ou certains paradigmes.

2. EFFETS DES INDICATEURS

Une seconde et importante part du travail critique consiste à examiner dans le détail les effets concrets de l'usage des différents instruments. Or, à cet égard, force est de constater, études empiriques à la main, que cet usage engendre d'importantes modifications des comportements des individus et des institutions, modifications que l'on doit sans aucun doute qualifier d'« effets pervers ». Ainsi en va-t-il par exemple de l'usage stratégique des citations (pour favoriser les indicateurs d'alliés et/ou défavoriser ceux de concurrents), du choix des objets de recherches (au profit de ce qui est le plus « vendable » auprès des revues bien classées) et du délaissement des préoccupations plus « locales » (qui peuvent plus difficilement intéresser de manière récurrente la communauté internationale¹). Du côté des institutions, on a pu observer des stratégies aussi éthiquement contestables que l'achat de publications (rémunération de l'auteur contre simple mention par lui d'une affiliation universitaire supplémentaire lors de la signature de ses articles) par des universités qui souhaitaient monter rapidement dans les classements².

3. CRITÈRES DE VALIDITÉ DES INDICATEURS

Une troisième part, essentielle, du travail critique réside évidemment dans l'étude de la validité des indicateurs utilisés et de la légitimité des prétentions à leur faire jouer tel ou tel rôle dans l'évaluation.

¹ Dans les années 1960-1970, les économistes canadiens, qui étudiaient jusqu'alors des problématiques locales et pratiques de l'économie canadienne (*policy papers*), ont, depuis les années 1990, pour publier dans les revues américaines, commencé à travailler sur des modèles théoriques universels. De la même manière, il est plus facile, en sociologie, de publier sur la scène internationale une série d'articles comparant Bourdieu et Luhmann plutôt qu'une série d'études sur l'électorat lepéniste dans quelques départements du sud de la France (sujet qu'on peut sans doute publier aux USA une fois mais pas tous les six mois). La chose est notamment sensible dans les sciences sociales, mais pas seulement ; sur les différentiels d'internationalité selon les disciplines, voir Yves GINGRAS, « Les formes spécifiques de l'internationalité du champ scientifique », in *Actes de la recherche en sciences sociales*, n^{os} 141-142, mars 2002, p. 31-45.

² Yves GINGRAS, « How to boost your university up the rankings », in *University World News*, 18 July 2014, p. 329. Disponible sur <http://www.universityworldnews.com/article.php?story=20140715142345754>

À cet égard, les « Principes de Berlin » définis en 2006 par l’(autoproclamé) « International Ranking Expert Group » énoncent un certain nombre d’exigences que devraient satisfaire les classements internationaux :

- définir clairement les buts visés ;
- être transparents dans leur méthodologie ;
- choisir des indicateurs adéquats et valides ;
- identifier clairement (et de manière stable) la pondération de chaque indicateur ;
- reconnaître la diversité des institutions.

Or, il est clair que ces exigences — qui passent par ailleurs sous silence l’importante problématique de la diversité disciplinaire — sont loin d’être satisfaites par le classement de Shanghai et par d’autres classements comparables.

D’une manière plus générale, un bon indicateur doit :

- être adéquat à son objet (mesurer ce qu’il prétend mesurer) ;
- être homogène dans sa mesure ;
- varier en conformité avec l’inertie propre de l’objet (un thermomètre qui, sans raison apparente, enregistre un brusque changement de la température d’une pièce est sans doute mal configuré ou défectueux) ;
- soutenir une relation monotone croissante entre le concept mesuré et l’indicateur choisi.

Or, une fois encore, nombre d’indicateurs actuellement utilisés dans l’évaluation de la recherche sont, au regard de ces exigences, totalement défaillants. Pour ne prendre que quelques exemples :

1/ l’indice h (h-index), qui est censé mesurer à la fois la production scientifique et sa qualité (mesurée d’après le nombre de citations que les publications ont suscitées), est corrélé à 0.9 avec le nombre de publications. C’est donc essentiellement la production scientifique qu’il mesure. Et il suffit de voir qu’un chercheur ayant publié 10 articles cités 10 fois a un h-index double d’un chercheur ayant publié 5 articles cités 60 fois pour voir que l’indice h ne mesure pas ce qu’il prétend mesurer. On a vu plus haut qu’il en est de même du Facteur d’impact, qui ne mesure pas les articles mais les revues.

2/ les classements d’université, qui sont constitués à partir de scores calculés par la somme pondérée d’indicateurs multiples et hétérogènes, ne sont clairement pas homogènes dans leur mesure. On ne peut donc pas y voir

des indicateurs valides de la qualité d'une institution, alors qu'ils reflètent des données multidimensionnelles (enseignement, recherche...) et multidisciplinaires qui ne sont pas uniformes : telle université globalement bien classée peut, dans le domaine de l'histoire, dispenser une formation nettement moins bonne ou mener une recherche nettement moins dynamique qu'une autre globalement moins bien classée. De plus, la pondération des variables est arbitraire et il est connu qu'il suffit de la modifier pour modifier les résultats des classements.

3/ lorsqu'ils constatent des progressions ou au contraire des dégringolades fulgurantes d'universités d'une année à l'autre, les classements sont manifestement mal configurés (ou font l'objet de stratégies d'adaptation contestables cf. *supra*). Les universités sont de gros bateaux inertiels, dont le cap et la position ne peuvent pas se modifier de manière spectaculaire en seulement deux années. Or, ces classements sont annuels et montrent des variations importantes d'une année à l'autre, preuve de l'invalidité de leurs mesures.

4/ si le nombre d'étudiants étrangers ou celui de professeurs étrangers peut, jusqu'à un certain point, donner une indication de l'ouverture internationale d'une université, ce n'est plus le cas au-delà d'un certain point : cela n'a pas vraiment de sens de dire qu'une université dont 100 % des étudiants ou des professeurs sont étrangers qu'elle est deux fois plus ouverte à l'international qu'une université dont seulement 50 % le sont. Bref, cet indicateur ne soutient pas une relation monotone croissante avec le concept prétendument mesuré et le sens de l'indicateur est différent s'il marque 15 % ou 90 %.

Ces quelques exemples montrent bien qu'il importe de connaître précisément le fonctionnement des indicateurs que l'on utilise. Il faut donc ouvrir ces « boîtes noires » car le diable est dans les détails, détails qu'ignorent le plus souvent ceux qui se fient aveuglément aux indicateurs.

Le fait que, parce qu'ils sont faciles à calculer et à utiliser, on continue à se fier à des indicateurs qui sont manifestement invalides au regard de critères de validité bien définis, relève de ce qu'on appelle parfois le « syndrome du lampadaire », lequel consiste à chercher ses clés non pas à l'endroit où on les a perdues mais là où c'est éclairé.

4. CONCLUSIONS

De ces brèves analyses, qui ne font qu'évoquer quelques résultats de recherches menées ailleurs de manière plus détaillée³, on peut tirer les conclusions suivantes :

- Les chercheurs et les professeurs doivent reprendre le contrôle académique du message public des universités et doivent éviter de donner, par leurs discours ou leurs (ré)actions, de la crédibilité à des classements improvisés et inadéquats à leur objet ;
- Plutôt que d'utiliser uniformément des indicateurs tout-terrain, les techniques d'évaluation doivent être adaptées à l'échelle d'analyse et aux disciplines ;
- De manière rationnelle on ne devrait utiliser que les instruments dont on a testé la validité technique ;
- Au lieu de faire usage d'indicateurs simplistes et uniques, il convient de s'assurer que les indicateurs de performance choisis soient multiples, bien définis et correspondent à la mission de l'institution évaluée ;
- Les données chiffrées peuvent être utiles comme aides au jugement et à la décision, mais elles ne remplacent pas l'évaluation par les pairs ;
- L'usage de mauvais indicateurs — copiés-collés depuis les sciences de la nature — est particulièrement néfaste pour la recherche en sciences humaines et sociales, qui a des enjeux spécifiques (objets locaux, cohabitation de paradigmes...);
- Face à « l'offre » accrue sur le marché des indicateurs en matière d'évaluation de la recherche menant à des décisions politiques, il est urgent que les gestionnaires de haut rang consacrent plus de temps à l'analyse de la signification exacte des indicateurs choisis au lieu de sauter dans le train des derniers « indicateurs » à la mode qui utilisent tout ce qui se trouve mesurable à portée de main (syndrome du lampadaire) ;
- Le manque de réflexions critiques sur les effets pervers de l'utilisation de la plupart des indicateurs existants sur la dynamique des institutions d'enseignement supérieur pourrait conduire à des

³ Yves GINGRAS, *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*, Paris, Éditions Raisons d'agir, 2014.

décisions politiques mal avisées affectant leur mission fondamentale de recherche et de formation de citoyens éclairés, originaux et critiques.

En somme, les indicateurs et autres « métriques » ne peuvent à eux seuls diriger les politiques en matière d'enseignement et de recherche. La valeur (positive ou négative) d'un indicateur n'a en elle-même aucune signification pratique. En cas de mauvaises performances d'un laboratoire ou d'une équipe de formation, faut-il la sanctionner ou au contraire la renforcer ? On voit ainsi qu'il est toujours possible, devant un résultat donné, de choisir entre deux voies opposées. Diriger c'est choisir et aucun algorithme ne peut le faire à notre place. Dire ou penser le contraire n'est qu'une façon peu subtile de cacher des choix sociaux et politiques derrière des chiffres.

Le H-index à l'index ?

HUGUES BERSINI

Directeur de l'IRIDIA (ULB),
Membre de l'Académie royale de Belgique

1. INTRODUCTION

Le H-index est un nombre chargé en théorie d'évaluer à la fois la productivité et la notoriété scientifique d'un chercheur. À la différence de nombreux indicateurs bibliométriques l'ayant précédé, il brille par sa simplicité et d'obtention et d'usage. Parce que trois clics souris suffisent à crucifier ou porter aux nues un chercheur, son existence et son exploitation croissante font l'objet de nombreuses réactions, levées de bouclier, discussions et polémiques. Je me propose de scinder ce bref article en trois parties. D'abord une rapide présentation de ce nombre, ensuite son analyse critique et finalement des recommandations d'usage.

Autant le dire d'emblée, je suis plutôt partisan de son utilisation, en tenant compte bien évidemment de toutes les recommandations d'usage que je m'apprête à énoncer. La raison première est qu'il rationalise et simplifie des débats sans fin sur la manière de prendre en compte et le nombre de publications et le succès référentiel de celles-ci (leur impact en volumes de citations). Depuis de nombreuses années que je participe à des réunions d'évaluation de chercheur sur base essentiellement de leurs publications, les modes de calcul se sont multipliés, tentant de prendre en compte à la fois le nombre et la qualité de ces publications en termes d'impact. La tâche est ardue et je me rappelle de discussions et de polémiques sans fin sur le classement des revues

dans lesquelles les chercheurs à évaluer avaient publié. Valait-il mieux une publication dans *Nature* ou deux dans *PNAS* ou dix encore dans une obscure conférence spécialisée dans la reproduction des crevettes roses dans la région du Yucatán ? S'ensuivaient alors de savants calculs d'apothicaires mêlant nombres de publications, participations aux congrès, aux ouvrages collectifs, facteurs d'impacts des revues, âge des grands-parents et code postal de la commune de résidence du chercheur... Nul ne s'accordait bien évidemment. Des colonnes de chiffres incomparables entre elles se multipliaient pour conduire le plus souvent au découragement des membres de la commission d'évaluation, à la colère ou à la décision pure et simple de recourir à un lancement de dés tellement plus garant de l'égalité.

Pour le moins, ce H-index cherche à conforter l'impact référentiel comme un critère essentiel de la qualité d'une production scientifique, ce qui, à titre personnel, me semble, dans une communauté scientifique idéale, exerçant la pratique de citation de la manière la plus constructive, honnête et apolitique qui soit, une excellente chose.

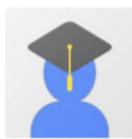
Ce sont donc les deux raisons essentielles pour lesquelles ce petit nombre, malgré tous ses défauts envers lesquels je m'empresserai de mettre en garde par la suite, a gagné mon adhésion : redonner un juste poids à l'impact référentiel d'un chercheur comme gage de sa créativité et de son talent, et grandement faciliter la tâche des évaluateurs pour autant qu'ils en fassent un usage parcimonieux et réfléchi.

2. DÉFINITION DU H-INDEX



Nombre de publications – Nombre de citations reçues par celles-ci.

Le H-index d'un chercheur a pour vocation de proposer un compromis ingénieux entre le nombre de ses publications et leur impact référentiel. Depuis toujours, l'on souhaite de tout chercheur qu'il soit productif, son nombre de publications produisant un témoignage assez éloquent de cette productivité, tout autant qu'influent, le nombre de citations qu'il a reçu, là aussi, pouvant en témoigner. Aussi, si le H-index d'un chercheur est de « X », c'est qu'il a



David Bowie

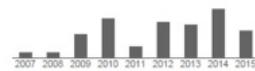
Associate Professor of Linguistics, Department of English, University of Alaska Anchorage
Sociolinguistics, Dialectology, Language and Age, Language and Religion, Utah English
Verified email at uaa.alaska.edu - Homepage



Google Scholar

Get my own profile

Citation indices	All	Since 2010
Citations	93	66
h-index	5	5
i10-index	3	2



Co-authors View all...

Yasuhiro Ozuru

Kenrick Mock

Title	1-20	Cited by	Year
THE EFFECT OF GEOGRAPHIC MOBILITY D Bowie University of Pennsylvania		26	2000
Religious affiliation as a correlate of linguistic behavior W Baker, D Bowie University of Pennsylvania Working Papers in Linguistics 15 (2), 2		19	2010
Early Development of the card-cord Merger in Utah D Bowie American speech 78 (1), 31-51		12	2003
The diphthongization of/ay: Abandoning a southern norm in southern Maryland D Bowie Journal of english linguistics 29 (4), 329-345		9	2001
Language change over the lifespan: A test of the apparent time construct D Bowie University of Pennsylvania Working Papers in Linguistics 11 (2), 3		8	2005
Acoustic Characteristics of Utah's CARD-CORD Merger D Bowie American Speech 83 (1), 35-61		5	2008
Perception and production in a series of related mergers D Bowie The LACUS forum, 297-305		3	2001
Was Mir Wisse: A Review of the Literature on the Languages of the Pennsylvania Germans D Bowie University of Pennsylvania Working Papers in Linguistics 4 (3), 2		3	1997

publié X articles dont chacun a reçu un minimum de X citations. La lettre H provient du nom du physicien, Jorge Hirsch, qui l'a proposé pour la première fois, il y a une dizaine d'années.

La manière de le calculer est illustrée par le cas de David Bowie ci-dessous dont le H-index est de 5.

La mise en œuvre algorithmique du calcul de ce nombre est des plus triviales. Cette facilité explique aussi en partie la raison de son succès. Il suffit pour l'obtenir d'ordonner les publications de David Bowie en nombre de citations décroissant. Ainsi, son premier article a recueilli 26 citations, bien supérieur à 1, son deuxième 19, bien supérieur à 2, son troisième 12 supérieur à 3, son quatrième 9 supérieur à 4, son cinquième 8, toujours supérieur à 5, puis enfin son sixième article, lui ne dépassant pas en citation le nombre 6. On peut donc arrêter là ce passage en revue descendant de ses publications pour aboutir au nombre 5.

Aujourd'hui de nombreux logiciels gratuits et facilement téléchargeables vous permettent de calculer aisément le H-index de qui que ce soit. Ce nombre est le dernier avatar d'une « googleisation » du monde par laquelle tout type

d'activité ou production humaine peut aisément faire l'objet d'une évaluation quantifiée et du classement qui s'ensuit (mais j'y reviendrai).

3. ANALYSE CRITIQUE

De quoi est fait ce H-index ?

La science est par définition une aventure humaine collective. Toute proposition scientifique reçoit son statut de vérité par approbation des pairs. C'est par cette adhésion collective qu'une expérience, un théorème ou un modèle gagne en crédibilité et se retrouve dans la colonne de la bonne science, loin du créationnisme, de l'astrologie ou du climato-scepticisme. Chaque écrit doit donc délibérément se prêter au jeu de l'évaluation par les pairs. Un pair est en général un scientifique dont vous souhaitez l'approbation, notamment parce que vous vous en êtes considérablement inspiré dans vos travaux. La condition minimale pour solliciter cette approbation est d'avoir préalablement référé ce collègue dans vos écrits, là où précisément son apport vous aura permis de progresser. Cette pratique de référence croisée rend compte de la stabilité du réel et de la progression douce de notre connaissance et compréhension de ce réel. Chacun fait la courte échelle à chacun pour voir un peu mieux derrière le mur de l'obscurantisme.

Et c'est bien parce qu'il en va de la déontologie de la pratique scientifique de citer les travaux qui furent nécessaires à l'accomplissement du vôtre (aucune « hard loi » ne l'impose mais la chose est devenue suffisamment normative), que le comptage de ces références rend compte assez fidèlement de l'impact de ces travaux. Bien avant l'existence du H-index déjà, un chercheur qui se reconnaissait dans les travaux d'un autre, sans voir apparaître son nom dans la bibliographie de l'article, avait légitimement de quoi s'offusquer et pouvait exiger « réparation ». De surcroît, l'obtention des prix et des récompenses en matière scientifique, nous sommes tous d'accord, se base généralement sur la paternité des idées originales qu'il devient donc d'autant plus important de reconnaître (autant jouer le jeu, car demain cela pourrait tout autant vous concerner). Dès lors, le comptage des références fait sens, justifiant d'autant l'importance que ce comptage prend dans l'établissement du H-index.

On pourrait être plus réservé sur le nombre de publications. Faut-il à tout prix chercher à maximiser ce nombre ? Là encore, rares sont les bons

chercheurs qui se sont contentés d'un nombre très restreint de publications. La recherche est une activité continue, incessante, toujours en cours, toujours dynamique, jamais tarie. Les idées arrivent et progressent vite, à une fréquence soutenue, et aucun ne veut tomber du train de ce voyage groupé, organisé, vers l'élucidation du réel. Le nombre de publications qui compose aussi en partie le H-index rend compte de cette envie, cette motivation et capacité du chercheur à rester dans la course (tous se reconnaîtront), à ne pas s'exiler de sa communauté, à maintenir sa participation dans le peloton des coureurs, si pas dans le groupe de tête.

Mais pourquoi classer les chercheurs ?

Il me semble nécessaire d'accepter ce type de classement pour deux raisons clés : l'identification des experts et la progression de carrière sur une base principalement méritocratique. Approfondissons quelque peu ces deux raisons.

Dans un monde qui se complexifie sans cesse, la décision politique, gouvernementale, se doit de reposer sur des prises d'expertise en matière technologique, scientifique. Or, il n'est pas simple, encore moins lorsqu'on est étranger au domaine, de repérer les experts reconnus de celui-ci. Parmi d'autres sources d'information, le H-index peut contribuer à cette identification. Au moins, il témoigne, et de l'activité persistante d'un expert et de l'attention soutenue que continuent à lui vouer ses pairs. Si, par ailleurs, ce même expert réussit à joindre à cette compétence reconnue, un brin d'intelligence sociale et quelques qualités pédagogiques, cela en fait le candidat idéal pour répondre à cette demande d'expertise. Par ce jeu de citations croisées que la science oblige, plus que dans beaucoup d'autres activités humaines, qui parfois exigent justement de se démarquer de ses pairs (l'art en cela est très différent des sciences, où les influences existent souvent de manière masquée), il est facile de repérer ceux qui semblent faire « la pluie et le beau temps » dans leur communauté et leur discipline. Ils auront alors la tâche ingrate de décharger le décideur d'éliciter les fondements rationnels de sa prise de décision.

Nous vivons une époque difficile où l'argent se raréfie sans cesse, l'argent public plus rapidement encore. Être chercheur ou professeur d'université à vie devient un luxe professionnel, un fonctionnariat honorifique, auquel de moins en moins de personnes peuvent aspirer. Du coup, les places sont rares

et la concurrence est vive. Sur quelles bases autres que méritocratiques, et qui ne provoqueraient davantage de frustration et de ressentiment, peut-on faire reposer la sélection des élus ? Le H-index ne rend explicite qu'un volet de la compétence des scientifiques mais un volet essentiel : sa productivité et son impact. Il laisse entièrement sous silence ses qualités sociales, son talent pédagogique (certainement celui qui s'exerce oralement), aspects essentiels pour un poste de professeur, mais sans doute moins pour un « chercheur pur ». Avec toutes les recommandations d'usage qui suivent, je trouverais donc personnellement regrettable que lors d'une promotion ou d'un échec à l'issue d'une candidature pour l'un de ces postes, ce résultat heureux ou malheureux ne fasse pas la moindre allusion au H-index du chercheur en question.

Avantages et défauts très pratiques du H-index

Clairement, un des avantages premiers du H-index est d'avoir tenté de ramener à un seul nombre ce que chacun a toujours considéré comme primordial dans l'évaluation d'une production scientifique : la quantité (le nombre de publications) et la qualité (leur impact référentiel). J'ai eu déjà l'occasion de rappeler les nombreuses autres vaines tentatives auxquelles il m'a été donné d'assister, cherchant à évaluer cette production à travers la valeur bibliométrique des revues ou des congrès. Avec le H-index, tout redevient plus simple, plus immédiat. On comprend assez aisément la séduction exercée par ce seul nombre magique. Ainsi, un nombre est plus facile à comparer à un autre que deux à deux autres. Au risque selon certains de mélanger des pommes et des poires, on évite, soulagé néanmoins, toute l'épineuse difficulté des décisions ou des classements multicritères. Il est aussi très facile d'additionner ou de moyenniser ce nombre avec d'autres, de sorte, par exemple, à calculer le H-index d'un labo (la somme ou la moyenne des H-index de ses chercheurs), le H-index d'une université (pareil) ou d'une revue (la somme ou la moyenne des H-index des auteurs qui y publient). Au-delà de son obtention facile, cette simplicité d'usage due à son « unicité » est l'autre grande explication des faveurs qu'il obtient auprès de ceux en charge de la sélection des « bons scientifiques ».

De nombreuses difficultés et des défauts sont néanmoins à épinglez quant à l'utilisation de ce seul nombre. Je me fais ici et pour l'essentiel le porte-parole ou le rapporteur de toutes les critiques que ce nombre a déjà essuyées.

Comme on l'aura constaté pour le cas de David Bowie, son obtention exige la mise à disposition d'une base de données qui reprend toutes les publications et toutes les citations reçues par un auteur. Or, plusieurs de ces bases de données existent, « Web of knowledge », « Scopus », « Google Scholar », et leur contenu diffère, débouchant sur des valeurs de H-index distinctes pour un même chercheur selon que l'on se base sur l'une ou sur l'autre. Sa facilité et sa gratuité d'usage rendent Google Scholar la plus utilisée jusqu'à présent (il semblerait que la valeur obtenue via cette base de données soit très souvent supérieure aux autres), mais il serait toutefois préférable lors d'une comparaison, soit d'utiliser la même base de données pour tous, soit de moyenner les valeurs obtenues par plusieurs bases.

Cet outil algorithmique est purement syntaxique, soyez donc sûr de l'orthographe du nom du chercheur que vous souhaitez évaluer. Ne le pénalisez pas à cause des multiples déclinaisons de son nom (assez caractéristique des noms composés, à rallonge ou simplement difficile à orthographier) ou ne le gratifiez pas de publications dont il n'est pas l'auteur (grâce à son homonyme... Rappelez-vous, David Bowie). Plus important encore, les traditions de publications et de citations sont très différentes d'une communauté scientifique à l'autre, ce qui conduit, par exemple, à des valeurs de H-index maximales étonnamment hautes dans certaines disciplines (comme la médecine, où des valeurs de 160 ne sont pas rares) et basses dans d'autres (comme la philosophie ou le droit). Tout d'abord, les communautés scientifiques peuvent être de taille très différente, les développements par principe concerner un public cible beaucoup plus local et restreint, le rythme de publication se trouver beaucoup plus lent dans certaines disciplines (en médecine et en physique, il faut publier vite, sans doute beaucoup plus vite qu'en philosophie) et moins suivies et régentées les pratiques de citation dans d'autres. Dès lors, il devient important de ne comparer par ce nombre que des chercheurs appartenant à une même communauté et même sous-communauté ou, si un besoin impérieux naît de comparer deux spécialistes de disciplines différentes, de normaliser les H-index au départ des valeurs minimales et maximales que l'on rencontre dans leur communauté.

Autre travers important de ce nombre, il se base sur les publications d'un chercheur qui est pourtant que très rarement seul à avoir écrites. Certaines publications peuvent être signées, cela arrive, par des centaines d'auteurs (on pense aux physiciens du CERN ou à des biologistes ou médecins dans des

grands centres de génétique ou de cancérologie). Or, le H-index ne prend en compte ni le nombre d'auteurs, ni leur positionnement dans la signature de l'article (on sait pourtant, tradition scientifique oblige, que ce positionnement reflète, d'une manière ou d'une autre, l'importance prise par le chercheur dans l'obtention des résultats relatés dans l'article). On peut aisément admettre qu'un H-index de 50 d'un auteur dont les publications l'unissent à beaucoup d'autres rend moins compte de sa valeur propre qu'un H-index de 30 obtenu par un chercheur qui a toujours choisit de signer seul ses articles. De nombreuses manipulations arithmétiques ont également été proposées de façon à prendre en compte le nombre et le positionnement des auteurs dans le bénéfice à retirer des citations de leurs confrères.

Finalement, cet index, de nature cumulative, est évidemment très sensible à l'âge du chercheur, autre aspect à ne pas négliger, et tend aussi à défavoriser les auteurs déjà très âgés, dont les articles ne sont pas référencés dans ces bases de données logicielles. En substance, l'âge du chercheur compte énormément dans l'importance qu'il faut accorder à ce nombre.

4. CONCLUSIONS EN FORME DE RECOMMANDATIONS D'USAGE

Dans un ouvrage que j'ai publié il y a quelques années de cela et intitulé *Haro sur la compétition*¹, je m'inquiétais de la montée du registre compétitif dans l'ensemble des possibles modes d'interaction humaines. Une des origines de cette montée était, de fait, la mise à disposition aisée de comparatifs informatisés amplifiant d'autant la concurrence parfaite réclamée par l'économie de marché. Il me semblait déjà, à l'époque, que la multiplication des classements, que les moteurs de recherche Internet encouragent par un simple clic, participait également de la propagation de la compétition dès que la moindre sélection se trouvait en jeu. Et effectivement, tout tend aujourd'hui à se hiérarchiser, à se présenter sous forme de classement, tout se « googleise » : les écoles, les universités, les cliniques, les villes, les lieux de villégiatures, les hôtels, les publications, les livres, les chercheurs, les œuvres d'art, les artistes, les salariés. Difficile d'imaginer aujourd'hui une seule facette de notre quotidien qui ne puisse faire l'objet d'une évaluation et d'un classement.

¹ Publié aux Presses Universitaires de France en 2010.

Vive Google, enfin parvenu à faire de chaque recoin caché de nos existences l'objet d'un classement !

Il me semble souhaitable de différencier classement et conséquence d'un tel classement. Une société plus soucieuse d'égalitarisme et de détresse pourrait partir de ce classement pour aider et encourager les perdants. Une société plus libérale, méritocratique, et sous l'emprise des mécanismes auto-régulateurs du marché, choisira de mettre tous ses œufs dans le panier des gagnants. Il en va ainsi du H-index, mais on peut comprendre que ceux qui n'arrivent pas à élever le leur peuvent craindre, à raison, des retombées négatives que cela risque d'avoir sur leur carrière. Ils tenteront alors de recourir à tous les moyens décents et indécents pour le faire décoller : se cantonner aux domaines scientifiques les plus populaires (et contribuer ce faisant à renforcer leur popularité par un mécanisme de rétroaction positive), publier à outrance (saucissonnage des publications, bien que cela ne favorise en rien la composante référentielle du H-index), tenter de coller aux « stars » de son domaine (chercher absolument à publier avec eux pour bénéficier de leur aura), suivre et forcer les collègues à vous citer (par mesure d'intimidation ou par le pouvoir que vous pourriez exercer comme « reviewer », ne référer positivement que les publications qui vous citent indépendamment de leur contenu). Certains petits malins sont mêmes parvenus à se créer une très haute valeur de H-index sans n'avoir jamais rien publié, en s'inventant de toute pièce des publications et des citations, mais le système est devenu suffisamment résilient pour éviter que cela ne puisse se reproduire.

On peut décemment se demander si l'influence d'indicateurs, tels le H-index, sur le comportement des chercheurs va dans le sens d'une amélioration du bien-être de ces mêmes chercheurs et plus encore de la qualité de leur production scientifique. Vu la grande jeunesse de cet indice, il est à ce stade encore difficile de se prononcer. On connaît les critiques généralement adressées aux situations d'intense compétition qui profitent aux gagnants et, idéalement, à la qualité de leur production, mais sont délétères pour les perdants qui souffrent des inégalités de traitement. Sur le long terme, la science a-t-elle à gagner de cette multiplication des classements ainsi que cette exhibition ostentatoire du succès de certains chercheurs ? Est-ce l'obSESSION des podiums qui conduira les chercheurs de demain à mettre au point le vaccin anti-sida ou le dispositif géothermique maintenant l'augmentation de la température moyenne du globe en-dessous de deux degrés ?

De manière à atténuer au mieux les effets négatifs que la comparaison des H-index est très susceptible d'induire, je terminerai, en sus des défauts du H-index déjà relevés dans les paragraphes précédents, par ces deux dernières recommandations d'usage :

- Surtout ne pas importuner ou perturber les jeunes chercheurs avec un tel indice. De par sa nature cumulative, il ne faut pas y accorder la moindre attention avant que le chercheur n'ait atteint l'âge d'une maturité scientifique suffisante (autour des 35 ans). Il est indispensable que notre chercheur ait eu vraiment l'occasion de faire ses preuves, de produire et se faire reconnaître dans sa communauté. Et cela prend le temps que cela prend...
- Surtout n'octroyez pas à cet index une valeur absolue mais toute relative au contraire. S'il est clair que l'on peut à juste titre considérer qu'un H-index de 50 soit à réellement différencier d'un H-index de 5, et qu'une telle différence reflète indéniablement une réalité : le premier chercheur est sans conteste plus productif et mieux reconnu, je ne suis pas du tout convaincu, à la lumière de tous les défauts déjà mentionnés : spécificité disciplinaire, âge, problème multi-auteurs, discipline plus populaire qu'une autre, parrainage ou non par des vedettes, etc., qu'une même volonté de différenciation doive s'exercer entre un H-index de 50 et de 30.

Réflexion sur l'évaluation de la recherche en physique et gestion

BARBARA CLERBAUX

Professeur au département de physique (ULB)
et spécialiste de la physique des particules élémentaires

MAREK HUDON

Professeur à la Solvay Brussels School (ULB)
et spécialiste d'économie sociale et éthique des affaires

Cette contribution a pour but de comparer les pratiques d'évaluation entre différentes disciplines de sciences sociales, sciences humaines et sciences exactes. Plus précisément, nous proposons un exercice de réflexion à deux voix entre science de gestion et science physique. Nous commençons par présenter les critères principaux utilisés pour l'évaluation dans les deux disciplines et leur utilisation différenciée. Par la suite, nous aborderons des constats d'inquiétude du système d'évaluation actuel partagés par les deux disciplines.

En guise d'introduction, nous souhaitons souligner que l'évaluation de la recherche et des chercheurs a énormément évolué ces 20 dernières années. Nous avons directement vécu ce changement dans les deux disciplines. Une modification des pratiques qui nous paraît importante concerne la transparence quant aux critères d'évaluation et à la composition des comités, à ses rotations, ainsi qu'à l'indispensable évaluation de l'évaluation. Ces nouveaux outils nous permettent de mieux juger le système, de le faire progresser et de tendre vers une procédure d'évaluation la mieux appropriée possible pour le

développement du savoir. Après évaluation, le retour à l'évaluer ou au projet, dans n'importe quel domaine, est également une évolution positive pour les postulants.

Un premier constat qui ressortait dans les deux disciplines en sciences de gestion et en science physique est la forte standardisation de l'évaluation. Il y a un phénomène d'alignement général sur les critères et standards internationaux avec un alignement vers les critères européens principalement. Nous nous sommes alors posé la question de savoir si les critères d'évaluation sont les mêmes pour les différents champs de recherche. En science physique, divers critères sont pris en compte. Un critère important est la diffusion et la portée des résultats, et en particulier dans les revues internationales, autant du point de vue du nombre de publications que de la renommée de la revue, le nombre de citations des articles est aussi regardé. D'autre type de critères sont utilisés comme les présentations des résultats lors de conférences et colloques, les collaborations internationales ou nationales reflétant le dynamisme du chercheur. Par exemple, les fameux « H index », ne sont pas systématiquement repris et ne sont pas toujours discutés.

En science de gestion, l'évaluation se fait de manière différente, dû en partie à un contexte historique particulier. Lorsque la nouvelle faculté (Solvay Brussels School) résultant de la fusion de l'ancienne école de commerce avec le département des sciences économiques, il n'y avait pas de mesure d'évaluation standardisée. Le besoin d'avoir un référentiel s'est rapidement fait sentir. La culture de recherche était fort différente entre les deux entités. Les méthodologies de recherche utilisées en science économique sont souvent de la modélisation et une approche empirique quantitative alors que la recherche qualitative est aussi présente en science de gestion, à cause des nombreux liens avec la sociologie, de la psychologie, ou d'autres sciences humaines ou sociales. La création d'un classement de revues fut alors décidé en agrégeant deux classements externes : le classement du CNRS et le classement anglais, et il a été décidé de donner des « points » à chaque publication en fonction de ces classements des revues. La « qualité » de la revue est ainsi définie par ce classement, et chaque publication est pondérée par ce facteur de qualité. Ce classement est utilisé pour les nominations à titre définitif mais aussi en partie pour évaluer les dossiers de promotion ou lors des recrutements. Nous avons également mis en avant des revues « d'excellence » qui sont fort bien notées dans les classements anglais et français, par exemple l'équivalent de

Nature qui, en science de gestion, est l'*Academy of Management Review*. Dans ce genre de système d'évaluation, il est donc important de publier dans les meilleures revues des champs considérés. Finalement, l'ensemble de ces nombres sont pris en compte et additionnés pour arriver à un indice final. Avec ce système, plus l'indice final est élevé, plus le chercheur est performant. Par exemple, un nombre minimum de points (25 à la Faculté Solvay) est requis pour être confirmé comme chercheur définitif. Un enjeu stratégique en découle donc quant au choix du domaine de recherche, au choix des revues dans lesquelles publier et quant à l'optimisation du nombre de publications ; cela crée une tension non négligeable pour certains chercheurs.

L'élément positif de ce système d'évaluation très codé et quantitatif est la diminution d'une certaine partie des conflits/tensions entre les cultures différentes des chercheurs en termes de valeur de la recherche et de la publication. Certains chercheurs estimaient qu'une de leur publication dans une excellente revue représentait un meilleur travail de recherche que d'autres confrères ayant effectué une succession de publications dans des revues légèrement moins bien classées, ce qui entraînait fréquemment un déficit de reconnaissance. Le système proposé définit donc l'échelle de comparaison. Cependant, des éléments négatifs du système sont apparus liés à la définition même de la procédure. Dans certains domaines il y a des publications bien cotées (par exemple cotées 10) alors que dans d'autres disciplines (par exemple, économie de l'éducation, finances historiques), il n'y a aucune revue notée au-delà de 6 ou 7. Pour des chercheurs travaillant dans ces sous-disciplines, il faut donc produire beaucoup plus de publications pour arriver à égaliser le total de points plus facilement obtenu dans d'autres disciplines. De même, pour un chercheur qui publie principalement dans des revues plus proches de disciplines liées aux sciences humaines, les facteurs d'impact sont typiquement plus faibles que dans le cas de disciplines plus proches des sciences exactes.

Nous observons ainsi que pour les deux disciplines mises en comparaison, en science de gestion et en science physique, même si les indicateurs de base sont similaires, ils sont utilisés très différemment dans les deux branches. Si en science de gestion la procédure est très codée, en science physique, davantage de liberté est donnée aux comités de sélection/aux évaluateurs.

Nous nous sommes également rendu compte qu'il y a une série de biais et de dangers dans le système d'évaluation, menant à une inquiétude partagée

généralement par un grand nombre de chercheurs, de toutes disciplines. Nous décrivons ci-dessous quelques uns de ces parents pauvres qui nous paraissent relevant.

Le premier danger qui nous est apparu concerne les chercheurs ou les projets avec un profil visant l'interdisciplinarité. Pouvoir faire des passerelles entre les disciplines est généralement génératrice de nouvelles idées ou de nouveaux développements. Pour illustrer ce point mentionnons François Englert qui, pour sortir d'une impasse en physique théorique, a appliqué des idées provenant d'un autre domaine de la physique (physique statistique et du solide), lui permettant de proposer une solution radicalement nouvelle, il a ainsi réussi à débloquer la problématique en question¹. Paradoxalement, les projets ou profils interdisciplinaires sont généralement aujourd'hui plutôt soutenus par le politique, les autorités et l'opinion publique. Cependant, on remarque que concrètement ces profils ne passent pas fréquemment les étapes d'évaluations de manière satisfaisante. L'évaluation est faite par les pairs et généralement à travers de commissions qui sont relativement disciplinaires aussi. Dans quelle commission faut-il alors évaluer ces projets ? Avec le danger que les experts ne se sentent pas compétents. Pour minimiser ce problème, des commissions mixtes ont été proposées dans certains cas, qui a l'intérêt de créer un débat entre les disciplines. La aussi, il y a la difficulté de questionner la discipline de l'autre à travers sa propre expertise qui est différente. Le même constat peut être fait par rapport à la question des classements de revues. L'interdisciplinarité peut y être incluse mais les revues interdisciplinaires sont souvent moins bien évaluées que les revues disciplinaires et le chercheur doit alors publier beaucoup plus pour arriver à un équivalent en terme de points. Nous remarquons aussi que dans une majorité de disciplines, les revues considérées comme de haut niveau, sont typiquement des revues américaines.

Un autre parent pauvre du système d'évaluation actuel concerne le problème de transfert de savoirs entre les scientifiques ou plus généralement les chercheurs et la société. Il nous semble important d'avoir une politique à long terme afin de combler le fossé des connaissances techniques entre scientifiques et société. Or actuellement, les activités liées à la vulgarisation ou « outreach », visant donc le grand public n'est typiquement pas ou

1 FR. ENGLERT and R. BROUT, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* 13, (1964), p. 321.

peu valorisé dans les évaluations. Liée à cette question, citons également la problématique du livre, un autre outil pour la diffusion du savoir. Or les classements mentionnés plus haut se basent sur les revues principalement, la publication de livres n'est donc que peu valorisée dans le système d'évaluation actuel. Ce dernier n'incite donc pas le chercheur à passer du temps et de l'énergie à écrire un ouvrage de référence ou un ouvrage de vulgarisation. Les livres permettent cependant d'aller beaucoup plus loin dans la réflexion et ceci pour toutes les disciplines de la recherche. A titre d'exemple, s'il y a un livre dont on parle beaucoup dans la presse ou entre collègues, c'est le livre de Thomas Piketty *Le Capital au XXI^e siècle*². Dans le domaine de science économique, ce livre serait repris avec une pondération nulle, en effet le système d'évaluation utilisé dans ce domaine ne considère que les publications dans l'une des dix meilleures maisons d'édition (Stanford, Harvard, etc.), où des points sont alors partagés entre les co-auteurs –contrairement aux articles. Comme on le voit, le problème est d'avoir une possibilité de publication qui permet d'aller plus loin dans la réflexion par rapport aux articles et de pouvoir les valoriser à leur juste valeur.

Un troisième constat qui nous a marqué se rapporte au danger de concentration et de standardisation, que ce soit en terme de méthodologie ou de thématique. S'il est logique de soutenir les groupes performants, définis comme tel d'après le système actuel, il nous semble important ne pas tomber dans la spirale de la concentration de la recherche, au détriment de sa diversité. Le danger est d'avoir un système qui s'auto alimente et s'auto évalue, créant alors un biais dans l'évaluation : avoir un support acquis aux gros groupes ou aux chercheurs stars qui accumulent les récompenses. Il est donc capital de rester vigilant quant à l'ensemble des biais possibles du système, notamment afin de ne pas nuire à la diversité et la créativité qui est le moteur de la recherche et du développement de nouvelles idées. Dans nos deux domaines respectifs, il nous semble que le plus pertinent est une évaluation par champ, par exemple le marketing ou la recherche opérationnelle pour la gestion, afin de tenir compte des spécificités de chacun d'entre eux.

Nous voudrions conclure par quelques pistes de réflexion. (i) Ne pourrait-on pas « inverser » le système d'évaluation en faisant davantage confiance au chercheur au départ? Celui-ci a intérêt à faire la recherche qui lui semble la plus prometteuse avec des étudiants de haut niveau. Avec le système actuel,

² PIKETTY, T., *Le capital au XXI^e siècle*, Paris, Le Seuil, 2013.

les chercheurs/projets sont davantage évalués en amont du travail proposé, mais plus vraiment par la suite. (ii) Nous ne protégeons pas suffisamment l'interdisciplinarité par des outils spécifiques, on la prône mais elle émerge peu. Il faut assurer des moyens afin de développer et pérenniser ces profils. (iii) Nous proposons de mettre à disposition une aide plus active et efficace pour les chercheurs dans l'accompagnement des projets à tous les niveaux (universitaires, régionaux, européens, etc.). Cela aiderait les scientifiques à effectuer leur mission première : la recherche, dans de bonne condition. Ces pistes de réflexion pourront être reprises et développées lors de la seconde journée de discussion dont le but est de travailler sur l'élaboration d'une liste de propositions concrètes pour améliorer le système d'évaluation actuel.

L'impact des rankings sur l'existence et la diversité des revues scientifiques et sur les pratiques disciplinaires. Le cas de la géographie

CHRISTIAN VANDERMOTTEN

Professeur de l'Université libre de Bruxelles

Membre de l'Académie royale de Belgique

Je vais focaliser cette réflexion sur le cas de la géographie, celui que je connais. Mais les conclusions que l'on peut tirer de ce qui va suivre ont sans doute une portée générale, le cas de la géographie amplifiant néanmoins vraisemblablement certains constats, parce que la discipline est très hétérogène, mal positionnée dans les classements disciplinaires par les instances responsables de la recherche et donc sujette à des procédures d'évaluation dans lesquelles une partie des évaluateurs peuvent être amenés à juger selon des critères qui ne correspondent pas aux réalités de la discipline, parmi lesquelles, pour une partie des géographes du moins, un ancrage fort dans les problématiques sociétales locales. Ceci n'exclut évidemment pas le recours des géographes aux théories générales. Mais dans le contexte de la géographie sociale en particulier, la question de l'usage de la langue n'est évidemment pas neutre, à la différence d'un chimiste ou d'un mathématicien, voire d'un géographe travaillant sur le *global change*.

1. LES PUBLICATIONS EN GÉOGRAPHIE AVANT LA SECONDE GUERRE MONDIALE

Outre les livres, les publications en géographie avant la Seconde Guerre mondiale relevaient de trois types de revues :

- des revues de sociétés savantes, nationales ou parfois locales, remontant pour beaucoup d'entre elles au XIX^e siècle ;
- quelques revues généralistes prestigieuses, parfois issues de comités nationaux ou ayant évolué au départ des publications de sociétés savantes, mais traduisant l'évolution de la géographie vers un positionnement scientifique (ex. les *Annales de Géographie*, fondées en 1891 par Vidal de la Blache ; *The Geographical Journal* ; les *Annals of the Association of American Geographers* ; les *Petermanns Geographische Mitteilungen*, fondées dès 1855, etc.) ;
- quelques revues spécialisées, portées au départ par l'un ou l'autre laboratoire de renom, comme la *Revue de Géographie Alpine*, fondée à Grenoble en 1913 par Raoul Blanchard, ou *Economic Geography*, fondée en 1925 à la Clark University, voire par des sociétés de géographie (comme le *Tijdschrift voor Sociale en Economische Geografie*, 1910, sous les auspices de la KNAG, la société de géographie néerlandaise).

2. LA SITUATION ACTUELLE

Comme dans les autres disciplines, le nombre de revues géographiques a aujourd'hui explosé, en même temps que le statut du livre s'affaiblissait et tendait à évoluer de plus en plus vers la formule du recueil de contributions, du moins en tant qu'outil de diffusion d'une connaissance nouvelle pointue. En reprenant les revues qui se réclament de la géographie ou de ses sous-disciplines et celles dans lesquelles publient souvent des géographes, j'ai recensé une liste de 761 titres ; l'Union géographique internationale a mené pour sa part un recensement plus large et arrive à 1310 titres.

Parmi ces revues, on note évidemment une hégémonie accrue de la publication en anglais et dans des revues aux mains d'un petit nombre d'éditeurs internationaux (Reuters et Thompson, Elsevier, Springer). Cette hégémonie assure aux publications de certains pays une reconnaissance dans les relevés

bibliométriques qui n'est en aucune manière proportionnelle à la qualité intrinsèque de leur production : ainsi, sur les 761 revues relevées, le H-index moyen de *Publish or Perish* est de 27,3 pour les revues publiées aux Pays-Bas, de 18,7 pour celles publiées en Grande-Bretagne, de 15,4 pour celles publiées aux États-Unis et seulement de 4,4 pour celles publiées en France ! Parmi les revues recensées par l'Union géographique internationale, 69 % sont recensées dans le *Science* ou le *Social Science Index* parmi les 723 revues publiées en anglais, 50 % parmi les 14 publiées en allemand (un effet Springer ?), 12 % parmi les 57 publiées en français, 4 % parmi les 240 publiées en chinois et 3 % parmi les 108 publiées en espagnol. Il est clair que les classements contribuent à imposer une forme d'impérialisme scientifique.

Avec la croissance du nombre de revues, on assiste à la multiplication des revues thématiques spécialisées, également publiées pour beaucoup d'entre elles par ces mêmes firmes internationales, dans lesquelles le terme « géographie » n'apparaît pas mais où publient des géographes à côté de personnes issues d'autres disciplines (environnement, transport, *global change*, etc.).

Enfin, ces grandes firmes ont aussi pris le contrôle de certaines revues publiées de longue date par des sociétés de géographie, en anglais (celles de la Royal Geographical Society, de la Société écossaise de géographie, de la Société norvégienne, de la KNAG aux Pays-Bas, etc.). Il en résulte que le H-index moyen des revues issues de sociétés de géographie reprises par les firmes commerciales mentionnées ci-dessus est de 15,9, alors qu'il n'est que de 6,1 pour les autres revues de sociétés.

En conclusion, la marchandisation de la publication scientifique s'est imposée. La conséquence en est une forte réduction des échanges gracieux, qui jadis fournissaient l'essentiel des revues dans les bibliothèques universitaires. Aujourd'hui à l'inverse, celles-ci ploient sous le coût d'abonnements, électroniques ou non, très coûteux et les universités et les pouvoirs publics paient trois fois pour le coût de la recherche : une première fois en la finançant, une seconde en payant la contribution demandée aux chercheurs pour publier, et une troisième en payant pour y avoir accès (il arrive que l'auteur lui-même doive payer pour consulter sa propre production !). Quant aux firmes internationales d'édition, elles ne rémunèrent même pas les *referees* qu'elles mettent à contribution (et qui pendant ce temps ne font pas de recherche) et elles renvoient de plus en plus vers les auteurs eux-mêmes la

charge de la mise en page. On sait que ces firmes génèrent des taux de profit particulièrement élevés.

3. LES IMPLICATIONS ÉPISTÉMOLOGIQUES ET SOCIALES

Cette évolution n'est pas neutre en termes d'épistémologie disciplinaire, ni en termes d'utilisation sociale de la recherche géographique.

Les revues à haut facteur d'impact diffusent largement une vision épistémologique hégémonique, auto-reproduisant des concepts et une pensée anglo-saxonne standardisée. On n'écrit pas et on ne problématise pas de la même manière un article dans une revue anglo-saxonne et dans la géographie française.

Il s'ensuit aussi un recul d'une vision holistique de la géographie, intégratrice de différents aspects de la dynamique sociale pour en examiner les impacts spatiaux globaux, difficile à publier dans des revues très spécialisées, et aussi un recul des études régionales, moins bien classées en termes de visibilité, puisqu'elles ne s'adressent par définition qu'à un public plus circonscrit. Ainsi, parmi les 761 revues recensées, le H-index moyen des revues généralistes de géographie est de 6,4, alors qu'il s'élève à 14,7 pour les revues classées *Urban studies*, à 16,8 pour les *Economic and transport studies* et à 17,7 pour les *Global change and biogeography*. Dès lors, les jeunes chercheurs en quête de reconnaissance académique tendent à préférer ces revues thématiques spécifiques, dont les H-index sont plus élevés, puisqu'elles s'adressent à un public issu de différentes disciplines, dont beaucoup sont numériquement plus puissantes que la géographie. Il en résulte une perte de l'identité disciplinaire, dont le maintien n'est certes pas un objectif en soi, mais qui peut éviter des dérives technocratiques aux dépens d'une vision sociétale intégrée des problématiques spatiales.

Ce recul de la vision globale est renforcé par la tendance au saucissonnage des publications (et de manière accessoire à la multiplication d'articles ajoutant chacun peu aux autres déjà publiés sur le même thème), lié aux contraintes du nombre de publications et du *ranking* dans lesquelles doivent s'inscrire les jeunes chercheurs pour promouvoir leur carrière (ou les plus âgés pour assurer la reconnaissance et le financement de leur laboratoire). Il faut y ajouter, on l'a déjà dit, la dévalorisation des livres, qui peuvent pourtant

permettre, en sciences humaines, l'expression d'une pensée plus riche et globale que dans les articles de revues.

Beaucoup d'articles multiplient les citations de quelques grands noms de la discipline, obligées pour apparaître crédible aux yeux des *referees* des revues dites internationales (et par mimétisme dans les autres revues). Ceci entraîne des dérives dogmatiques (on cite le nom de manière non critique, incantatoire, à la limite sans avoir lu ses travaux, comme une preuve de sérieux scientifique, sans que cette citation ne constitue en aucune manière le renvoi à une démonstration ; pensons par exemple en géographie urbaine au nombre de renvois purement révérentiels ou formels à des noms comme Sassen, Castells, Florida, Taylor, etc., souvent appelés pour soutenir l'énoncé d'évidences).

En outre, ce renvoi systématique obligé (parfois d'ailleurs demandé par les *referees* et les éditeurs) à quelques « vedettes » de la littérature internationale, en fait anglo-saxonne, y compris dans des études locales à finalité appliquée, contribue, avec le poids croissant des bureaux d'étude internationaux et les pratiques de *benchmarking* qu'ils imposent, dans le référentiel néo-libéral dominant, à uniformiser les « solutions » proposées localement aux autorités politiques par la recherche appliquée. Ainsi, en études urbaines appliquées, le recours systématique aux grandes références de la mise en réseau des villes et de l'internationalisation des économies urbaines, contribue à vulgariser auprès des responsables politiques la généralisation des recettes néo-libérales de la rénovation urbaine, qui de fait contribuent à creuser les disparités sociales dans la ville et à en expulser les pauvres indésirables.

Les auteurs sont amenés, consciemment ou non, à se soumettre, pour satisfaire aux exigences des *referees*, aux critères qui sont ceux d'une certaine pratique anglo-saxonne de la géographie, parfois aseptisée, en général moins dialectique que la géographie française, ou à des normes provenant d'autres disciplines dans les revues thématiques. Ils sont amenés aussi à sacrifier aux « manies » de coteries scientifiques contrôlant certaines revues prestigieuses. Ainsi, sans que cela ne soit écrit nulle part et en dépit d'avis favorables des deux *referees*, un collègue s'est vu refuser par l'éditeur un article dans une revue de géographie politique qui se veut pourtant critique, *Antipode*, parce que s'appuyant sur l'analyse de cartes et de tableaux, donc sur l'empirie. Pauvre théorie si elle ne peut être articulée avec l'empirie ! Ou encore, les revues « établies » imposent implicitement que le sujet soit « à la mode »

pour accepter un article, ce qui brime évidemment toute possibilité de publication originale ou hors des normes, et contribue ainsi au lissage du contenu de la publication scientifique. On connaît la boutade selon laquelle Einstein n'aurait pas été publié dans une revue de physique selon les critères contemporains, outre le fait qu'il n'avait pas un facteur d'impact qui lui aurait permis aujourd'hui d'être recruté par une université.

4. ATTITUDES ET RÉACTIONS

En Belgique, les tendances qui viennent d'être évoquées semblent mieux acceptées (ou mieux subies) du côté néerlandophone, qui ne publie pratiquement plus qu'en anglais et où les universités ont encore plus intégré que du côté francophone le primat des logiques de sélection imposées par le *ranking* et les logiques de la soi-disant excellence internationalisée ; le monde universitaire francophone a plus de relations avec la France.

Du côté français, on connaît les oppositions politiques à l'hégémonie anglo-saxonne. Une réaction contre celle-ci est organisée sous l'égide du CNRS, qui a contribué au financement du site *revues.org*. Celui-ci abrite et met en ligne gratuitement de nombreuses revues électroniques, pas seulement françaises ; *Belgeo*, la revue belge de géographie, fait partie du lot. Cette politique a favorisé la diffusion de revues électroniques gratuites, tant pour les lecteurs que pour les auteurs, dans les autres pays latins (Espagne, Portugal, Roumanie, pays latino-américains). Parmi les revues reprises dans notre recensement, la part du *full open access* (FOA) est de 78 % pour les revues latino-américaines, de 64 % pour les revues ibériques, de 60 % pour les revues françaises, de 23 % pour les revues canadiennes ; à opposer à un pourcentage de 9 % aux États-Unis, de 8 % en Allemagne, de 2 % aux Pays-Bas et de 0,5 % seulement en Grande-Bretagne. Il y a donc clairement deux modèles qui s'opposent et, au-delà, une stratégie scientifique et culturelle sous-jacente, promue par le gouvernement français. L'université française a préféré, au recours au classement des revues par le *Web of Science* (*Science et Social Science Index*) (le plus restrictif et le plus hégémoniquement anglo-saxon, qui dépend de Reuters et Thompson) ou, moins restrictif, par *Scopus* (qui est produit par Elsevier), la prise en considération de tout article publié dans une revue reconnue par la liste de l'AERES pour la discipline, mise en place par des commissions d'universitaires et non hiérarchisée. Les commissions

françaises de recrutement examinent donc le contenu des articles plutôt que de recourir à des critères bibliométriques automatiques. Cette politique de publication en ligne gratuite initiée par le CNRS a fait bondir la diffusion effective des contenus des revues, ce qui ne se traduit pas nécessairement dans les classements bibliométriques « reconnus ». Dans notre recensement, le H-index moyen des 154 revues en FOA n'est que de 4,6, alors que la moyenne générale est de 12,7. La revue française de géographie la plus lue, *Cybergéo*, de grande qualité, est une revue gratuite en ligne, avec de l'ordre de 100 000 visites mensuelles ; mais elle n'a qu'un H-index de 9 et, si elle est reprise par *Scopus*, elle ne l'est pas dans le *Web of Science*.

En Belgique (francophone), il faut souligner la sensibilité de l'Académie, du moins à la Classe des Lettres, et semble-t-il aussi du FNRS, à la question de l'hégémonie imposée par les *rankings* et de ses conséquences sur les publications.

Il faut aussi mentionner les réactions de plus en plus nombreuses de scientifiques de différentes disciplines au monopole imposé par les revues commerciales anglo-saxonnes et à ses conséquences sur l'évolution épistémologique des disciplines et sur la manière de travailler (voire de frauder les résultats) dans les laboratoires et d'en publier les résultats (multiplication des auteurs de complaisance, saucissonnage des publications, manque d'originalité de celles-ci, publications redondantes, etc.). Ces réactions s'expriment dans des mouvements comme *Slow Science*, ou des déclarations comme celle de San Francisco.

Soulignons enfin que les classements bibliométriques « reconnus » passent totalement à côté de l'utilité sociale de la production scientifique. Des revues très lues dans les milieux professionnels ou sociaux n'apparaîtront pas dans les classements et ne tenteront donc pas les chercheurs en quête de reconnaissance académique officielle, alors qu'une publication dans ces revues contribuerait sans doute bien plus à leur reconnaissance (et à leur utilité) sociale que celle dans des revues lues par un très petit nombre de lecteurs, quand ce n'est pas pour certains articles par les seuls *referees*. On peut dès lors regretter une position quelque peu schizophrénique des universités, au moment où pourtant elles affirment vouloir promouvoir leur « troisième mission », qui implique nécessairement aussi une ouverture vers le local. Pensons, à l'ULB, aux efforts menés, en coopération avec la VUB et Saint-Louis, pour promouvoir la recherche appliquée sur Bruxelles, articulée

aux demandes du monde politique et concrétisée par la création du BSI (*Brussels Studies Institute*), et au succès et à l'impact social d'une revue scientifique gratuite en ligne comme *Brussels Studies*, revue référée mais évidemment sans facteur d'impact reconnu dans les classements internationaux, bien que très largement consultée. Les universités disent aussi vouloir mettre en avant les qualités pédagogiques.

Pourtant, dans la pratique des commissions de sélection et de promotion, pour qu'un candidat soit retenu ou promu, il faut avant tout avoir publié dans les revues commerciales dites internationales dont il a été question, quitte à ne pas se soucier de l'adéquation de ces publications aux réalités scientifiques et sociétales concrètes, en fonction de ces critères bibliométriques abstraits qui encouragent les candidats à recourir à toutes les « ruses » évoquées ci-dessus pour publier beaucoup au bon endroit plutôt que bien.

QUELQUES RÉFÉRENCES

- BREMBS B., BUTTON K. & MUNAFO M. (2013), « Deep impact : unintended consequences of journal rank », *Frontier in Human Neuroscience*, 7, pp. 1-12.
- DASSA M., KOSMOPOULOS C. & PUMAIN D. (2010), « JournalBase. Comparer les bases de données scientifiques internationales en sciences humaines et sociales (SHS) », *Cybergeo. European Journal of Geography* [En ligne], Dossiers, Les revues en sciences humaines et sociales, cybergeo.revues.org/22864.
- GINGRAS Y. (2008), *La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs*, Note de recherche 5, Centre interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal, Université de Montréal, Université de Sherbrooke.
- GUNESAKARAN S. & ARUNACHALAM S. (2014), « The impact factors of open access and subscription journals across fields », *Current Science*, 107, pp. 380-388.
- KITCHIN R. (2005), « Disrupting and destabilizing Anglo-American and English language hegemony in geography », *Social and Cultural Geography*, 6, pp. 1-17.
- LEONELLI S., SPICHTINGER D. & PRAINSACK B. (2015), « Sticks and carrots : encouraging open science at its source », *Geography and Environment*, 2, pp. 12-16.
- MEADOWS M., DIETZ A.J. & VANDERMOTTEN C. (2016), « Geographical journals : ranking, publishing houses and open access », *Geography and Environment*, 3, 1.
- PAASI A. (2005), « Globalisation, academic capitalism, and the uneven geographies of international journal publishing spaces », *Environment and Planning A*, 37, pp. 769-789.
- PONTILLE D. & TORNY D. (2010), « The controversial policies of journal ratings : Evaluating social sciences and humanities », *Research Evaluation*, 19, pp. 347-360.
- REISSE J. (2011), « L'expertise scientifique et le choix des experts en question », *Revue électronique de la Classe Technologie et Société de l'Académie royale de Belgique*.
- SHORT J.R., BONICH A., KIM Y. & LI LI P. (2001), « Cultural globalisation, global English and geography journals », *The Professional Geographer*, 53, pp. 1-11.
- SLOW SCIENCE ACADEMY (2010), *Slow Science Manifesto*, slow-science.org.
- VANDERMOTTEN C. (2012), « Cotation des revues de géographie, impérialisme scientifique anglo-saxon et culture de l'excellence marchandisée », *Belgeo*, 1-2, belgeo.revues.org/7131.
- VAN LEEUWEN T.N., MOED H.F., TIJSEN R.J.W., VISSER M.S. & VAN RAAN A.F.J. (2000), « First evidence of serious language-bias in the use of citation analysis for the evaluation of national science systems », *Research Evaluation*, 8, pp. 155-156.

Faire une carrière scientifique aujourd'hui

Quelques clés de lecture et critiques

BERNARD FUSULIER

Professeur de sociologie à l'Université catholique de Louvain et maître de recherches du FNRS

Les réflexions sur le monde scientifique et académique sont anciennes et sans cesse renouvelées. Max Weber formulait ainsi dès 1917 le problème du sens de la science pour celui qui a décidé d'en faire sa profession¹. En 1942, Robert-King Merton réfléchissait à la question de l'ethos scientifique (repris en 1973) et Pierre Bourdieu s'interrogeait sur le champ scientifique (1976) et la production d'un *homo academicus* (1984). Dans le même temps, Bruno Latour et Steve Woolgar (1979) questionnait la construction de la science dans le quotidien d'un laboratoire. Toutefois, si la sociologie des sciences est devenue ample et bouillonnante (Gingras, 2013), celle des scientifiques et des carrières scientifiques demeure un domaine de recherche peu développé (Prpic *et al.*, 2014).

Faire carrière² comme chercheur (ou enseignant-chercheur) suppose, hier comme aujourd'hui, un engagement temporel et subjectif important qu'implique l'intégration dans un champ social, c'est-à-dire un espace de luttes pour l'accès à la reconnaissance scientifique et aux positions valorisées selon des critères et capitaux spécifiques à ce champ, dont la définition constitue

¹ Voir sa conférence prononcée à Munich en novembre 1917 et traduite en français en 2005 – Weber, 2005.

² C'est-à-dire, s'engager dans la compétition pour obtenir un poste stable puis participer au champ scientifique en vue d'accéder progressivement aux positions valorisées au sein de ce champ.

un enjeu d'un jeu de pouvoir entre les intervenants (Bourdieu, 1976). Dans une perspective à caractère fonctionnaliste, la socialisation professionnelle des chercheurs conduit à l'apprentissage (l'acceptation et, bien souvent, la reproduction), d'un *ethos* proche d'autres métiers hautement qualifiés (Blair-Loy, 2003 ; Hochschild, 1997). Cet *ethos* qui exige la démonstration d'une vocation et un investissement entier dans la carrière professionnelle (Beaufays, Kraiss, 2005 ; Dany, Louvel, Valette, 2011) émane également d'une *greedy institution* (institution gourmande — Coser, 1974). Il apparaît dans les récits que les chercheurs font de leur travail³ et dans les documents institutionnels où les universités explicitent leurs attentes, à l'image de l'Université catholique de Louvain qui, dans une note interne de son Conseil rectoral datée de 2012 indiquait : « *du professeur totalement investi dans son travail et profitant pleinement de sa liberté académique, on est passé au professeur toujours entièrement investi, mais beaucoup moins libre académiquement, vu l'accroissement de la charge administrative et les nombreuses réformes qui se sont succédé* ». L'université est donc cette *greedy institution*⁴ où la figure du chercheur, répondant à toutes les sollicitations, ignorant les préoccupations familiales, accédant à un poste stable selon un calendrier relativement serré⁵, apparaît comme un « *promotion script* » (Dany, Louvel et Valette, 2011), un étalon auquel se comparent les chercheurs dans l'élaboration de leurs perspectives professionnelles. Elle l'est peut-être encore plus depuis le déploiement, ces dernières années, d'une nouvelle régulation du travail scientifique : mesure de la productivité, exigence de diffusion rapide des recherches, injonction à la coopération et à la mobilité internationales (Ackers, 2010 ; Acker, Armenti, 2004 ; Barry *et al.*, 2001), compétition accrue sur le marché du travail et croissance de la précarité des statuts d'emploi (Ylijoki, 2010), « *accountability* » du travail du chercheur, le tout générant un rapport plus entrepreneurial à la recherche (Lamy, Shinn, 2006).

³ On trouvera en France un exemple récent de l'entretien de cette image dans le récit réalisé par un chercheur du Centre National de la Recherche Scientifique en physique (Balibar, 2014). Sébastien Balibar y présente une activité supposant une disponibilité et réactivité permanente motivée par la quête de la Vérité scientifique.

⁴ C'est ainsi qu'elle est décrite dans plusieurs recherches. Voir par exemple Currie *et al.* (2000), Hendrickson *et al.* (2011).

⁵ Grant *et al.* indiquent ainsi « *in addition to making claims for undivided loyalty, scientific careers in academia also have a normative clockwork, embodying expectations of the benchmarks that should be attained by specific points on a pre-determined timeline* » (2000 : 65).

1. DES INÉGALITÉS ENTRE LES HOMMES ET LES FEMMES

Ces modalités de fonctionnement de l'université produisent des inégalités de genre décrites dans les expressions « *glass ceiling* » (redéfini en « plafond de fer » — Fassa, Kradolfer, 2010) et « *leaky pipeline* » (par ex. Alper, 1993 ; Meulders *et al.*, 2012 ; Dubois-Shaik, Fusulier, 2015). Les causes spécifiques de ces inégalités ne relèvent plus tant d'une discrimination directe et explicite (notamment lors des recrutements — Musselin, Pigeyre, 2008), que d'une *gendered organization* (Acker, 1990) se traduisant dans une gestion de l'université proche de celle d'un *old boy's club* (Case, Richley, 2012 : 14), un effet Matilda (Rossiter, 1995 ; Fassa *et al.*, 2012) pénalisant les femmes devant leurs productions scientifiques (contre un effet Saint Matthieu — Merton, 1969 — pour les hommes) ou encore dans l'injonction à donner toute priorité au travail sur la vie privée, le chercheur apparaissant comme un « *lonely hero* » (Benschop, Brouns, 2003), entièrement engagé dans son travail et donc supposément libéré des contraintes domestiques par un *carer* (ce qui exprime un certain modèle d'articulation travail/famille selon un ordre genré, voir notamment Crompton, 1999). À cet égard, dans la suite et en lien avec d'autres travaux (e.g. Etzkowitz *et al.*, 2000, Marry, Jonas, 2004 ; Case, Richley, 2013), nos recherches montrent qu'une cause supplémentaire des inégalités entre les sexes s'observe dans l'articulation entre le privé et le professionnel, ce que nous avons thématiqué comme un « filtre caché » (Fusulier, del Rio Carral, 2012) bien que, comme le souligne Nicky Le Feuvre (2010), ce phénomène ne suffise pas à lui tout seul pour expliquer l'« évaporation » des femmes au fur et à mesure du parcours professionnel dans le monde de la recherche.

Selon la posture que nous prôtons, le rapport au travail se fabrique dans des dynamiques non seulement professionnelles et sociales (Avril *et al.*, 2010) mais aussi familiales. À partir du cas des chercheurs post-doctorants qui, à la manière de nombreux travailleurs, se meuvent dans un univers compétitif et incertain, nous avons pu reconstituer différentes logiques d'interférence synthétisant des formes d'inter-structuration de la vie professionnelle et de la vie familiale : renforcement, apaisement et recadrage (Barbier, Fusulier, 2015). Si l'on observe une interférence parentalité-travail chez les hommes *et* les femmes, on ne l'observe pas chez les femmes *comme* chez les hommes. Les hommes éprouvent certes eux aussi l'imposition d'une limite par la vie familiale augmentant le coût d'accès à la carrière scientifique et altérant le sens de

l'engagement dans le travail. Toutefois, cette altération ne s'accompagne pas, pour les hommes rencontrés en tout cas, d'une crainte trop prononcée d'imposer à la vie familiale une contrainte professionnelle. Le recadrage porte le plus souvent sur l'investissement dans le travail et s'accompagne rarement d'inquiétudes envers la parentalité. Avec ce constat d'une différence significative dans l'appréciation des coûts de l'interférence entre travail et famille sur l'entourage familial, on retrouve chez les chercheurs un résultat important de la sociologie du genre : la culpabilité du « temps volé » à la famille ou des contraintes supplémentaires imposées au conjoint constitue une « spécificité » largement féminine (Jarty, 2009) produite par l'assignation sexuée de la prise en charge de la vie domestique. Mais le cas des chercheurs révèle un facteur pouvant alourdir ce sentiment de culpabilité : la flexibilité et l'autonomie dans l'organisation du travail (Brannen, 2005 ; Negrey, 2012) qui permettent de décider du lieu et du moment de travail (même si cela est variable en fonction des objets d'études, plus que des disciplines d'ailleurs). Dans un contexte d'incertitude d'emploi, de régulation temporelle du travail faible et d'une injonction sociale à la prise en charge de la vie domestique par les femmes (retraduite quotidiennement dans les relations de travail) peuvent alourdir, chez les femmes, les risques d'être exposées à des rappels à l'ordre domestique et donc, éventuellement, à une forme de tiraillement voire de conflit (Perista, Perista, 2014).

Force est de reconnaître qu'il existe une prise de conscience de ce phénomène par les responsables des grandes institutions scientifiques puisque la Charte européenne des chercheurs intègre la proposition suivante : les employeurs et bailleurs de fonds « devraient viser à fournir des conditions de travail qui permettent aux chercheurs tant féminins que masculins de combiner la famille et le travail, les enfants et la carrière⁶ ». Cette « bonne intention » percole manifestement au niveau plus local ; nous en voulons pour exemples récents, parmi de nombreux autres, la Charte pour l'équilibre des temps de vie signée en septembre 2015 par l'École des hautes études en sciences sociales de Paris ; ou l'inscription en 2015, dans les axes stratégiques 2020 de l'Université de Louvain, d'une politique du personnel qui veille « à favoriser l'équilibre et la conciliation entre vie privée et vie professionnelle ».

Probablement qu'une politique de genre ambitieuse, avec des mesures *family-friendly*, peut soutenir cette quête d'équilibre et combattre les

⁶ http://ec.europa.eu/euraxess/pdf/brochure_rights/eur_21620_en-fr.pdf

asymétries dans la carrière scientifique entre les hommes et les femmes, qui proviennent de raisons qui n'ont rien à voir avec la qualité du travail scientifique. Toutefois, dans le contexte actuel, l'institution scientifique s'est dotée de nouveaux outils d'évaluation qui relèvent de ce que nous appelons un régime comptable-productiviste court-termiste. Or ces outils et ce régime accentuent la pression sur les chercheurs et, par effets de conséquence, renforcent la tension travail/famille ce qui va à l'encontre des intentions proclamées. Il y a par ailleurs un sérieux doute au sein de la communauté scientifique quant au bien-fondé de ce système d'évaluation par rapport à la nature même de la recherche fondamentale. Un pont peut ici être jeté entre une critique du nouveau mode de régulation de la science et un projet d'égalité entre les sexes, le tout au profit de la qualité de la production scientifique.

2. L'ÉVALUATION DE L'EXCELLENCE DANS UN RÉGIME COMPTABLE-PRODUCTIVISTE

Comme nous l'avons dit plus haut, la régulation de la science et des carrières scientifiques s'appuie traditionnellement sur une exigence de disponibilité des chercheurs pour se faire une place durable et reconnue dans un champ très concurrentiel, se montrer impliqués dans une institution gourmande, obtenir des financements et accomplir leurs tâches selon une certaine éthique du travail. Déjà exigeante, cette régulation a en outre intégré explicitement une nouvelle norme : l'excellence scientifique. Norme extrêmement puissante car difficile à contester.

Concept à charge qualitative, l'excellence semble pourtant opérer à partir de critères d'efficacité mesurables qui, en transmutant le « qualitatif » en « quantitatif », favorisent des mises en équivalence des chercheurs, de leurs travaux et des institutions à l'échelle mondiale (Burrows, 2012). Par conséquent, le mode de régulation crée une assimilation entre le *bon* et le *plus* où ce qui est attendu d'un bon chercheur c'est toujours plus de publications, plus de projets, plus de crédits, plus de mobilité... dans un laps de temps le plus court possible. Certes, ce régime comptable-productiviste court-termiste met sous haute tension les chercheurs et induit des comportements productivistes. Mais est-ce pour un mieux ?

Si nous regardons du côté des publications, force est de constater que le climat de grande incertitude et de concurrence croissant où la bibliographie

et sa mesure bibliométrique sont utilisées et perçues comme un critère de jugement de la qualité scientifique et de sélection — pour des financements de projet, des mandats, des promotions, des positions dans les *rankings*, etc. —, génère une course individuelle à la publication qui a parfois pour unique but de « *bodybuilder* » le *curriculum vitae*. En d'autres termes, la mise en équivalence via une logique « métrique » stimule des stratégies de maximisation de signaux comptabilisables, ce qui peut transformer le rapport à la publication :

- formatage du sujet et du protocole de recherche en fonction d'une maximisation du facteur d'impact anticipé plus qu'en fonction d'un intérêt primordial,
- accélération du processus de production et de publication au détriment de la profondeur analytique et d'une réelle avancée dans les connaissances,
- saucissonnage d'articles et multiplication des variations sur un même résultat,
- dilution des progrès réels dans un magma d'informations, etc.

En outre, certains chercheurs n'osent même plus publier en dehors des listes des revues prises en considération par les instances qui les évaluent, alors que leur savoir pourrait irriguer le débat scientifique et la société à travers d'autres canaux de diffusion plus adaptés.

Un autre exemple dans la fabrique du CV scientifique : l'inscription internationale. Tous les chercheurs connaissent et reconnaissent les immenses avantages de la mobilité (formatrice, enrichissante, nouvelles techniques, réseautage...). Cependant, la mobilité (surtout de longue durée) est également appréhendée comme une condition à devoir remplir coûte que coûte uniquement pour compléter la case « mobilité » dans le CV. Alors des ruses diverses sont employées pour contourner cette difficulté (comme déclarer une longue durée et la fractionner par des retours au pays). Une mobilité longue affecte directement la vie privée : on ne s'expatrie pas avec la même facilité quand on est célibataire ou en famille, surtout si on est une femme. Pourtant, ne pas accéder à une mobilité longue n'empêche pas de s'inscrire pleinement dans une dimension internationale par des séjours courts, des échanges skype, des réseaux virtuels..., et de faire de la bonne science en somme.

Le problème ne provient pas uniquement d'une fabrique des CV. L'accent mis sur la productivité pousse des chercheurs à recycler leurs travaux sous de nouvelles formes, tout simplement parce que la recherche fondamentale demande du temps et de l'approfondissement alors qu'ils sont soumis à l'injonction de produire et d'innover rapidement. Se développe ici une rhétorique de la nouveauté...

Ce que nous observons aussi c'est une tension entre un idéal de la recherche fondamentale comme activité créatrice et le régime comptable-productiviste court-termiste. Des chercheurs vivent alors de plus en plus un sentiment d'imposture, tant vis-à-vis du système évaluateur qu'en leur for intérieur. Pour des chercheurs, leurs excellents indicateurs peuvent leur apparaître comme une illusion lorsqu'ils se comparent à d'autres chercheurs à moindre productivité mais dont ils estiment, sans nécessairement le dire publiquement, la valeur de leur intelligence et la profondeur de leurs travaux ; pour ceux-ci, lorsqu'ils se comparent aux premiers, ils peuvent avoir le sentiment de ne pas être à la hauteur de la mesure attendue. Cela pèse sur le moral, crée de l'anxiété pour le futur, certains finissent par se demander s'ils sont vraiment à leur place. Et des chercheurs prometteurs quittent la recherche fondamentale aussi pour cette raison.

Ces quelques exemples d'effets pervers ne sont pas sans interroger la validité du système d'évaluation en place.

3. VERS UNE ALTER-EXCELLENCE ?

La critériologie qui sert à mesurer l'excellence dans le régime comptable-productiviste court-termiste relève davantage des mondes de l'ingénierie, de la gestion, du commerce, de la communication d'entreprise, que de celui de la recherche fondamentale. La liberté de chercher, c'est comme la liberté de créer : l'activité de recherche comporte par définition de grandes incertitudes, et il est même quasi-impossible de juger de la « valeur » d'un résultat, puisque celui-ci se révèle souvent ultérieurement, dans le contexte de l'évolution de la discipline : on ne peut savoir à l'avance ce qui va en sortir et entrer dans l'Histoire. Un retour aux fondamentaux de la recherche fondamentale s'impose. Faite de patience, de rigueur, de bifurcations et de sérendipité, sans parler du rôle de l'erreur dans la fécondité scientifique, elle est quasi

incommensurable. Les critères standards de mesure de l'excellence sont de ce fait peu adéquates.

Dès lors, des scientifiques ont proposé de déconstruire la norme de l'excellence par l'idée de désexcellence⁷, puis d'alter-excellence⁸. Une série d'actions ont été menées, des dispositifs mis en place, dont la *Déclaration de San Francisco* (2013) où des chercheurs critiquent point par point les facteurs d'impact⁹. À Berlin s'est créé l'Institut *Slow-science.org* qui a aussi lancé son manifeste, le *Slow Science Manifesto*, plaidoyer pour que le temps dû à la recherche soit respecté¹⁰. La révision des critères va dans le sens d'une autre définition de l'excellence dont l'enjeu est de cultiver un atout du métier qui perdure au-delà de réformes plus ou moins « liberticides » : l'amour du travail scientifique, qui va de pair avec le sentiment de liberté du chercheur, son désir de comprendre, son goût du « bel ouvrage » et le plaisir qu'il en retire. Le système d'évaluation se doit alors de faire primer la qualité sur la quantité, de soutenir la prise de risque intellectuelle et de tenir compte du rythme de la production scientifique. Simultanément, il permettrait aux chercheurs de mieux contrôler leur vie professionnelle et de l'articuler plus harmonieusement à leur vie privée, et ainsi de favoriser le déploiement de carrières scientifiques plus neutres au plan du genre.

⁷ Voir l'article en ligne d'Olivier GOSSELAIN, *Slow Science — La désexcellence* : <http://www.sisyphes.upmc.fr/~lemoine/docs/La%20Desexcellence.pdf>

⁸ La perspective d'une alter-excellence est défendue par le Comité Femmes et Sciences de la Fédération Wallonie-Bruxelles dans une note remise début 2015 au ministre en charge de l'enseignement supérieur et de la recherche scientifique. La critériologie de l'excellence fait aussi l'objet d'une analyse spécifique dans le projet européen GARCIA (*Gendering the Academy and research combating Career Instability and Asymmetries* — www.garciaproject.eu — Voir Herschberg, et al., 2015).

⁹ Texte complet : <http://www.ascb.org/dora/wp-content/uploads/2015/07/SFDeclarationFINAL.pdf>
Site : <http://www.ascb.org/dora/>

¹⁰ <http://slow-science.org>

BIBLIOGRAPHIE

- ACKER, J., 1990. « Hierarchies, jobs, and bodies : A theory of gendered organizations », in *Gender & Society*, 4, p. 139-58.
- ACKER, S., ARMENTI, C., 2004. « Sleepless in academia », in *Gender and Education*, 16 (1), p. 3-24.
- ACKERS, H.L., 2010. « Internationalisation and Equality. The Contribution of Short Stay Mobility to Progression in Science Careers », in *Recherches sociologiques et anthropologiques*, L (1), p. 55-76.
- ALPER, J., 1993. « The pipeline is leaking women all the way along », in *Science*, 260, p. 409-411.
- AVRIL, C., CARTIER, M., SERRE, D., 2010. *Enquêter le travail*. Paris, La Découverte.
- BALIBAR, S., 2014. *Chercheur au quotidien*, Paris, Le Seuil.
- BARBIER, P., FUSULIER, B., à paraître. « L'interférence parentalité-travail chez les chercheur-e-s en post-doctorat : le cas des chargé-e-s de recherches du Fonds National de la Recherche Scientifique en Belgique », in *Sociologie et Sociétés*.
- BARRY, J., CHANDLER, J., CLARK, H., 2001. « Between the ivory tower and the academic assembly line », in *Journal of Management Studies*, 38 (1), p. 87-101.
- BEAUFAYS, S., KRAIS, B., 2005. « Femmes dans les carrières scientifiques en Allemagne : les mécanismes cachés du pouvoir », in *Travail, genre et sociétés*, 14 (2), p. 49-68.
- BENSCHOP, Y., BROUNS, M., 2003. « Crumbling Ivory Towers : Academic Organizing and its Gender Effects », in *Gender Work and Organization*, 10 (2), p. 194-212.
- BLAIR-LOY, M., 2003. *Competing Devotions*. Cambridge (MA), Harvard University Press.
- BOURDIEU, P., 1976. « Le champ scientifique », in *Actes de la recherche en sciences sociales*, 2(2-3), p. 88-104.
- BOURDIEU, P., 1984. *Homo academicus*, Paris, Éditions de Minuit.
- BRANNEN, J., 2005. « Time and the Negotiation of Work-family Boundaries. Autonomy or Illusion ? », in *Time and Society*, 14 (1), p. 113-131.
- BURROWS, R., 2012. « Living with the h-index ? Metric assemblages in the contemporary academy », in *The Sociological Review*, 60, p. 355-372.
- CASE, S.S., RICHLEY, B.A., 2012. *Barriers to Women in Science: Examining the Interplay Between Individuals and Gendered Institutional Research Cultures on Women Scientists Desired Futures*, Work and Family Researchers Network Conference, New York, June 14.
- CASE, S.S., RICHLEY, B.A., 2013. « Gendered institutional research cultures in science : the post-doc transition for women scientists », in *Community, Work & Family*, 16(3), p. 327-349.
- COSER, L., 1974. *Greedy Institutions*, New York, The Free Press.
- CROMPTON, R., 1999. *Restructuring Gender Relations and Employment: the Decline of the Male Bread Winner*, Oxford, Oxford University Press.
- Currie J., Harris, P., Thiele, B., 2000. « Sacrifices in Greedy Universities : Are they gendered ? », in *Gender and Education*, vol. 12, Issue 3, p. 269-291.
- DEL RIO CARRAL, M., FUSULIER, B., 2013. « Jeunes chercheurs face aux exigences de disponibilité temporelle », in *Temporalités*, 18 [En ligne : <http://temporalites.revues.org/2614>].
- DUBOIS-SHAIK, F., FUSULIER, B., VINCKE, C., 2015. « Belgium », in Herschberg, Channah, Yvonne Benschop and Marieke van den Brink (eds.), *Constructing excellence : the gap between formal and actual selection criteria for early career academics*, Garcia Working Papers, 2, Trento, University of Trento, p. 43-109.
- DUBOIS-SHAIK, F., FUSULIER, B., (ed.) 2015. *Academic Careers and Gender Inequality : Leaky Pipeline and Interrelated Phenomena in Seven European Countries*, Garcia Working Papers, 5, Trento, University of Trento.
- FASSA, F., KRADOLFER, S. (dir.), 2010. *Le plafond de fer de l'université. Femmes et carrières*, Zurich, Seismo.
- FASSA, F., KRADOLFER, S., PAROZ, S., 2012. *Enquête au royaume de Matilda. La relève académique à l'Université de Lausanne*, LIVES Working Papers, 11, Lausanne-Genève.
- FUSULIER, B., DEL RIO CARRAL, M., 2012. *Chercheur-e-s sous haute tension !*, Louvain-la-Neuve, Presses de l'Université de Louvain.
- GINGRAS, Y., 2013. *Sociologie des sciences*, Paris, Presses universitaires de France, coll. « Que sais-je ? ».
- HENDRICKSON, R., MUELLER, M., STRAND, J., 2011. « Policital Science Careers at Comprehensive Universities : Building Careers at 'Greedy' Institutions », in *Political Science & Politics*, n° 44, p. 129-134.

- HERSCHBERG, C., BENSCHOP, Y., VAN DEN BRINK, M. (eds.), 2015. *Constructing excellence : the gap between formal and actual selection criteria for early career academics*, Garcia Working Papers, 2, Trento, University of Trento.
- HOLLANDER, J.A., 2013. « 'I Demand More of People' : Accountability, Interaction, and Gender Change », in *Gender & Society*, 27/1, p. 5-29.
- HOCHSCHILD, A., 1997. *The Time Bind*, New York, Henry Holt and Company.
- JARTY, J., 2009. « Les usages de la flexibilité temporelle chez les enseignantes du secondaire », in *Temporalités*, 9 [En ligne : <http://temporalites.revues.org/1057>]
- GRANT, L., KENNELLY, I., WARD, K.B., 2000. « Revisiting the gender, marriage, and parenthood puzzle in scientific careers », in *Women's Studies Quarterly*, 1-2, p. 62-85.
- LAMY E., SHINN T. (2006). « L'autonomie scientifique face à la mercantilisation », in *Actes de la recherche en sciences sociales*, 4 (164), p. 23-50.
- LATOUR, B., WOOLGAR, S., 1979 (trad. fr. 1988, rééd. 1996). *La Vie de laboratoire. La production des faits scientifiques*, Paris, La Découverte.
- LE FEUVRE, N., 2010. « Les carrières universitaires à l'épreuve du genre : éléments de conclusion », in F. Fassa, S. Kradolfer (dir.), *Le plafond de fer de l'université: femmes et carrières*, Zurich, Seismo, p. 225-242.
- MARRY, C., JONAS, I., 2005. « Chercheuses entre deux passions », in *Travail, genre et sociétés*, 14 (2), p. 69-88.
- MERTON, R., 1969. « The Matthew Effect », in *Science*, 159 (3810), p. 56-63.
- MERTON, R.K., 1973. *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, University of Chicago Press.
- MEULDERS, D., SIMEU, N., O'DORCHAI, S., 2012. *Alma Mater, Homo Sapiens II: Les inégalités entre femmes et hommes dans les universités francophones de Belgique*, Département d'Économie appliquée, Université libre de Bruxelles.
- MUSSELIN, C., PIGEYRE, F., 2008. « Les effets des mécanismes du recrutement collégial sur la discrimination : le cas des recrutements universitaires », in *Sociologie du travail*, 50, p. 48-70.
- NEGREY, C., 2012. *Work Time, Conflict, Control, and Change*, Cambridge, Polity Press.
- PERISTA, H., PERISTA, P., 2014. *Mapping the population, careers, mobilities and impacts of advanced degree graduates in the social sciences and humanities (POCARIM), The Impact of Partnering, Parenting and other Caring Responsibilities on SSH Work and Careers*, Policy Report, 12, CESIS and the University of Salford.
- PRPIC, K., VAN DER WEIJDEN, I., ASHEULOVA, N., 2014. *(Re)searching Scientific Career*, IHST/RAS/SSTNET-ESA, St. Petersburg, Publishing House Nestor-Historia.
- ROSSITER, M., 1993. « The Matthew Matilda Effect in Science », in *Social Studies of Science*, 23 (2), p. 325-341.
- YLIJOKI, O.H., 2010. « Future orientations in episodic labour : Short-term academics as a case in point », in *Time & Society*, 19 (3), p. 365-86.
- WEBER, M., 2005. *La Science, profession et vocation* (conférence de novembre 1917), trad. fr. de Kalinowski I., Marseille, Éditions Agone.

La méthode BVLM

Formes et enjeux des modes d'évaluation de la recherche

LAC

L'Atelier des Chercheurs pour une désexcellence
des universités — ULB

Ce texte est à peu de chose près celui qui fut présenté oralement lors du séminaire « Penser la Science » du 9 mai 2015. Il est le fruit de la réflexion d'un collectif d'enseignants-chercheurs de l'ULB réunis depuis 2011 au sein de l'Atelier des Chercheurs (LAC). Notre action consiste à comprendre et à s'opposer aux dérives contemporaines de la recherche et de l'enseignement universitaires. Dans ce contexte, nous avons notamment rédigé et diffusé une *Charte de la désexcellence*¹.

Pourquoi prônonons-nous la désexcellence ? Parce que, conjointement à l'appel obsessionnel à l'Excellence de la recherche (comme de tout, par ailleurs), nous voyons s'imposer dans nos métiers une **nouvelle raison évaluatrice (NRE)**, qui, à nos yeux, a des traits inquiétants puisqu'elle mine les fondements mêmes de la recherche, telle que nous l'envisageons, c'est-à-dire une recherche qui soit libre, plurielle et engagée.

- La fonction cardinale de cette NRE est d'assurer la mise en concurrence et la compétition généralisées des chercheurs et de leurs productions, sur un marché mondial de la connaissance.
- La sélection qui s'opère sur ce marché compétitif repose sur des critères rarement explicites mais qui relèvent essentiellement de

¹ Notre charte, ainsi que d'autres outils et textes, sont accessibles sur le site du LAC : <http://lac.ulb.ac.be/LAC/home.html>

la performance et de la productivité des chercheurs, des centres de recherche et des universités.

- La mise en œuvre de ces critères de performance et de productivité se réalise à travers le quadrillage de nos pratiques par des dispositifs de marketing, d'évaluation permanente et de classement.
- Comme il s'agit d'établir la productivité de chercheurs, centres et universités, qui sont conçus comme indifférenciés, la nouvelle raison évaluatrice conduit à réduire l'activité scientifique à ce qui peut être MESURÉ, et disqualifie *de fait* le recours à tout autre critère de jugement. Les grilles d'évaluation standardisées, dont les items sont rarement explicités et jamais discutés, en sont le marqueur le plus frappant.

Chacun sait qu'une pratique, quelle qu'elle soit, est toujours transformée par les instruments qui ambitionnent de la mesurer. Notre récit se propose donc de rendre compte des transformations des pratiques d'évaluation et de ses effets sur la recherche au cours des 20 dernières années, au travers du prisme de situations réellement vécues par nous et qui nous ont souvent laissé-e-s sans voix. Nous avons ponctué notre intervention de dialogues vécus. Ainsi, notre intervention a pour ambition de proposer des clés de lecture de cette NRE, de ce qui la fonde et de ce qu'elle façonne comme milieu professionnel.

L'instrument que nous avons choisi pour décrire et comprendre cette raison évaluatrice, c'est celui de la polémique et du décalage ; notre grille de lecture sera un slogan repéré sous un pont d'autoroute à Drogenbos : « Beaucoup, vite, loin, mal » ou pour faire vite la « méthode BVLM ».



Fig. 1 Mur du pont du Ring de Bruxelles, Drogenbos

Voyons, suivant cette grille, comment les nouvelles modalités de l'évaluation affectent nos pratiques de recherche, tant dans leur quantité, leur temporalité, leur spatialité que leur qualité ; comment elles nous affectent en tant qu'individu évalué et évaluant, et comment plus généralement elles affectent la recherche elle-même.

1. LA MÉTHODE BVLM DANS L'ÉVALUATION DE LA RECHERCHE

Beaucoup

Toutes les grilles d'évaluation auxquelles nous sommes soumis, nous invitent à « produire » beaucoup. La productivité du chercheur est au centre des dispositifs d'évaluation. Si bien que certaines pratiques qui permettent de se conformer au mieux à ce modèle productiviste ont fait leur apparition.

Donnez un marteau à un enfant et il en frappera tout ce qu'il rencontre. Mettez l'outil bibliométrique au service du nouveau management universitaire et celui-ci l'appliquera uniformément à tout ce qu'il touche. Sous son influence, le désormais ancien « publish or perish » s'est mué en frénésie publicationnelle. À l'instar des chaînes d'information continue, les chercheurs sont invités à publier sans trêve.

La tactique du « one idea, one paper », qui consiste à saucissonner les résultats d'une recherche en plusieurs publications en lieu et place d'articles de synthèse reposant sur des recherches à plus longue échéance, était déjà connue et pratiquée. Pourquoi, en effet, n'écrire qu'un seul article alors qu'on peut en comptabiliser deux ou trois ? La voilà qui connaît un nouvel emballement : on parle désormais de « one idea, ten papers ». Il ne s'agit plus de saucissonner, mais de répéter. Écrire plusieurs fois la même chose, mais pour des publics légèrement différents, en variant l'habillage (le titre, l'introduction), éventuellement dans différentes langues... Un phénomène qui est à la fois exploité et renforcé par les éditeurs scientifiques privés, engagés dans une multiplication des titres de revues et des champs de recherche.

On voit aussi apparaître des articles « programmatiques » annonçant une recherche, une méthode, sans aucun résultat : « no idea yet, a paper already » ; ou encore des articles dont certains auteurs n'ont pas participé à la recherche mais ont monnayé leur signature sous la forme d'un billet d'avion

par exemple : « no idea, cashed paper ». Des sociétés commerciales se sont même spécialisées dans le soutien à la rédaction d'articles scientifiques, garantis sans plagiat.

Une autre pratique courante est celle du « zombie ». Vous prenez un débat clôturé et vous en relancez les arguments les plus contestables dans une publication que l'on justifiera éventuellement comme « thought-provoking ». Dans les mois qui suivent, les réactions de vos pairs contribueront inmanquablement à faire grimper votre indice de citation.

Nous sommes désormais évalués « au poids ». Beaucoup de citations, beaucoup de publications, beaucoup de congrès, beaucoup de promotions de thèses à son actif, beaucoup de membres dans son centre, beaucoup de chercheurs intérimaires, beaucoup de projets, beaucoup d'argent extérieur... Un usage qui réveille d'étranges pratiques cannibales :

« Au fait, Jean-Michel, j'ai appris qu'Emile était le co-promoteur de votre gros projet lié à la création de la Nouvelle Maison de l'Excellence... »

« C'est exact, il est très enthousiaste pour ce projet, on va faire du bon boulot ensemble ! »

« Justement, il faudrait alors qu'on s'arrange pour se partager ses publications... car s'il rejoint votre projet il ne pourra plus apparaître dans les productions de mon centre... et donc, pour les évaluations ça risque de poser problème : tu te rends compte un peu du manque à gagner pour les statistiques du centre ? »

« Pierre, je t'avoue que je n'avais pas pensé à cela : c'est un argument de poids !... »

Vite

Le second principe de la méthode BVLM est le « Vite », forcément indissociable du « Beaucoup » car pour faire davantage dans le même laps de temps, il faut forcément faire vite, et ce dans toutes les temporalités : celles de nos carrières, de nos mandats, de nos jours (mais aussi de nos nuits et week-ends).

Produire beaucoup et vite, signifie aussi produire jeune, et donc si possible déjà avant la thèse. Ensuite, au cours des années de thèse, il est opportun de publier aussi, tout en faisant son doctorat en 4 ans, ou mieux en 3 ans... pour pouvoir publier encore la 4^e année, surtout s'il reste une année de financement. Une thèse excellente n'est donc plus le fruit d'une patiente maturation intellectuelle. C'est une, ou plutôt un ensemble de publication(s).

Surtout, il faut *communiquer* vite. Les concours « Ma thèse en 180 secondes » qui se sont multipliés ces dernières années expriment la quintessence de ce principe. Mais au-delà du grand show auquel sont convié-e-s doctorantes et doctorants de toutes disciplines, quel est l’avenir qui leur est proposé dans le monde académique ? Pour la plupart d’entre eux, aucun².



Fig. 2 Appel à candidature du concours « Ma thèse en 180 secondes »
<http://www.u-bordeaux.fr/mt180s>

Pour ce qui est de la temporalité des jours, cela signifie que nous nous retrouvons sans cesse dispersés, soumis à des frictions insolubles entre différents types de tâches qu’il nous faut cumuler, additionner, multiplier : publier, communiquer, réseauter, projeter, évaluer, réunir, administrer, réformer, encadrer... enseigner.

« Dis, il faudrait qu’on se voit de façon urgente pour finir notre papier commun. T’aurais pas 5 minutes d’ici la fin de la semaine ? »

« Euh... Pour tout te dire, ça va être compliqué. Tu vois, pour demain je dois envoyer à la faculté le projet de réforme de notre master. Demain, pas de chance, je suis à Paris pour la soutenance d’une thèse, que je n’ai d’ailleurs pas encore terminé de lire. Jeudi, je donne cours toute la journée et j’ai un conseil de département sur l’heure de midi. Vendredi matin, j’ai un comité d’accompagnement du gros projet européen dont je te parlais la semaine passée. J’enchaîne ensuite coup sur coup 3 réunions avec des mémorands. Sans parler de ce foutu

² Sylvestre HUET, « Ma thèse en 180 secondes. Et après ? Rien », in *Libération*, 30 juin 2014.

règlement d'ordre intérieur que je dois aussi rédiger. Pour notre article, je te propose donc de nous retrouver vendredi en fin de journée. 18h dans mon bureau. C'est OK pour toi ? »

Les conséquences du Beaucoup et du Vite ? Parmi d'autres :

- L'impossibilité de développer l'interdisciplinarité, car celle-ci impose de maîtriser la langue et les concepts d'autres cultures de recherche ;
- La tentation de reproduire du familier plutôt que de prendre des risques ;
- La production d'indicateurs idiots, sans plus aucun lien avec le contenu.

Loin

La raison évaluatrice aujourd'hui dominante modifie aussi nos pratiques de recherche dans leurs spatialités. En un mot comme en cent, il faut être IN-TER-NA-TIO-NAL, toujours et partout. Et le faire savoir.

International dans nos trajectoires personnelles de chercheur-e-s, d'abord. Un CV qui montre qu'on a obtenu ses diplômes de BA, de MA et de doctorat dans la même université devient suspect. Et sans post-doc à l'étranger, interdiction de postuler à certains postes. La pression s'intensifie même sur les étudiants, au travers de la sur-promotion des échanges Erasmus en particulier.

International dans nos modes de collaboration, de partage et de diffusion de nos recherches, aussi. Présenter un papier à un colloque à Los Angeles ou à Tokyo par an devient une norme, même si c'est pour s'y retrouver entre compatriotes poussés pour les mêmes raisons à faire le même déplacement. Les collègues, les étudiants ou les organisateurs de séminaires locaux, eux, devront attendre que nous soyons rentrés.

International dans nos pratiques de publication, encore. Il vaut toujours mieux publier loin, c'est-à-dire en anglais et dans une revue contrôlée par une multinationale de l'édition scientifique. Même si celle-ci, ravie de posséder ainsi un monopole sur les « revues qui comptent », ne se privera pas en retour de faire payer le prix fort aux lecteurs et aux universités. Les revues plus locales et indépendantes, dédaignées par la raison évaluatrice, permettent pourtant, souvent, d'atteindre un lectorat réel, important, intéressé et bien plus large.

International, également, dans nos choix d'objets de recherche. Envisager de postuler à une ERC, par exemple, demande de « bien » choisir son objet : sans dimension transnationale, aucune chance.

International, toujours, dans nos recrutements. Comme nous le savons tous, un candidat qui vient de loin est nécessairement plus excellent qu'un candidat d'ici.

International, enfin, dans nos engagements sociaux. Là aussi, tout le prestige va aux grandes causes mondiales. Les luttes ancrées dans des territoires locaux, ici et maintenant, où la réflexion s'enrichit de l'action concrète, elles, attendront.



Fig. 3 L'excellence vient toujours d'ailleurs

Bien sûr, nous ne sommes pas bêtement « contre » toute dimension internationale. Nous n'avons que faire des frontières dès lors qu'il s'agit de dialoguer avec des collègues, de confronter nos idées à d'autres contextes, de partager nos questionnements et nos résultats. Mais la nouvelle raison évaluatrice, elle, n'a que faire de tout cela. Pour elle, l'international est un fétiche : franchir des frontières devient gage de qualité *en soi*. L'affichage international prime sur tout le reste, et peu importe que l'on ait passé son « séjour de recherche à l'étranger » entouré d'autres expatriés, eux aussi

délaissés par l'institution d'accueil bien davantage préoccupée de gonfler ses statistiques « d'attractivité internationale » que de veiller à la qualité réelle des échanges.

Mal

L'objectif annoncé de la nouvelle raison évaluatrice promet des lendemains qui chantent. Il s'agirait en effet d'améliorer sans cesse la qualité de la production scientifique. À bien y regarder, pourtant, par sa fonction première — la mise en concurrence des chercheurs — et ses outils — la mesure de la productivité —, elle nous pousse concrètement à Mal faire notre travail. En ce sens, l'évaluation contemporaine produit ce que le sociologue Christian Morel appelle une « erreur radicale et persistante » : une suite de décisions collectives qui produisent un résultat exactement inverse à celui qui est escompté³. Au nom de l'Excellence, c'est une bouillie de plus en plus insipide et parfois toxique que nous sommes amenés à produire.

Contraints de produire « Beaucoup » et de plus en plus « Vite », nous sommes désormais engagés dans une course sans fin. Nous pouvons certes nous appuyer sur des outils qui accélèrent notre travail : des ordinateurs, des logiciels et des moyens de communication de plus en plus puissants... Reste que l'élaboration de connaissances nouvelles repose aussi sur l'imagination, la curiosité, l'échange des idées, la réflexion, le rêve... toutes activités qui demandent du *temps*. Paradoxalement, en dépit des nouvelles technologies, nous en perdons, car comme l'a montré le philosophe Hartmut Rosa⁴, la croissance du nombre de tâches que nous accomplissons dépasse l'accélération de la réalisation de ces tâches.

Par ailleurs, faire Vite ouvre une voie royale à la malhonnêteté scientifique. À moins d'être aveugle ou sourd, on ne peut que constater l'augmentation des fraudes scientifiques, par falsification de données ou par plagiat, surtout dans le chef des chercheurs... masculins. Une étude publiée dans *PNAS (Proceedings of the National Academy of Sciences — USA)*, montre par exemple, sur base d'une analyse de la principale base de données bibliographiques en sciences du vivant (PubMed), que la proportion d'articles rétractés

³ CHR. MOREL, *Les décisions absurdes. Sociologie des erreurs radicales et persistantes*, Paris, Gallimard, 2002.

⁴ H. ROSA, *Aliénation et accélération. Vers une théorie critique de la modernité tardive*, Paris, Éditions de la Découverte, 2012.

a été multipliée par dix entre 1975 et 2012 et que deux tiers des rétractions relèvent de la fraude scientifique⁵. Une autre étude, réalisée sur base des données collectées par l'Office of Research Integrity (USA), souligne pour sa part une nette surreprésentation des hommes parmi les fraudeurs dans le domaine des sciences du vivant⁶.

Une évaluation fondée pour l'essentiel sur des indicateurs quantitatifs a aussi pour effet pervers de nous encourager à n'être attentif qu'aux performances obtenues à l'aune de ces indicateurs. En ce sens, elle nous conduit à concevoir nos travaux uniquement en fonction des « places à gagner ». Comment faire croître mon *h-index*? Comment allonger facilement ma liste de publications? Comment améliorer ma/nos positions dans les classements? Dans cette optique compétitive, peu importe ce que l'on produit, du moment que ce qui est produit permet de rester dans la course... si possible en faisant par la même occasion le *buzz* médiatique.

Cette logique quantitative encourage également le rétrécissement et l'homogénéisation des domaines et des objets de recherche, et ce malgré l'explosion du nombre de chercheurs au cours des dernières décennies. Pour rester compétitif, il vaut mieux ne pas prendre de risque, ne pas s'écarter des chemins balisés, car le tâtonnement, la spéculation, les erreurs font perdre du temps et des points dans le CV. Cela aura des conséquences catastrophiques à long ou moyen terme en matière de développement et d'innovation.

Compte tenu de la primauté de la performance sur le contenu, le montage des dossiers repose dorénavant sur des principes stratégiques d'ajustement aux exigences (réelles ou supposées) des commissions d'évaluation et aux modes et marottes du moment. À l'arrivée, nous engendrons une profusion de chercheurs indistincts, interchangeableables, jetables...

⁵ F.C. FANG, R.G. STEEN ET A. CASADEVALL, « Misconduct accounts for the majority of retracted scientific publications », in *Proceedings of the National Academy of Sciences of the United States of America*, 16 Octobre 2012, n° 109. <http://www.pnas.org/content/109/42/17028.full?linkType=FULL&resid=109/42/17028&journalCode=pnas>

⁶ F.C. FANG, J.W. BENNETT ET A. CASADEVALL, « Males are overrepresented among life science researchers committing scientific misconduct », *mBio* 4:1-3, 2013 (<http://mbio.asm.org/content/4/1/e00640-12>).



Fig. 4 Des doctorant-e-s interchangeables, jetables ?
(Image extraite de Pink Floyd, *The Wall*, GB, 1982, 91 min)

« Tu te souviens de Lola ? Cette doctorante qui travaillait sur la nostalgie de l'intime ? »
« Ah, non, ça c'était Seb. Le gars d'à-côté. C'est pas plutôt de Laura dont tu parles ? Elle faisait un post-doc sur l'intimité nostalgique. »
« Non, non : l'intimité nostalgique c'était le sujet de thèse de Gilles ! Mais c'est pas loin : Lola travaillait sur le rôle de la nostalgie dans les mondes post-intimes »
« Hmm, possible. Mais pourquoi tu parles d'elle ? »
« Je l'ai croisée l'autre jour. Elle vient d'ouvrir un magasin de vêtements de seconde main. »

Faire de la mauvaise recherche donc. Faire de mauvaises évaluations, tout autant.

La transformation des formulaires d'évaluation, par exemple, est évidente : bardés de leurs grilles, les formulaires *new look* nous condamnent à produire un jugement terriblement réducteur sur les candidat-e-s et leurs projets. Ils nous forcent à être triviaux, en répondant par oui ou non, en utilisant des échelles ordinales de type A, B, C, ou des poncifs de type SWOT, et en limitant drastiquement la longueur autorisée des commentaires libres. Les grilles étouffent les différences qualitatives et ne sont en mesure que de saisir du déjà connu. Elles engendrent ainsi la déresponsabilisation de ceux/celles qui les remplissent.

Par ailleurs, la séparation croissante entre les lieux de la recherche et les lieux de l'expertise, avec pour seuls liens des grilles à remplir en ligne,

favorise la paranoïa et la désévaluation. Au nom de la visibilité et de l'équité, l'opacité n'a jamais été aussi grande.

« Tu sais comment remplir ces nouvelles grilles, toi ? »

« Ben, ça dépend : tu espères que la personne passe ou non ? »

« Qu'elle passe. Je crois vraiment que c'est un bon dossier. Et puis c'est quelqu'un de chouette, avec qui on a envie de travailler. »

« Alors tu mets des A partout ! »

« OK. Mais comment je les justifie ? »

« Oh, le bla-bla habituel : cohérence... gnagnagna... haut niveau d'excellence... gnagnagna... profil international... etc. De toute façon il n'y a presque pas de place dans les cases. »

« Mais tu crois pas que ça va faire un peu trop tous ces A ? »

« T'as qu'à mettre un A' quelque part... Genre pour « responsabilité de gestion » : tout le monde s'en fout »

La raison évaluatrice contribue aussi à NOUS FAIRE DU MAL.

Pour tenter de rester dans la course, nous allongeons notre temps de travail, réduisons nos heures de sommeil comme celles consacrées à nos proches, à nos amis ou à nos loisirs. Parallèlement, le sens même de notre métier s'érode. Le plaisir et la créativité font place aujourd'hui à la détresse. Détresse d'autant plus forte que les logiques mêmes de l'évaluation supposent que nos performances s'améliorent continuellement et que le « bien faire » n'est jamais stabilisé. Inévitablement nous sommes amenés à nous demander : « sommes-nous encore bons à quelque chose ? ».

Pour les cohortes de jeunes collègues cette détresse se double trop souvent de la plus rude des sanctions : celle de la précarisation puis de l'éviction du monde académique, en dépit de l'importance du temps et de l'énergie consacrés à la recherche.

Fatigue et perte de sens produisent de la souffrance au travail. Dans le monde universitaire, les symptômes les plus répandus en sont la dépression et le burn-out... bientôt, peut-être aussi, le suicide ? En la matière les témoignages se multiplient. Les services médicaux en charge du personnel universitaire constatent d'ailleurs une croissance de la prise des antidépresseurs, des brusques variations de poids, des bouffées délirantes ou des addictions féroces.

La méthode BVLM appliquée tout azimut pèse sur nos « manières de faire », guette chacun de nos pas : que ce soit dans l'évaluation de nos « carrières », dans l'évaluation de nos projets (qu'on nous presse à produire), dans l'évaluation de nos enseignements, dans l'évaluation de nos

programmes, dans l'évaluation de nos centres de recherche. Une évaluation permanente qui, en privilégiant le quantitatif, fait du mal à nos esprits et à nos corps... et donne naissance à un problème de santé publique.

2. AUX ORIGINES D'UNE ÉVALUATION SELON DES PRINCIPES ET INDICATEURS DE PERFORMANCE

Il ne s'agit pas seulement de recenser les transformations des modes d'évaluation et d'exprimer l'intensité avec laquelle elles affectent nos pratiques. Nous cherchons également à comprendre ce qui les nourrit et les rend si naturellement désirables.

Il nous semble que trois dimensions du système universitaire connaissent des mutations radicales, offrant l'environnement favorable à une conversion collective à la quête de l'excellence. La première est géo-démographique, la deuxième concerne les modalités de l'action publique, et la troisième est idéologique.

Premièrement, la population des étudiants, et, par conséquent, des cadres universitaires a connu une croissance fulgurante depuis les années 1980. Cela a engendré un phénomène de dévaluation relative des diplômes universitaires sur les marchés du travail, ceux-ci devenant de plus en plus nombreux et donc communs. Pour les étudiants désireux de se distinguer de la masse, il devient alors logique de chercher à se donner un parcours universitaire original, ce qui signifie concrètement la multiplication de diplômes, à décrocher dans des institutions de pays différents. Ce n'est désormais plus le diplôme à lui seul qui fait office de sésame professionnel, mais bien la combinaison de passages par des universités, si possible les plus prestigieuses. Pour les universités elles-mêmes, l'enjeu devient alors d'attirer ces étudiants de plus en plus mobiles. L'excellence — et tous les packagings qui peuvent être inventés pour la vendre — est alors non pas la qualité dorénavant supérieure d'une recherche mieux menée, mais le signal-prix censé attirer les meilleurs clients du marché de l'enseignement supérieur. Le phénomène s'amplifie d'autant plus que le processus de Bologne a transformé les universités européennes en un grand marché à la monnaie équivalente — l'ECTS — sur lequel des offreurs tentent d'attirer des demandeurs par l'excellence supposée — et standardisée — de leurs produits.

... High teaching standards

With its 13 faculties, school and institute, the ULB covers **all disciplines**, closely combining teaching and research. From natural sciences to applied sciences, from humanities and social sciences to health sciences, courses are offered at all levels, with over 40 Bachelor programmes, 245 Master programmes (Masters and Specialized Masters) and over 20 doctoral schools involving over 1800 PhD students...

Evening courses, lifelong learning, validation of professional experience - These courses targeting new publics - working professionals - are offered alongside programmes that are more traditional or involve cutting-edge research.



... A tradition of excellence

Four Scientific Nobel Prizes (Jules Bordet, Albert Claude, Ilya Prigogine and the recent Nobel Prize in Physics awarded to **François Englert**, Emeritus Professor, in 2013), the Nobel Peace Prize (Henri La Fontaine), one Fields Medal, three Wolf prizes, two Abel prizes, and numerous other awards bear witness to the University's **long tradition of excellence**.

François Englert, Nobel Prize in Physics 2013 and Emeritus Professor at ULB

... Ground-breaking research

Actively involved in the **European Research Area**, the ULB coordinates or participates in many Belgian and European projects and networks. The university has been awarded 17 European Research Council (ERC) grants in a range of domains (neuroscience, cancer, physics, mathematics, artificial intelligence, information technology, earth sciences, economics, political science, sociology). Moreover, its Institute for European Studies is recognized as a **Jean Monnet European Research Centre** for its work on European integration. A key player in the **regional development** of both Brussels and Wallonia, the ULB is active in finding business applications for the results of its research; it manages a portfolio of 170 families of patents, 31 active spin-offs, 3 incubators, an investment fund and a biotechnology business park, the **Biopark Charleroi Brussels South**.

... A hospital network

In the health sector, alongside the **Erasmus teaching hospital**, the ULB continues to develop a vast hospital network in many cities.

... Internationalization, a reality

In addition to its many student exchange programmes, the ULB has set up **integrated training programmes** with foreign partners, which can lead to **double Master and PhD diplomas**. Several of these programmes have received the prestigious "**Erasmus Mundus**" accreditation.

Fig. 5. Le signal-prix dans une brochure de l'Université libre de Bruxelles

Une deuxième transformation structurelle s'observe dans les modalités de l'action des autorités publiques, en charge de la recherche. Dans un contexte d'austérité budgétaire prolongée, de médiatisation intense des modalités de la gestion publique, et d'une exacerbation de la concurrence électorale, les responsables politiques sont amenés à privilégier des retours sur « investissements décisionnels » à court terme et fortement visibles. L'exigence de plus en plus pressante d'une *accountability* de l'action publique (c'est-à-dire la pression à rendre des comptes clairs — et convaincants — aux citoyens) amène à privilégier, dans le domaine de la recherche, l'affectation des ressources aux « champions » déjà confirmés, aux stakhanovistes de la publication et aux chevaux qui ramènent le plus de prix. La recherche de longue haleine, ou celle qui travaille sur des problématiques ou avec des populations médiatiquement marginales, bref celle qui ne peut ou ne veut se vêtir des paillettes de l'excellence, cette recherche-là rapporte peu aux politiques en quête de rentabilité électorale. L'empressement de nombreux ministres à venir poser sur la photo aux côtés de François Englert lors de son retour à Bruxelles avec son prix Nobel a quelque chose, pour le moins, d'anachronique : l'environnement institutionnel dans lequel il a produit les travaux récompensés était complètement différent de celui que ces ministres incarnent aujourd'hui, un

environnement dont l'intéressé lui-même dit qu'il ne pourrait y survivre, au regard des critères actuels de l'excellence...

Enfin, les nouvelles modalités de l'évaluation de la recherche universitaire sont également promues pour des raisons idéologiques. Ce qui est en jeu ici, c'est le passage d'un État-Providence à une société organisée selon les principes du néo-libéralisme. Concrètement, cela signifie qu'est progressivement — et plus ou moins explicitement — remis en cause le principe d'une pérennité du financement public de la recherche. Ce qu'on observe, c'est la diffusion massive de l'idée d'une contractualisation de ce financement : ce qui est dénoncé par certains comme une « rente » de la recherche universitaire devrait à terme disparaître pour laisser aux chercheurs et à leurs équipes la tâche de décrocher eux-mêmes les budgets qu'ils réclament. D'où l'intérêt d'être excellents, selon les critères en vigueur ; d'où l'intérêt, donc, de faire de la recherche à grande vitesse sur du court terme, en grande quantité, reconnue dans le monde entier, et de manière standardisée. En suivant cette logique, c'est un certain intérêt idéologique qui est en même temps servi. D'aucuns l'expriment d'ailleurs explicitement, en indiquant que ce qui distingue un *Centre d'Excellence* d'un quelconque autre centre de recherche, c'est sa capacité, à long terme, à se passer de financement public.

Action pour des "centres d'excellence" de dimension européenne

I - Qu'entend-on par centres d'excellence en RDT ?

Il s'agit d'une notion intuitive difficile à définir...

Dans quasiment tous les domaines et disciplines, l'Europe dispose de centres publics ou privés faisant de la recherche et du développement technologique (RDT) de très haut niveau, et souvent de niveau mondial. On peut identifier intuitivement de tels "centres d'excellence" par le fait qu'ils comprennent et attirent d'excellents chercheurs et développeurs, et qu'ils acquièrent la réputation d'être une ressource importante pour le progrès scientifique et technologique, et la diffusion de l'innovation. Pour montrer le rôle que les centres d'excellence peuvent jouer, il suffit de citer, aux États-Unis, l'université de Stanford, qui a créé la Silicon Valley, le MIT (avec les retombées "route 128"), ou l'université de Princeton. En Europe, on peut citer le magnifique exemple du CERN.

La notion de centres d'excellence est interprétée et utilisée de multiples façons en Europe. On pourrait proposer la définition simple suivante : "Un centre d'excellence est une structure dans laquelle la RDT est de niveau mondial en ce qui concerne la production scientifique mesurable (notamment la formation) ou les innovations technologiques". De toute manière, il semble possible d'énumérer un certain nombre des caractéristiques principales que cette notion devrait englober :

- une "masse critique" de scientifiques ou de développeurs technologique de haut niveau;
- une structure bien définie (reposant le plus souvent sur des structures existantes) possédant un calendrier de recherche propre;
- un centre d'excellence doit être capable d'intégrer des domaines liés et d'associer des savoir-faire complémentaires;
- il doit être capable de maintenir un niveau élevé d'échange de ressources humaines qualifiées;
- il doit jouer un rôle dynamique dans le système d'innovation dans lequel il se trouve (ajouter de la valeur à la connaissance);
- il doit posséder un niveau élevé de visibilité internationale et de connectivité scientifique ou industrielle;
- un centre d'excellence doit présenter une stabilité acceptable au niveau financier et au niveau des conditions d'exploitation à long terme (base sur laquelle on peut investir dans du personnel et construire des partenariats);
- les sources de financement ne doivent pas dépendre à long terme des pouvoirs publics.

Fig. 6. Une définition des centres d'excellence publiée par le European Research Area (ERA), dans une publication vouée à la promotion de la recherche et du développement technologique
http://ec.europa.eu/research/era/index_en.htm

3. CONCLUSIONS

Nous avons assisté ces deux dernières décennies à l'émergence d'un nouveau « régime d'évaluation » — pour reprendre l'expression de Nicolas Dodier⁷—, un régime dont nous contestons les principes et les outils qu'il mobilise, comme nous déplorons les effets qu'il produit sur la recherche et le sens même de la communauté des chercheurs.

Ceci ne signifie nullement que nous sommes réfractaires à toute forme d'évaluation, ce qui serait assez absurde dans un monde d'activités fondées sur la permanence des échanges, des confrontations et des controverses. Simplement, nous nous opposons à une politique de l'évaluation s'appuyant sur l'application de critères artificiels, standardisés, transversaux et conçus selon la logique, externe, du « management de la qualité ». Nous serions bien plus favorables à un régime d'évaluation collégiale, commandé par les exigences internes de la recherche et animé par le souci heuristique de la faire progresser tout en cherchant à donner vitalité, par le dialogue, à une pluralité de styles de recherche, à une diversité épistémologique — une « épistémodiversité ».

Ce régime d'évaluation collégiale doit reposer selon nous sur la construction permanente, jamais totalement achevée, de référents communs propres à un champ de recherches et non pas sur des critères externes que des experts éloignés viendraient imposer sans discussion. À ce propos, il est piquant de constater que les partisans de l'évaluation de l'excellence définissent généralement leur dispositif d'« expérimental », ouvert à la critique, mais cette ouverture ne s'applique en général qu'aux marges, sur certains critères, et ne tolère aucune discussion sur le cadre lui-même. C'est la logique que l'on a vu opérer lors de l'évaluation de la recherche comme certains d'entre nous l'ont vécue. Rappelons-nous le discours des autorités : « Ne craignez rien il ne s'agit que d'une expérience et elle doit vous être utile » ; verdict final : aucun effort de compréhension de la part des experts externes, aucune discussion sur le fond. À une collègue qui s'est permise de demander au chef de l'escouade experte sur quel référentiel, sur quelle base, il comptait mener la discussion évaluative, la réponse a fusé : « Madame, ici c'est nous qui posons les questions ! ». Une interaction humiliante, un manque de respect pour le moins ; et

⁷ N. DODIER, « Penser un régime d'évaluation de la recherche scientifique », mai 2009, in <http://evaluation.hypotheses.org/452>.

de la part des autorités universitaires une potentielle manipulation, en tout cas un procédé vécu comme pervers par beaucoup des chercheurs évalués.

À l'inverse de cette expérience, les évaluateurs du régime collégial s'efforceraient de comprendre de l'intérieur la pertinence et l'originalité des recherches et des parcours scientifiques, plutôt que de comptabiliser les publications qui en résultent. Ils s'ouvriraient à la pluralité des styles de recherche plutôt que de leur appliquer un principe d'équivalence qui en arase les spécificités respectives.

On voit que ce qui se joue dans les politiques d'évaluation, c'est également un certain sens de la communauté scientifique. Quand il s'agit d'établir des critères communs en phase avec un certain esprit de la recherche et au-delà des épistémologies locales, c'est l'appartenance à une communauté scientifique qui s'éprouve. Ce n'est pas la finalité poursuivie par l'application des grilles.

La construction collective de référents communs, l'attention portée au sens et aux apports des recherches sont affaire exigeante, cela demande des formats appropriés, cela prend du temps, beaucoup de temps, et à chaque fois, peut générer des tensions, inévitablement. Mais ces fragilités méritent à nos yeux que nous y consacrons plus de temps et d'énergie qu'à nous soumettre à un régime dominant qui impose à la recherche les critères établis par ailleurs dans le champ du management et de la publicité. Ce régime d'évaluation collégiale sera sans nul doute au programme de l'un des chantiers que notre atelier va ouvrir, comme pour répondre à ceux qui nous rétorquent que nous ne serions que nostalgiques d'un passé révolu. Ces chantiers représentent pour nous le temps fort de notre positionnement, ils doivent être l'expérimentation d'une recherche collective guidée par la pertinence, le souci de la durée, le plaisir et le sens du travail bien fait...

Ainsi, ce n'est pas seulement un constat que nous dressons, c'est aussi un appel à la désobéissance. Face à des injonctions managériales qui nous poussent à être les acteurs de notre propre destruction, le maître mot est résistance. Il s'agit d'occuper et d'exploiter les innombrables interstices. De mettre à profit cette liberté académique qui nous est encore accordée pour préserver et cultiver la biodiversité au sein de l'Université. Pas seulement pour dire ou dénoncer, mais pour agir concrètement. Pour mettre nos valeurs en pratique dans nos activités d'enseignement et de recherche et pour les partager avec d'autres — collègues, doctorants, étudiants. C'est le sens de

notre « charte de la désexcellence », projet collectif issu du terrain et non des sphères dirigeantes, une suite de propositions à s'approprier, adapter, transformer en fonction des contextes et des profils.

C'est tout le contraire de la standardisation aveugle et de l'absence de sens que l'on nous vend sous la bannière de l'excellence.



Fig 7. Le sort réservé à l'excellence à l'Université Autonoma de Madrid

Comment nous désirons être évalué-e-s ?

ISABELLE STENGERS

Professeur à l'Université libre de Bruxelles

Désirer être évalué. Il y a quelque chose d'un peu paradoxal dans cette expression. Sauf si la question porte sur le « comment ». En ce cas, en tant qu'enseignants, nous savons que la question de la pertinence, la manière dont nous commentons, discutons le travail proposé par un-e étudiant-e peut — pour certains — avoir plus d'importance que la note. Une évaluation, aussi quantitativement généreuse qu'elle soit, n'y changera rien : s'il n'y a pas de retour pertinent sur la valeur de leur travail ces étudiant-es se sentiront méprisé-e-s ou alors nourriront un certain mépris pour l'enseignant qui maltraite son métier.

L'évaluation dite objective nous met dans une situation de ce genre, où le mépris prévaut — sauf que dans ce cas, nous savons que la pertinence n'est pas le moins du monde à l'ordre du jour, seulement l'injonction à « jouer le jeu ». Ce n'est pas que les critères manquent de pertinence, c'est que, avec ce que Dominique Pestre a appelé la gouvernance néo-libérale, ils doivent ne pas l'être. L'efficacité de l'évaluation est ici d'autant plus grande qu'ils ne le sont pas. Plus ils sont absurdes au regard de la situation ou de l'activité concrète de la personne évaluée, mieux ils signalent que leur enjeu est de faire prévaloir un impératif de soumission à une mobilisation qui demande flexibilité, absence de scrupule, opportunisme. Au sens fort, il s'agit de « démoraliser », c'est-à-dire de casser, de faire « perdre la boussole ». La boussole pointe en effet vers des valeurs qui pourraient signifier une récalcitrance insoumise,

voire même des comportements « irrationnels » de solidarité — l'équivalent d'un pacifisme alors que la guerre (compétition) généralisée doit prévaloir.

En d'autres termes, poser la question de la pertinence peut alors sembler utopique, voire romantique. Cette question me semble cependant avoir un sens, et ce pour deux raisons. D'abord parce qu'il s'agit de cesser de se plaindre, ou de protester comme si les chercheurs universitaires avaient droit à un traitement spécial, au vu de la valeur éminente de leur activité — tous les « métiers » impliquant que toutes les manières de faire ne se valent pas sont soumis à la même attaque. Ensuite parce que les pratiques de résistance éventuelles à cette attaque demandent, dans chaque cas, une imagination propre, une réappropriation qui ne se fasse pas seulement au nom du passé qui est en train de basculer dans le périmé, mais aussi d'un avenir qui intègre la manière dont ce passé a été vulnérable à l'attaque — car c'est de nos faiblesses, nos inerties, nos « je sais bien mais... » que cette attaque a, dans chaque cas, profité.

Il s'agit, dans notre cas, d'un « nous » qui désigne des chercheurs-enseignants appartenant à une institution qui rend solidaires les deux rôles. L'université moderne s'est inventée à partir de l'idéal de communautés de recherche qui prennent en charge tant « l'avancée des connaissances » dans leur champ que leur propre reproduction¹ — elle a inventé un mode d'enseignement dont réussite par excellence, l'idéal, est la formation de futurs chercheurs, d'éventuels collègues. Avons-nous pris la mesure du défi que constituait la démocratisation des études supérieures ou avons-nous maintenu le cap désigné par l'idéal ? Nous autres qui travaillons à l'université et avons en ce sens reproduit cet idéal, pensons-nous assez à ceux qui nous valaient bien et qui n'ont pu le faire ? Nous avons été évalués comme « conformes » — jusqu'où est allée notre conformité ?

Mais il ne s'agit pas ici de se sentir coupables, seulement de ne pas nous réputer victimes innocentes. Et il en est de même pour l'activité de recherche elle-même ainsi que pour son mode d'évaluation.

Certes dans les champs de recherche où la notion de « fait » est associée à une réussite vérifiée par une pratique collective exigeante, soumise aux objections de « collègues compétents », partageant les mêmes exigences disciplinaires et des questions « concurrentes », « courant ensemble dans

¹ Voir pour cette invention et ses conséquences, I. STENGERS, « Plaidoyer pour une science 'slow' », in *Une autre science est possible !*, Paris, Les Empêcheurs de openser en rond/La Découverte, 2013.

la même direction », le modèle classique d'évaluation par les pairs offre une garantie de pertinence évidente. Car ces collègues sont ceux pour qui la fiabilité du fait en question importe parce qu'il pourra le cas échéant — s'il surmonte les objections — permettre d'envisager la possibilité de poser de nouvelles questions, de concevoir de nouveaux dispositifs. Malheur à qui propose un fait ouvrant de nouvelles perspectives, qui devrait donc faire une différence pour ses collègues, mais ne suscite pas d'objection de la part de ceux que ce fait devrait intéresser, dont il devrait modifier ou mettre en question la recherche. Si nul ne prend la peine de le mettre à l'épreuve, c'est que nul n'envisage sérieusement l'éventualité de sa prise en compte : il restera alors dans les limbes, même pas réfuté. Et c'est cette forme d'intelligence collective, où l'objection est bienvenue et contribue à la stabilisation et à la création de reproductibilité des faits, que les chercheurs ont défendue lorsqu'ils ont réclamé l'autonomie de leur pratique contre les interférences d'intérêts « étrangers ». Ceux-ci seront accueillis, voire activement recherchés, mais ils ne doivent pas intervenir ouvertement, faire argument, avant que la réussite ait été reconnue par la communauté concernée.

On le sait, c'est ce modèle qui est aujourd'hui démantelé par ce qu'on appelle l'économie de la connaissance, avec le partenariat institué avec le privé, le secret commercial, la prise de brevet. L'intelligence collective se défait si la réussite n'a plus besoin de passer par les pairs. Cependant ce même modèle, généralisé tel qu'il l'a été, à tous les champs de recherche, est-il généralement synonyme d'intelligence collective ? Que signifie l'objection dans des champs plus conflictuels, où la notion même de « fait » pose problème, où il n'y a pas concurrence mais conflit. Un économiste « classique » sera-t-il évaluable par un économiste marxiste ? Un psychologue cognitiviste par un psychologue phénoménologue, voire psychanalyste ? Et que signifie-t-elle dans les champs où les « faits » sont avant tout produits par une instrumentation sophistiquée (neuro-imagerie) et s'accumulent sans correspondre pour autant à une démarche cumulative, une démarche telle que ce qui est proposé par l'un a des conséquences pour les autres ? Et qui sont mes pairs à moi, philosophe, alors que les philosophes pratiquent plutôt un art du dissensus, posant et reposant sans fin la question « qu'est-ce que la philosophie » ?

En d'autres termes le principe d'évaluation par les pairs, aujourd'hui battu en brèche, n'a jamais eu une pertinence égale dans tous les champs. Il traduit

plutôt la manière dont l'institution scientifique a érigé en modèle les sciences pour qui et par qui il a été inventé (rappelons qu'en France tout groupe de recherche est nommé « laboratoire »).

D'autre part, même dans ces sciences la mise en œuvre de ce principe a mal résisté à l'explosion du nombre des publications, des spécialités et des chercheurs. Dans les années 70' déjà le *publish or perish* était dénoncé, les référés, c'est-à-dire collègues compétents en première ligne, étaient surchargés, devenus peu fiables, et étaient mis en place des instruments bibliométriques (nombre de citations) permettant d'évaluer des candidats inconnus sur un mode qui n'exige pas la compétence des évaluateurs — ainsi que des tentatives de déceler les renvois d'ascenseurs et autres manœuvres auxquelles invitait la non-compétence de l'évaluation. De manière plus générale, les pratiques d'évaluation impliquant aussi bien le couplage entre évaluation et compétence activement intéressée à la pertinence de cette évaluation que la confiance interpersonnelle entre chercheurs qui « se connaissent et s'apprécient » étaient déjà mises à mal. On pourrait dire de ce point de vue que l'évaluation dite objective a transformé en norme impérative ce qui était déjà dénoncé et/ou déploré comme dérives pathologiques.

C'est pourquoi la question « comment désirons-nous être évalué-e-s » ne devrait pas se borner à l'évocation nostalgique d'un passé où la recherche était respectée, ni non plus évoquer un « nous » incluant indifféremment tous les chercheurs. La question constitue d'abord une épreuve adressée aux chercheurs en tant que situés par leur pratique, par ce qu'elle vise, par la question de ceux et celles à qui elle s'adresse, pour qui elle compte. Son enjeu est la capacité, pour celles et ceux que cette question concerne, de se situer autrement que par des références plus ou moins grandiloquentes à la cause de la connaissance ou du progrès. Comment favoriser, régénérer ou réinventer une intelligence collective qui puisse fonctionner en tant qu'instance immanente d'évaluation est sans doute un point commun puisque c'est ce que la gouvernance néo-libérale détruit quel que soit le champ de recherche. Mais il s'agit aussi, posant la question de l'adresse, de mettre en cause une certaine irresponsabilité revendiquée quant à la manière dont les savoirs qu'ils produisent seront « valorisés ». Pour qui les savoirs produits comptent-ils ?

Il ne suffit donc pas de protester contre la « marchandisation » des savoirs. Comme Dominique Pestre l'a bien montré², tout « régime de savoir »

² D. PESTRE, *Science, argent et politique*, Paris, INRA éditions, 2003.

inclut la question du mode de valorisation de ces savoirs. La prétention à faire de « l'avancée des connaissances » une cause en soi, qui transcenderait toute demande de compte et constituerait une vocation que toute société se devrait de nourrir de manière désintéressée, est une aimable façade construite à l'époque même où des sciences comme la physique, la chimie, la géologie entraient dans des rapports de symbiose particulièrement intimes avec le « développement des forces productives ». Elle affirme certes la liberté consentie à des communautés de recherche de déterminer leurs questions de manière autonome. Cependant, cette autonomie a un prix : les scientifiques doivent se garder comme la peste de se préoccuper de la manière dont les réponses qu'ils donnent à leurs questions « purement scientifiques » seront « valorisées », c'est-à-dire acquerront d'autres valeurs, seront évaluées selon d'autres critères. S'en préoccuper serait sortir de leur rôle. On entend dire par certains : « nous respectons la démocratie ; il appartient à 'la société' de décider ». Mais la question de savoir comment cette « société » est équipée pour prendre de telles décisions n'est pas leur affaire. Et lorsqu'une innovation industrielle issue des recherches scientifiques est contestée — je pense aux OGMs notamment — les scientifiques qui s'en font les promoteurs le font en toute impunité. Leurs arguments ne suscitent pas la moindre surprise indignée de la part de leurs collègues, alors même qu'ils ignorent superbement les limites de leur compétence. D'une manière ou d'une autre, la théorie du « déficit », impliquant que c'est par manque d'information ou par peur du changement que le « progrès impulsé par la science » peut être contesté par le public, règne en maître sur l'imaginaire des scientifiques.

Je parle d'imaginaire et non d'imagination : lorsque Marc Van Montagu, par exemple, présente les OGMs comme solution au problème de la famine dans le monde, il écarte d'un revers de main la question des conséquences socio-économiques et écologiques de l'extension massive de monocultures brevetées et ne fait pas montre de beaucoup d'intérêt pour les explications que ses collègues socio-économistes donnent des famines. L'OGM « abstrait », sorti des laboratoires, résoudra un problème auquel il ne connaît pas grand-chose : cela ne le regarde pas en tant que biologiste ce qui ne l'empêche pas de plaider que la « politisation » du problème fait obstacle à ce qui lui semble rationnellement évident. Des scientifiques cohérents avec leurs prétentions démocratiques (« c'est à la société de décider ») devraient en appeler de toute leur force à une mise en politique de la manière dont leur savoir est valorisé

« hors science ». Mais le silence écrasant de la communauté scientifique lorsque l'un des siens en appelle aux valeurs de la rationalité et de l'objectivité pour dénoncer une telle « politisation » n'a pas attendu l'économie de la connaissance.

La revendication d'irresponsabilité des scientifiques, quant aux conséquences de leur recherche, est indissociable de l'autonomie qui leur a été concédée dans le passé. L'avertissement quasi-rituel « Ne tuez pas la poule aux œufs d'or ! N'essayez pas de vous approprier directement cet or ! Laissez-nous pondre en paix et vous pourrez ensuite valoriser nos œufs, ils vaudront pour vous de l'or ! » a toujours impliqué que les poules académiques s'adressent d'abord à ceux qui ont le pouvoir de faire de l'or avec leurs œufs, c'est-à-dire plaident que ces intérêts de ces derniers seront mieux servis par l'autonomie qu'elles demandent. Le fait que l'avertissement ne soit plus entendu signifie peut-être que la fiabilité scientifique des œufs n'importe plus. Dans une économie désormais axée sur la promesse et la spéculation, tout ce qui brille peut faire l'affaire. Mais l'imaginaire de la poule, tout à la fois irresponsable et fière de ce que ses œufs puissent assurer le progrès de l'humanité, est toujours aussi puissant tant en ce qui concerne l'élevage des poulets, c'est-à-dire la formation des étudiant et des chercheurs, que la pratique des éleveurs. L'idéal de l'évaluation par les pairs — ceux qui sont censés vérifier la solidité des œufs — reste officiellement la norme, même si la plupart des pairs dépendent désormais pareillement de partenariats avec le privé, et même si les objections ne sont plus bienvenues si elles mettent en cause ce partenariat³. Cet imaginaire n'est pas mis en cause car il traduit la canalisation de l'imagination des chercheurs, active lorsqu'il s'agit d'enjeux qui concernent leur compétence, littéralement anesthésiée lorsqu'il s'agit de questions étiquetées « non scientifiques », c'est-à-dire politiques, éthiques ou idéologiques. De manière générale l'accusation de « politisation de la science » disqualifie les chercheurs qui « perdent leur temps » à s'intéresser à des questions qui « ne font pas avancer le savoir », alors que l'arrogance superficielle et naïve des jugements qui tiennent lieu de réponse à ces questions est tolérée : elle authentifie leur auteur comme un « vrai scientifique ».

³ De ce point de vue l'affaire Séralini est exemplaire. Mettant en cause l'innocuité du Round-up de Monsanto, comme aussi les tests censés garantir cette innocuité, il a eu affaire non à des collègues intéressés mais à de véritables ennemis. Voir à ce sujet F. CHATEAURAYNAUD, « L'histoire des OGM n'est pas une controverse ratée mais un conflit réussi », in *Logiques argumentatives*, 2010. L'article est en ligne : <http://socioargu.hypotheses.org/1807>.

Parler de responsabilité des chercheurs et des chercheuses ne signifie pas bien évidemment leur demander qu'ils garantissent la fiabilité des prétentions à servir le « progrès humain » qui accompagnent ce qui sort de leur laboratoire. Il s'agit seulement de leur demander de ne pas s'en laver les mains, d'être sérieusement, c'est-à-dire collectivement, concernés par cette fiabilité. Cela suffirait pour que Van Montagu et consorts, par exemple, ne puissent plus parler impunément de famine sans compétence vérifiable en la matière. Mais cela implique — ce qui n'est pas le cas dans les programmes officiels visant à vendre au public l'idée d'une science désormais responsable — qu'ils acceptent de « perdre leur temps » à s'intéresser au monde qui commence aux portes de leurs laboratoires. Et cet intérêt à son tour implique une formation qui active un appétit imaginatif, une intolérance aux types de généralités qui servent aux scientifiques à s'exonérer de toute responsabilité. Un appétit imaginatif et une intolérance qui rendraient normal que, chargés d'expertise, ils exigent la présence active des co-experts qui soient capables d'envisager les aspects d'une situation qui ne sont pas de leur compétence et les empêchent de faire prévaloir le point de vue qu'il considère comme scientifique sur ce qui, anecdotique, est censé s'arranger ou doit être accepté comme le « prix du progrès ».

Il est inutile de souligner que l'institution scientifique, ici, ne peut être séparée des instances que sont loin de déranger la figure du scientifique revendiquant fièrement son irresponsabilité et le mode de formation et d'évaluation des chercheurs qui les formate sur ce mode. Et je parle ici non seulement du privé, mais aussi des institutions étatiques qui ne sont pas précisément des « amies » de la mise en politique de questions qu'il s'agit de gérer en lieu clos. Inversement l'exercice d'évaluation de l'expertise que sont susceptibles d'accomplir des citoyens dépourvus de compétence particulière mais qui, tirés au sort, ont accepté de consacrer plusieurs week-ends à la question autour de laquelle ils ont été réunis, suffit par sa lucidité, sa pertinence et son exigence imaginative, à manifester que les experts scientifiques ont quelques raisons de craindre de tels modes d'évaluation. Ici la rengaine des scientifiques « c'est à la société de décider » prend sens, non que ces citoyens décident, mais parce que la manière dont ils écoutent et interrogent des experts d'horizons multiples aboutit à ce qui, d'une manière ou d'une autre, devraient être un préalable obligé à la décision : un déploiement véritablement « désintéressé » de la question et une évaluation critique des

arguments autorisés⁴. L'institution scientifique, qui prétend tenir aux seules valeurs de la rationalité, a cependant vécu comme une catastrophe l'événement OGM, lorsque le problème de l'agriculture de demain a été déployé et les arguments autorisés battus en brèche. Elle préfère poser la question (apolitique) du « public » ou de « l'opinion », qu'il s'agit de réconcilier avec « sa » science, à coup de fêtes, de journées portes ouvertes et autres opérations de propagande.

Il devrait aller sans dire que cette institution, telle qu'elle a fonctionné depuis plus d'un siècle et demi en symbiose avec un « progrès » censément supra-politique n'est pas seule responsable du caractère radicalement insoutenable de ce qu'on appelle « développement ». Il reste que, à part les états d'âme des physiciens ayant participé au développement de l'arme nucléaire, on ne peut pas dire que les scientifiques dans leur ensemble aient prêté une oreille très attentive à ceux qui, nombreux et depuis bien longtemps, mettaient l'accent sur le caractère irresponsable de ce « développement », ou aient collectivement protesté contre ceux de leurs collègues qui disqualifiaient toute mise en cause un peu radicale comme manifestation d'irrationalité. Aujourd'hui encore, je ne suis pas sûre que la question du rôle joué par la référence à la rationalité dite « scientifique » dans la mise sous silence de la contestation soit bienvenue dans les milieux scientifiques, et que, notamment, le rêve d'une science apportant une solution miracle, de type géo-ingénierie, au désordre climatique ne soit pas vivace. Il me semble parfois que la menace d'une perte de confiance dans « la science » est vécue comme plus redoutable que celle qui affecte notre avenir. Mais il est hors de question d'admettre que la perte de confiance puisse être légitime. La question de savoir comment la regagner est aujourd'hui confiée aux sciences sociales : « à vous de vous charger de faire valoir que nous sommes désormais ouverts aux questions qui préoccupent le public mais sans nous demander de perdre notre temps, qui doit être consacré tout entier à l'avancé des questions vraiment scientifiques ».

Poser la question « comment désirons-nous être évalué-e-s ? » ne signifie évidemment pas que toutes les voies de recherche aient à affronter le même type d'épreuve. La question de l'évaluation, que je voudrais renvoyer

⁴ C'est ce que la Fondation Sciences citoyennes propose d'appeler convention citoyennes, et qu'elle associe à un protocole indépendamment duquel la convention peut devenir un n+unième dispositif destiné à aider à l'acceptabilité d'une innovation. Voir, en ligne <http://sciencescitoyennes.org/dossier-de-presse-les-conventions-de-citoyens-cdc/>

à la question « cui bono ? » — « pour le bien de qui ? », c'est-à-dire « à qui manquerions-nous ? », ou « pour qui ce que nous faisons importe-t-il ? » — se pose bien évidemment différemment dans les champs technoscientifiques producteurs de brevets et de spin off, en archéologie ou en philologie. Cependant, elle réclame de tous les champs une épreuve commune, qui est de renoncer à la prétention de la poule académique, qui prétend servir la « cause » de la connaissance désintéressée, quitte à fabriquer quelques généralités bâclées qui reviennent à affirmer que l'humanité entière manquerait du savoir produit. Certes, tous les champs de recherche ne peuvent être réputés « coupables » d'avoir travaillé pour le bien de certains en oubliant les conséquences qu'auront à subir d'autres, mais le problème ne passe pas par le contraste coupable/innocent. Aucun champ n'est innocent dans la mesure où tous ont bénéficié de l'institution de « la science », de la définition de l'avancement de nos connaissances comme « cause » n'ayant pas besoin de justification. Tous ont bénéficié de la légitimité de principe concédée à « la science », même ceux qui se sont fait un devoir de critiquer les sciences dites objectives. Car ceux-là même l'ont fait sans trop s'interroger pour autant sur l'autonomie relative dont ils bénéficiaient.

Il s'agit donc de penser comment nous désirons être évalué-e-s à partir d'une commune non-innocence — celle dont témoigne notamment la perplexité d'avoir, pour ce faire, à se rendre capable de répondre à la question « cui bono » — « pour qui cette recherche est-elle pertinente ? » ; « à qui s'adresse-t-elle ? ». Donna Haraway propose d'ailleurs d'entendre le terme « responsabilité » au sens de répondre-abilité, capacité de répondre sans invoquer une légitimité toute faite, auto-évidente. Il s'agit d'accepter la légitimité d'une question qui fait épreuve parce que ce qui est évalué n'est pas, alors, la simple compétence, la reconnaissance par les pairs, l'adéquation aux normes du champ, et parce que cette question exige l'abandon de la position héroïque du refus d'être « asservi au désir de l'autre », de gens incapables, paraît-il, de comprendre le sens de ce qui nous fait œuvrer, et qui doivent donc être tenus à distance. *Mais pour qui nous prenons-nous ?*

Et ce qui frappe alors est que nos institutions académiques n'ont pas, ou peu, de « milieu associé », ce que j'appellerais volontiers de milieux « amateurs » au sens où il faut entendre par amateur non l'incompétence mais le goût cultivé, l'amour qui sait discriminer, l'expérience qui permet

d'évaluer, qui, éventuellement, ne s'en laisse pas compter⁵. Les amateurs sont ceux qui savent objecter si on s'adresse « mal » à eux, si on les prend pour des gogos — que l'on pense aux amateurs de musique ou de technologie informatique, et même de sports. J'ai souvent imaginé des journalistes scientifiques qui seraient capables de prendre exemple sur les journalistes sportifs et qui sauraient exiger qu'on leur parle « vrai » au lieu de leur livrer une propagande édifiante.

Certes, on pourrait dire que nous sommes toutes et tous « amateurs », dès qu'il s'agit d'un champ qui n'est pas le nôtre, mais pas au sens, malheureusement, que je viens de présenter. La tolérance indifférente, le mépris à peine secret, l'ignorance crasse sont monnaie courante. Comme aussi les mauvais coups hypocrites : ne sommes-nous pas en situation structurelle de rivalité pour les financements, mais tenus aussi de ne pas mettre en question ouvertement le travail d'un « collègue » car ce serait « affaiblir l'institution », ouvrir la porte à ceux qui doivent être tenus à distance, qui doivent respecter la façade lisse de « la science ».

J'en reviens ici à la question de l'enseignement. La singularité de l'université, telle que je l'ai entendue célébrer, est que c'est le seul lieu d'enseignement où l'enseignant sait que parmi ses étudiants se trouve peut-être un futur collègue. Et je ne nierai pas ce qui est à la fois un privilège et une source d'exigence. Mais c'est aussi une responsabilité si la formation est axée sur l'idéal de la poule aux œufs d'or. Et, en tout état de cause, il faut aussi dire que c'est aujourd'hui une fiction et un énorme gâchis. Je ne plaide évidemment pas pour que l'université fabrique des diplômés « prêts à l'emploi », dotés d'une formation en « adéquation avec les besoins du marché du travail ». Je dis seulement que la démocratisation de l'université n'a pas été vécue comme un défi, mais subi, l'œil fixé sur le « niveau », qui ne devait pas baisser. Vous êtes les bienvenus, avons-nous dit aux cohortes d'étudiants issus de milieux « nouveaux », mais après = avant, et il faudra vous adapter. Nous allons continuer comme avant, vous former de telle manière que certains puissent mériter de, peut-être, nous succéder. Quant aux, si nombreux désormais, autres, ils pourront toujours bénéficier, sur le marché du travail, de l'accès réservé aux postes demandant un diplôme universitaire. Et tant pis s'ils ont pris le goût de ce qu'ils ont rencontré à l'université. La logique méritocratique qui nous imbibe les en exclut.

⁵ Voir I. STENGERS, « Pour une intelligence publique des sciences », in *Une autre science est possible*, op. cit.

Nous n'avons pas inventé un enseignement qui contribue à créer un « milieu associé », qui ouvre à ceux que nous formons la possibilité de maintenir un engagement actif quoiqu'amateur dans la spécialité dont ils auraient pris le goût — la possibilité non pas simplement de se « tenir informé » mais d'être habilités à participer activement à la valorisation des savoirs qui sont produits et transmis à l'université. Et si l'université devait être évaluée non certes dans ses rapports avec le marché du travail mais dans sa capacité à former de futurs amateurs, capables de critiquer, d'apprécier, d'objecter, de situer ? Bref de former ceux qui forceraient les chercheurs « élus » à se rendre capables de s'adresser « pour de vrai » non seulement au milieu raréfié de leurs seuls « collègues compétents » mais aussi à ceux qui, très légitimement, n'accepteront pas de se voir traités en incompetents ? On parle si souvent de « culture scientifique, et l'université s'enorgueillit de ne pas former de simples « techniciens compétents ». Mais ce n'est pas quelques cours de « culture générale » qui feront l'affaire si le défi est un rapport « cultivé » au savoir, une capacité de situer la pertinence de ce que l'on sait et d'en répondre. Cela demande une formation dont la première réussite serait la capacité à s'adresser à, ou à faire partie d'un, milieu d'« amateurs exigeants », ce milieu que l'université ne sait pas entretenir, n'imagine même pas qu'elle devrait entretenir.

Aujourd'hui, plus que jamais, pourtant la question de la pertinence de nos savoirs académiques est une question vive, une question transversale qui devrait faire épreuve pour les enseignants-chercheurs. Les étudiants qui entrent à l'université appartiennent à une génération qui sera confrontée à des défis que nous ne pouvons imaginer. L'université peut-elle devenir capable de nourrir, dans un monde où rien n'ira plus de soi, leur imagination contre la rigidité du *business as usual*, y compris le business que constitue l'avancée de savoirs aveugles à la question de leur pertinence ?

Gilles Deleuze a écrit que la honte était un motif puissant de penser. La honte et non la culpabilité qui constitue plutôt un attracteur faisant revenir toute pensée à la même rengaine doloriste. Ce que j'ai présenté n'a évidemment rien ni d'une revendication ni d'un programme. Je n'entends disqualifier en rien la souffrance de bien des chercheuses et chercheurs qui se sentent trahis par ceux qui étaient censés les défendre. Mais il s'agissait pour moi de tenter de penser sans nostalgie pour le passé, devant celles et ceux dont, d'une manière ou d'une autre, nous avons entériné l'exclusion. La question

« comment désirons-nous être évalué-e-s ? » doit alors être comprise comme une épreuve, demandant que nous nous situions dans un milieu plus dense, que nous apprenions à imaginer pour qui nous faisons ou pourrions faire une différence susceptible de compter, une différence dont ils seraient en retour susceptibles d'évaluer la pertinence.

Il faut l'accepter, ceux qui ont prétendu nous protéger de la perte de temps qu'entraînerait la création de rapports vivants (non « pédagogiques ») avec ce milieu, ceux qui nous ont encouragé-e-s à le tenir pour irrationnels, ne sont plus, et n'ont sans doute jamais été, des amis de ce que nous appelons rationalité. Et nous-mêmes l'avons assez souvent déshonorée, cette rationalité, en en faisant un principe de démarcation permettant de sélectionner ce qui, dans une situation, nous regarde. Nous avons peu résisté à la tentation facile de faire de la rationalité une bannière territoriale alors que je n'en connais pas d'autre définition transversale que celle de se rendre capable d'entendre ce que demande une situation, sans la soumettre à nos habitudes de pensée, sans rêver que ce que nous ne prenons pas en compte s'arrangera de soi-même. Nous vivons une époque où les comptes de toutes nos négligences, de toutes nos soumissions nous sont présentés sur un mode tel que nous savons que cela ne s'arrangera pas de la sorte. Il est temps, plus que temps, d'apprendre l'art des rapports avec ceux dont nous avons besoin si nous désirons être rationnels.

Éléments pour construire une évaluation alternative

MARYSE SALLES (IRIT/SIG)

Université de Toulouse 1-Capitole

GABRIEL COLLETIS (LEREPS)

Université de Toulouse 1-Capitole

1. INTRODUCTION

Le présent texte a été produit suite à un colloque qui s'est tenu à Bruxelles le 16 octobre 2015 accueilli par l'Université libre de Bruxelles (ULB), « L'évaluation de la recherche en question », dans le cadre des séminaires Ilya Prigogine « Penser la science ».

L'objectif général de cette journée, la seconde du colloque, consistait en une réflexion sur les *alternatives* possibles aux pratiques actuelles d'évaluation.

Les chercheurs rassemblés étaient donc invités à produire des propositions nouvelles en matière d'évaluation.

Les propositions que nous avons présentées ont été conçues, d'une part, en nous basant sur le rapport AFEP *Pour une nouvelle démarche d'évaluation des laboratoires de recherche, des enseignants-chercheurs et des chercheurs en économie* (2011)¹, d'autre part, en utilisant un outil d'ingénierie qui permet

¹ AFEP, (2011), *Pour une nouvelle démarche d'évaluation des laboratoires de recherche, des enseignants-chercheurs et des chercheurs en économie*, Rapport de la Commission Évaluation de l'Association Française d'Économie Politique, coordonné par G. Colletis et Th. Lamarche. En

de construire le chaînage allant des grandes représentations des objets concernés (recherche, évaluation...), jusqu'aux procédures et normes à vocation opérationnelle.

En préalable nous noterons que l'activité des acteurs de la recherche est encadrée par des *normes*. Les indicateurs, les listes de revues classées, des objectifs de publication, des procédures diverses (d'évaluation, d'avancement, de dépôt de projet de recherche...), sont des exemples de ces normes. Ces normes se présentent généralement comme simplement techniques, neutres, garantes d'efficacité.

Les travaux présentés ici se fondent sur une double hypothèse. La première est que les normes ne sont jamais naturelles, neutres ou simplement techniques. Elles participent toujours d'un dispositif de pouvoir, et sont le résultat et la codification d'un rapport de forces. À ce titre, elles sont bien situées dans le temps et dans l'espace social, et véhiculent des visions du monde, des représentations, des *doxai*. Celles-ci peuvent, par exemple, concerner le rôle ou la nature de l'exercice d'évaluation, les missions de l'Université, de l'Enseignement supérieur, ou de la Recherche ou encore les motivations des chercheurs.

La seconde hypothèse est que si les représentations (*doxai*) donnent leur sens profond aux normes, elles restent très globales. La production des normes est rendue possible par l'existence d'un niveau intermédiaire : les *principes*. Ceux-ci vont permettre de rendre les représentations opérationnelles, au travers d'instruments à portée générique. Au sein des organisations, les technologies de gestion se situent à ce niveau intermédiaire, de même que, plus largement, ce que Foucault (2001)² désignait comme les « technologies gouvernementales ».

Dans une première partie, nous présenterons l'outil d'ingénierie Doxai-Principes-Normes (DPN) (Salles, 2007³, 2015⁴).

ligne : http://assoeconomiepolitique.org/pour_une_nouvelle_demarche_devaluation_des_laboratoires_de_recherche_des_enseignants-chercheurs_et_des_chercheurs_en_economie/

² FOUCAULT M., (2001), *Dits et écrits II. 1976-1988*, Paris, Gallimard, coll. Quarto.

³ SALLES M., (2007), Présentation du dossier « Représentations, modèles et normes pour l'entreprise », in *Droit & Société*, n° 65. En ligne : <http://www.cairn.info/article.php?REVUE=droit-et-societe&ANNEE=2007&NUMERO=1&PP=21>

⁴ SALLES M., (2015), *Décision et système d'information*, Coll. Systèmes d'information avancés, vol. 2, ISTE-WILEY. En ligne : <http://iste-editions.fr/collections/systemes-d-information-web-et-informatique-ubiquitaire/products/decision-et-systeme-d-information>

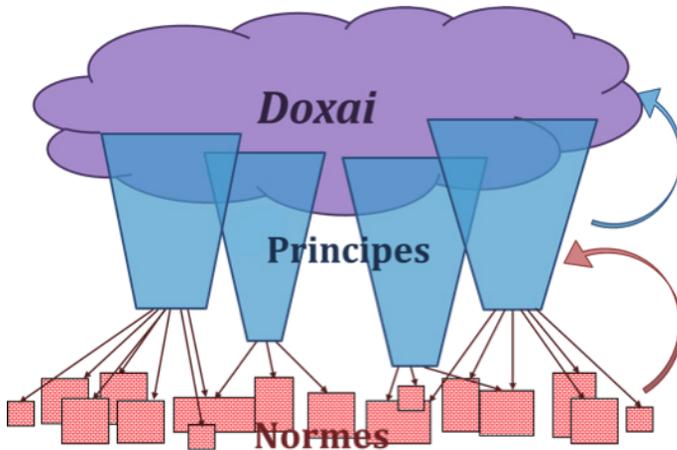
En nous basant sur le rapport AFEP (2011)⁵, nous nous appuyerons sur cet outil, dans un premier temps pour analyser l'évaluation actuelle de la recherche en économie (deuxième partie), puis, dans un second temps, pour présenter les éléments d'une évaluation alternative (troisième partie).

2. L'OUTIL D'INGÉNIERIE DPN

Pour analyser une situation existante, dans notre cas les pratiques actuelles d'évaluation de la recherche, l'outil DPN permet de révéler la chaîne qui va des doxai (représentations, paradigme, niveau ontologique), aux normes (niveau opérationnel), en passant par le niveau intermédiaire (et déterminant) des principes.

Mais DPN est aussi un outil qui peut aider à construire une alternative complète, en déclinant les trois niveaux D-P-N, et en reconnaissant le rôle central du niveau des principes dans la mise en œuvre de la doxa que l'on veut promouvoir (Salles & Colletis, 2013)⁶.

Le schéma suivant présente les trois niveaux de DPN, distincts mais indissociables. Nous les détaillons dans la suite.



⁵ *Op. cit.*

⁶ SALLES M., COLLETIS G., (2013), « Déconstruire la doxa dominante, construire une pensée politique alternative. Du lien entre les représentations, les principes et les normes », in *LoSguardo*, The Instruments of the Power, XIII, 10. En ligne : http://www.losguardo.net/public/archivio/num13/articoli/2013_13_Gabriel_Colletis_Maryse_Salles_Deconstruire_la_doxa_dominante.pdf

Le niveau des doxai

Le niveau des *doxai* est celui des représentations générales existant dans l'organisation concernée (ou plus largement le système considéré), et promues par elle. Les *doxai* peuvent connaître divers niveaux de formalisation : opinions faiblement structurées, croyances, idéologies, écoles de pensées relativement homogènes et cohérentes. Au sein d'une organisation les *doxai* vont prendre la forme de grands choix paradigmatiques sur les missions de l'organisation, et de grandes options politiques ou stratégiques.

Elles vont déterminer le statut ontologique des grands objets impliqués dans la vie de l'organisation. Dans le cadre de l'évaluation de la recherche, le niveau des *doxai* définira ce que sont la recherche et ses missions, l'évaluation et ses objectifs, un laboratoire, une production scientifique, etc.

Ces représentations générales sont le plus souvent à reconstruire, elles ne se donnent pas aisément à voir, et il faut les découvrir dans le fonctionnement de dispositifs concrets. La force des *doxai* tient en partie à cette existence latente, qui renforce leur caractère *in-discutable*, que soulignait Althusser (1976)⁷ à propos de l'idéologie dominante : « C'est en effet le propre de l'idéologie que d'imposer (sans en avoir l'air, puisque ce sont des « évidences ») les évidences comme évidences, que nous ne pouvons pas ne pas reconnaître (...) ». Les *doxai* sont donc à *révéler* à partir de leurs manifestations visibles, c'est-à-dire, dans notre modèle, à partir des principes et des normes.

Le niveau des principes

Les *doxai* sont des visions générales, qui ne peuvent être utilisées pour organiser et contrôler le fonctionnement d'une organisation ou d'un système social. Des dispositifs plus structurés sont alors nécessaires.

Les *principes* vont préciser à l'aide de quels dispositifs, concepts, objets et méthodes les orientations du niveau précédent vont être rendues plus opérationnelles. Ils vont ainsi par exemple s'incarner dans des éléments du langage, des conventions, des catégories du droit, des méthodes de management, des modes de prise de décision, des modes de définition des nomenclatures et des principes de classement, des méthodes de définition des indicateurs d'évaluation, la structure des systèmes d'information...

⁷ ALTHUSSER L., (1976), « Idéologie et appareils idéologiques d'État. (Notes pour une recherche) », in *Positions (1964-1975)*, Paris, Les Éditions sociales, p. 67-125.

L'influence des « instruments de gestion » dans la trajectoire des entreprises, a été reconnue il y a déjà une trentaine d'année par Berry (1983)⁸, qui notait « (...) les instruments de gestion sont souvent des éléments décisifs de la structuration du réel, engendrant des choix et des comportements échappant aux prises des hommes, parfois à leur conscience ». Il s'inquiétait des effets qu'ils engendraient : « (...) réduction de la complexité, mise en place d'automatismes de décision, division de la vigilance, régulation des rapports sociaux et maintien de la cohérence », dont il soulignait le caractère caché et le rôle essentiel sur la mise en œuvre des stratégies. Berry concluait en notant : « c'est l'intendance qui commande ».

Les principes constituent ainsi selon nous, par essence, des « machines de pouvoir » au sens de Foucault (2001)⁹, par lesquelles « le pouvoir domine et se fait obéir ». Nous les qualifions de *machines à produire des normes*, dont ils fournissent le cadre conceptuel et méthodologique.

Il s'agit donc d'un niveau déterminant, qui cependant est fréquemment ignoré, confondu avec celui des normes.

Le niveau des normes

Le niveau des normes représente le niveau d'application concrète des doxai. Le niveau des normes est produit par l'application des principes du niveau précédent, qu'elles rendent opérationnels. La fonction des normes est d'encadrer concrètement l'activité. Elles sont en très grand nombre.

Contrairement aux doxai et à certains types de principes (comme les conventions), les normes sont visibles, formalisées ou formalisables : lois, règlements et normes, nomenclatures, procédures, listes de critères, listes de contrôle, paramètres, logiciels...

À titre d'illustration, un principe de définition du classement des revues génère (ou participe à la génération de) un ensemble de normes : les listes de revues classées elles-mêmes bien entendu, mais aussi la procédure d'attribution de certaines primes, celle de recrutement d'enseignants-chercheurs...

Comme cela a été indiqué précédemment, les normes se présentent généralement comme purement techniques, issues d'une logique de

⁸ BERRY M., (1983), *Une Technologie invisible ? L'impact des instruments de gestion sur l'évolution des systèmes humains*, Centre de recherche en gestion (CRG) de l'École Polytechnique, Paris.

⁹ Op. cit.

rationalisation, d'amélioration des pratiques, de « modernisation », et ne pouvant donc faire l'objet d'aucun débat.

3. ANALYSE DES PRATIQUES ACTUELLES D'ÉVALUATION EN ÉCONOMIE

Cette section se propose d'utiliser la grille méthodologique que nous venons d'exposer en l'appliquant au champ de l'évaluation de la recherche en économie. Comme indiqué précédemment, nous nous baserons pour cela sur les éléments du rapport produit par l'AFEP (2011)¹⁰.

La doxa dominante dans l'évaluation de la recherche en économie

Il est possible d'identifier un ensemble de composants de la doxa dominant actuellement l'évaluation de la recherche en économie, concernant les missions de la recherche, l'intention de l'évaluation, la vision des chercheurs ou encore le rapport à la quantification (Colletis, 2010)¹¹.

À la question « quelles sont les missions de la recherche en économie ? », la doxa dominante répond que la recherche doit être mise principalement au service des entreprises.

L'intention de l'évaluation est de vérifier la conformité à une cible, à un modèle prédéfini. Les différences sont ainsi des frottements à éliminer, et il faut viser l'homogénéité.

Pour cette doxa, évaluer (et au-delà, gérer la recherche) consiste à comparer, classer, sélectionner, surveiller, et punir ou récompenser.

C'est une doxa basée sur le contrôle (et non la confiance), qui considère le chercheur sur le modèle de l'*homo œconomicus* (calculateur, égoïste, recherchant un profit personnel maximal), et promeut la compétition comme seul mode de rapport entre les chercheurs, les laboratoires, les territoires, les États...

Une autre croyance de cette doxa est que tout peut être quantifié et, par suite, que tout peut être comparé. Par ailleurs, elle considère que la

¹⁰ *Op. cit.*

¹¹ COLLETIS G., (2010), « Pour un débat sur les représentations en économie », in *Revue Esprit*, n° 12. En ligne : https://www.cairn.info/article.php?ID_ARTICLE=ESPRI_1012_0150&DocId=66841&hits=5639+10+

quantification quantifie le réel (le monde « tel qu'il est »), de façon « objective », technique, « neutre ». La doxa dominante le plus souvent ignore ainsi, ou omet de reconnaître, que toute quantification implique l'existence de conventions. Desrosières (2012)¹² souligne ainsi que « le verbe *quantifier*, dans sa forme active (*faire du nombre*), suppose que soit élaborée et explicitée une série de conventions d'équivalences préalables, impliquant des comparaisons, des négociations, des compromis, des traductions, des inscriptions, des codages, des procédures codifiées et répliquables, et des calculs conduisant à la mise en nombre. La mesure proprement dite vient ensuite, comme mise en œuvre réglée de ces conventions ».

Pour la doxa dominante, qui occulte ce travail de formation des conventions d'équivalences, la quantification ne saurait donc être remise en question, voire même être discutée.

Notons que cette doxa, dans sa globalité, est née au sein des entreprises privées. Elle s'étend aujourd'hui aux institutions (dont l'université), aux services publics, au secteur associatif...

Quelques principes de la doxa dominante

Quelques exemples de principes sont exposés ci-dessous, regroupés en cinq grandes catégories : langage, conventions, mode d'élaboration du processus d'évaluation, contenu du système d'évaluation, systèmes d'information.

1. Langage et discours dominants dans l'évaluation

Le langage de l'approche dominante de l'évaluation est issu du management des entreprises. Nous retrouvons ainsi des termes comme : leadership, excellence, performance, prix, visibilité internationale, compétitivité, ambition.

On observera que ce qui est excellent est « international », ce qui ne l'est pas ou l'est moins est « national » ou, pire, « local »...

2. Conventions

La seule production scientifique recevable se limite aux articles publiés dans certaines revues, les instances d'évaluation ne sont pas évaluées.

¹² DESROSIÈRES A., (2012), « Est-il bon, est-il méchant ? Le rôle du nombre dans le gouvernement de la cité néolibérale », in *Nouvelles perspectives en sciences sociales : revue internationale de systémique complexe et d'études relationnelles (NPSS)*, vol. 7, n° 2, mai, p. 261-295.

3. Mode d'élaboration du processus d'évaluation

Les évaluateurs ne sont pas élus, leur choix dépend pour l'essentiel de l'instance d'évaluation.

4. Contenu du processus d'évaluation

Le principe de base, *de facto*, est celui de non lecture des travaux. La lecture est « sous-traitée » aux revues qui évaluent les articles. Un usage extensif est fait des indicateurs bibliométriques.

5. Systèmes d'information

Les systèmes d'information de l'évaluation emploient différents types d'indicateurs bibliométriques en fonction du choix de certains modes de constitution des bases de données utilisées en bibliométrie.

On observera que ces bases de données sont dans leur très grande majorité produites hors France et pays francophones et ne reconnaissent que certaines langues et certains pays. Certains domaines ou courants de l'économie y sont très bien représentés, quand d'autres le sont fort peu voire pas du tout. par ailleurs, les listes de mots-clefs utilisées dans ces bases de données excluent pour l'essentiel les concepts représentatifs des écoles de pensées n'appartenant pas au *mainstream*.

Quelques normes de la doxa dominante

Les normes étant par définition en grand nombre, nous ne citons dans la suite que quelques exemples concernant l'évaluation actuelle :

- les jeux d'indicateurs ; notons que les indicateurs, par leur fonction de réduction extrême de la complexité du réel, représentent la norme par excellence,
- les algorithmes de calcul des notations et des classements des chercheurs, des laboratoires, des universités,
- les règlements (ex., les règlements des concours de recrutement),
- les procédures d'octroi des primes liées à la recherche,
- les procédures d'avancement,
- les listes de revues et leur classement,
- les logiciels pour déposer une demande d'avancement ou un rapport de laboratoire, pour consulter un rapport d'évaluation et y répondre,

- la liste des mots-clés des banques de données (qui excluent de fait certains champs ou certains courants de pensée),
- et demain : les algorithmes de calcul du service d'enseignement (modulation) ?

4. QUELQUES ÉLÉMENTS D'UNE ÉVALUATION ALTERNATIVE

Comme indiqué plus haut, nous présenterons dans cette section la proposition de l'évaluation alternative élaborée par l'AFEP, selon les trois composants de DPN : doxai, principes, normes.

Une doxa alternative de l'évaluation de la recherche en économie

Nous commençons la présentation d'une évaluation alternative par le niveau des doxai, en précisant d'abord les finalités potentielles de la recherche et celles de l'évaluation qui lui serait associée.

Selon le rapport de l'AFEP, les finalités de la recherche devraient consister en :

- une avancée des connaissances et une production de savoirs utiles à la société,
- avec une attention particulière portée aux questions/problématiques/ domaines émergents,
- la promotion de la diversité, reconnue comme seule source de vitalité de la recherche.

Les finalités de l'évaluation selon l'AFEP devraient se décliner comme suit :

- être garante d'un ensemble de valeurs : démocratie, diversité (des paradigmes, points de vue, approches, méthodes, objets...), dimension collective, interdisciplinarité ;
- identifier des compétences (ou leur absence), des savoirs, dans le but de les développer, les promouvoir ;
- aider les acteurs collectifs et individuels de la recherche dans la promotion de leurs objectifs librement élaborés (évaluation conseil et non évaluation sanction) ;

- faire progresser la professionnalisation des acteurs de la recherche (équipes de recherche, chercheurs, réseaux, instances d'évaluation), en identifiant les difficultés rencontrées, les avancées ;
- favoriser les synergies.

Après avoir indiqué les finalités potentielles de la recherche en économie et celles de l'évaluation qui lui seraient liées, nous donnons dans la suite quelques éléments pour caractériser l'évaluation de la recherche.

L'évaluation devrait être considérée comme un composant de la politique nationale de recherche, et non uniquement un outil de management des acteurs de la recherche. Le rapport AFEP précise ainsi : « Parler de politique suppose d'inscrire les choix collectifs dans un projet de société publiquement exprimé et justifié (lui substituer celui de management contient implicitement l'idée qu'il n'y a pas de débat possible — ou nécessaire — sur les fins, mais seulement une expertise à développer sur les moyens) ».

L'évaluation devrait être reconnue comme un exercice situé dans le temps et dans l'espace. Un collectif de recherche a en effet une histoire, il est inscrit dans un contexte institutionnel. La recherche s'inscrit dans le temps long.

La *quantification* doit être mise en débat, car tout n'est pas réductible à des valeurs numériques, de même que toute valeur n'est pas assimilable à une valeur monétaire¹³.

En d'autres termes, il convient d'affirmer avec force que l'évaluation peut être autre que quantitative.

Les « objets » concernés par l'évaluation (laboratoire, chercheur, instance d'évaluation) pourraient, selon l'AFEP, être appréhendés comme suit.

Un *laboratoire* devrait être considéré comme un système (et non une somme de chercheurs « producteurs »).

Ses finalités devraient être :

- L'élaboration de questions de recherche ;
- La production de dispositifs de réponse à ces questions mobilisant les différentes dimensions de la recherche : épistémologique, théorique, conceptuelle, méthodologique, empirique ;
- La formation des chercheurs par et pour la recherche ;
- La fertilisation de l'enseignement par les méthodes et résultats de la recherche ;

¹³ Dans la même logique, il convient de rappeler que tout n'est pas comparable : un laboratoire d'économie ne peut être comparé à un autre, sauf de manière très (trop) globale.

- La diffusion des connaissances et du savoir vers des publics différenciés.

Un *chercheur* devrait être considéré comme libre et autonome, inséré dans un ou des collectifs de recherche. Sans négliger ses intérêts personnels, le chercheur, contrairement à ce que suggère la doxa dominante, n'est pas nécessairement un être calculateur, intéressé de façon principale par l'optimisation de sa carrière et son accès aux primes. Un chercheur ne saurait être réduit à une somme de publications « valides » au regard des critères usuels.

La ou les *instances d'évaluation* ne devraient pas être extérieures à l'évaluation. Ces instances devraient être elles-mêmes évaluées.

L'AFEP suggère enfin que soit posée la question des autres acteurs : acteurs « invisibles » tels que les réseaux, angle mort de l'évaluation actuelle...

Quelques principes de la doxa alternative

Après avoir évoqué les éléments d'une *doxa* alternative de l'évaluation de la recherche en économie, il convient à présent d'appréhender le niveau des principes de cette *doxa*.

Si la liste des principes évoqués ici est classée selon les mêmes catégories que pour la *doxa* dominante, leur sens et leur contenu diffèrent, bien sûr, sensiblement.

1. Langage

Il s'agit ici de revenir au langage de la recherche par l'emploi de termes comme : collégialité, pairs, coopération, projet, complémentarité, participation, demande sociale, éthique...

2. Conventions

Ce qui est reconnu comme étant une production scientifique inclut, selon le rapport de l'AFEP la prise en compte la plus large possible :

- de tous les types de publication
- des autres formes de « production »
- des activités mixant formation et recherche
- des pratiques de diffusion de la recherche.

Une question posée et non tranchée par le rapport de l'AFEP est celle de la pertinence d'une hiérarchie des productions scientifiques.

Une autre question est celle des conventions de quantification : quelles équivalences ? Sur quelles bases (qui doivent dans tous les cas être explicites et publiques) pondérer les éventuels comptages ?

3. Mode d'élaboration du processus d'évaluation

Le rapport de l'AFEP s'interroge quant à la nécessité d'une instance spécifique d'évaluation et est enclin à considérer qu'une telle instance n'est pas utile, que, pour la France, le Conseil National des Universités (CNU) suffit.

Si une telle instance doit être maintenue, alors elle devrait respecter les exigences suivantes :

- indépendance,
- transparence,
- être elle-même évaluée.

La question qui demeure posée est qui participera à la définition du processus d'évaluation et des exigences qu'il respectera ? Pour l'AFEP, ce processus doit inclure toutes les parties prenantes. Le choix des évaluateurs doit être fait par les pairs, élus.

4. Contenu du processus d'évaluation

Pour l'AFEP, le processus d'évaluation devrait :

- être pluraliste,
- démocratique,
- transparent,
- réellement contradictoire (incluant un droit de réponse des acteurs évalués),
- prendre en compte de la dimension longue du temps de la recherche,
- intégrer la dimension éthique de la recherche ; « dans quelle mesure les recherches menées sont-elles toujours acceptables éthiquement, en particulier dans les usages que l'on peut en escompter ? » s'interroge le rapport.

Pour ce qui est des méthodes d'évaluation, le rapport AFEP prône :

- la stabilité des méthodes, des critères et des procédures,
- la lecture des travaux.

Les types d'indicateurs utilisés doivent limiter le nombre et l'importance des indicateurs bibliométriques.

Le système de préférence devrait aller :

- au soutien aux thématiques de recherche innovantes,
- au soutien aux activités en réseau,
- à la prise en compte des apports méthodologiques,
- au soutien aux collectifs en difficultés ou en démarrage.

5. Structure des systèmes d'information

Rappelons que les bases de données utilisées dans la bibliométrie prêtent à certaines interrogations. En particulier se posent les questions de leur représentativité, de la hiérarchie qu'elles instaurent de fait entre les différents courants de pensée, les différentes régions du monde.

Les fonctions du système d'information doivent permettre de :

- préparer l'évaluation (venue du comité...),
- construire le rapport du laboratoire,
- consulter le rapport des évaluateurs,
- répondre au rapport des évaluateurs...

Quelques exemples de normes de la doxa alternative

Pour compléter les éléments que nous avons rassemblés d'une approche alternative de l'évaluation de la recherche en économie, nous donnons à présent quelques exemples succincts de choix à opérer au niveau des normes.

Un premier exemple est celui de la liste des *types* de productions de recherche à prendre en compte dans l'évaluation des chercheurs. L'examen du rapport de l'AFEP permet de dresser la liste suivante :

- publication : articles, chapitres d'ouvrages et ouvrages, rapports publics, etc., en fonction des projets et programmes scientifiques,
- autres formes de « production » : enquêtes, bases de données etc., dès lors qu'elles sont scientifiquement validées,
- activités mixant formation et recherche : encadrement, production de manuels et de documents pédagogiques...,
- pratiques de diffusion de la recherche, de l'action dans le débat public.

La question qui reste posée est, bien sûr, celle, déjà signalée plus haut, de la pondération éventuelle de chaque type de production.

Un deuxième exemple porte sur la liste des activités de recherche autres que celles de production scientifique, à prendre en compte dans l'évaluation des chercheurs. Pour l'AFEP, les activités qui devraient ainsi être prises en compte concernant :

- les responsabilités collectives de toutes sortes (coordination de contrats de recherche collectifs,
- la coordination de publications,
- la participation à des comités éditoriaux,
- la participation à la direction de masters, de laboratoires, etc.

Un troisième exemple concerne la norme de transparence. Pour l'AFEP, les évaluateurs doivent rendre compte obligatoirement de *tous* les critères d'évaluation dans leur délibération et leur rapport.

Un dernier exemple a trait à la procédure pour le choix des évaluateurs. Selon quelles normes et quels critères ces évaluateurs devraient-ils être choisis ? Le rapport de l'AFEP ne se prononce pas vraiment sur cette question.

5. CONCLUSION

L'évaluation actuellement dominante peut apparaître comme difficile à remettre en cause, et la promotion d'une évaluation alternative comme un long combat. Pour mener à bien celui-ci, la question essentielle, selon nous, est de déterminer le ou les aspects (ou composants) sur lesquels il faut se concentrer.

Notre position est que bien souvent la remise en cause d'une situation existante (ici, les pratiques dominantes d'évaluation) se concentre par trop sur le niveau des doxai (et ce n'est alors qu'une bataille rhétorique), ou, à l'inverse, sur telle ou telle norme, saisie isolément (par exemple le reclassement d'une revue donnée).

Dans la logique de l'outil que nous avons présenté, il conviendrait de se centrer sur le niveau des principes, et peut-être, plus secondairement, sur quelques normes.

La démarche que nous proposons pour ce faire consiste à identifier, par un travail collectif, les *principes verrous* qui ont provoqué des irréversibilités dans

la mise en œuvre de l'évaluation actuelle, et les *principes leviers* (très structurants) qui seraient le mieux à même de fonder l'évaluation alternative, en promouvant automatiquement des principes complémentaires et en générant un nombre conséquent de normes.

Ce travail reste à faire, mais les réflexions menées au sein de ce colloque et dans les ateliers qui l'ont suivi constituent un apport précieux dans cette perspective.

Ateliers d'élaboration de propositions

ATELIER 1: QUEL USAGE FAIRE DES INDICATEURS ?

ANIMÉ PAR HUGUES BERSINI (ULB) ET BRUNO LECLERCQ (ULG)

Rapidement, la discussion autour du choix des indicateurs et des usages plus ou moins pertinents qu'on peut en faire pour l'évaluation de la recherche a fait apparaître d'importantes différences entre disciplines ainsi qu'entre types d'évaluations (financement de projets, mandats de recherche doctorale, mandats de recherche postdoctorale, postes permanents de chercheurs ou d'enseignants-chercheurs...). Toutes les suggestions suivantes doivent donc être lues dans le cadre plus général d'une volonté de rendre aux commissions d'évaluation plus d'autonomie sur le choix des données utiles à éclairer leur décision, ainsi que sur la manière même dont ces données doivent être prises en compte. Modulés par les différentes commissions, en fonction des disciplines et des types d'évaluation, les critères d'évaluation devraient toutefois être définis dans une certaine stabilité à moyen ou long terme pour donner aux chercheurs une vision claire de ce qui est attendu d'eux.

La perspective dans laquelle nous nous situons récuse la possibilité d'une évaluation qui serait strictement « technique » ou neutre. Nous considérons que les indicateurs d'évaluation pertinents dépendent de *l'intention* de l'évaluation : non pas tant vérifier la conformité à une norme, sélectionner, *a fortiori* gérer la pénurie, mais bien conseiller, soutenir, favoriser les synergies.

Dans cette optique, sont formulées les 4 propositions concrètes suivantes :

Principe de diversité

Les commissions disciplinaires ou multidisciplinaires doivent pouvoir se prononcer (avant même le lancement des appels à candidatures) sur les informations qui doivent leur être fournies (et sur celles qui ne doivent pas l'être). Cela veut dire qu'elles doivent pouvoir agir sur les formulaires mêmes qui sont adressés aux candidats et, par exemple, ajouter ou au contraire supprimer certains cadres aux formulaires utilisés dans d'autres disciplines. Certaines commissions exigeront de connaître le *h-index* des candidats, d'autres non ; certaines commissions exigeront de connaître la maîtrise de langues étrangères, d'autres non... Et ces demandes peuvent évidemment varier d'un type d'évaluation à l'autre (le *h-index* pourrait par exemple, dans certaines disciplines, importer à partir des postes post-doctoraux et pas avant).

Principe de compensation du regard quantitatif par un regard qualitatif

Pour éviter la course au « toujours plus », les commissions devraient envoyer des messages clairs quant au fait qu'au-delà d'exigences minimales en termes de quantité (exigences que chaque commission pourrait préciser en fonction du type d'évaluation concernée), elles seront surtout attentives à départager les candidats d'après des critères qualitatifs plutôt que quantitatifs. En bref, les candidats qui auraient atteint un certain niveau de « production » scientifique (sur un ou plusieurs indicateurs) ne seraient pas tant encouragés à apporter des contributions toujours plus nombreuses qu'à apporter des contributions importantes. Cette dimension qualitative des contributions ne serait pas tant fournie par le nombre de citations auxquelles elles donnent lieu que par certaines appréciations d'experts (cf. proposition 3). Ici, signalons seulement la proposition que les commissions fournissent à l'avance des seuils quantitatifs*¹ — indicatifs ou impératifs — (sur différents indicateurs) à partir desquels les candidats sont éligibles et au-delà desquels il ne leur est pas (concurrentiellement) nécessaire de produire toujours plus de publications, dans la mesure où la sélection se fera aussi et surtout sur d'autres critères.

¹ Pour la fixation même des seuils minimaux d'éligibilité, il conviendra sans doute que les commissions pointent des objectifs tout à la fois scientifiquement pertinents et humainement raisonnables tout en veillant à éviter les mesures discriminantes à l'égard de certaines catégories de chercheurs, telles que les femmes qui ont ou voudraient avoir des maternités (cf. recommandations de la commission 3).

Principe de lecture des travaux pour s'assurer de leur originalité et de leur profondeur

Pour permettre cette évaluation qualitative, les experts pointus, sollicités par la commission pour donner avis sur les candidats, devraient se voir poser à leur égard des questions plus précises que celle de savoir ce qu'ils pensent du candidat d'après son CV. Ce type de sollicitation amène les experts à reproduire largement une évaluation quantitative qui peut déjà être faite au niveau de la commission (moins experte sur la recherche spécifique du candidat). Les experts pointus devraient être moins interrogés sur le CV des candidats que sur la qualité de leurs travaux dans le domaine. Pour ce faire, ils pourraient recevoir trois articles choisis par le candidat et, après lecture, se prononcer sur des questions telles que : « Les travaux de ce chercheur sont-ils originaux ? Apportent-ils des éléments susceptibles de renouveler le domaine sur un plan théorique ou méthodologique ? » ; « Les travaux du chercheur se distinguent-ils par leurs prises de risque théoriques ou méthodologiques ? » ; « Les travaux du chercheur se distinguent-ils par leur profondeur d'analyse ? » ; « Les travaux du chercheur se distinguent-ils par leur clarté d'exposition et leur capacité à se faire comprendre du public spécialisé ? du public savant non-expert dans le domaine ? d'un public plus large ? » ; ... La liste de ces questions pourrait évidemment varier d'une commission à l'autre et d'un type d'évaluation à l'autre.

Principe de responsabilité des chercheurs À l'égard des effets de leur recherche (et d'explicitation de leurs propres stratégies de recherche)

Dans les évaluations, les chercheurs devraient être autorisés, voire incités, par des cadres spécifiquement prévus dans les formulaires de candidatures ou d'évaluations (ou dans certains cas par un entretien oral avec la commission), à se prononcer sur :

- les enjeux sociétaux de leur recherche, qui ne se réduisent évidemment pas à la question de sa valorisation économique (cf. suggestions d'Isabelle Stengers sur cette question, notamment sur la responsabilité du chercheur à l'égard des usages potentiellement négatifs des résultats de sa recherche) ;
- leur parcours antérieur (et ce qui pourrait apparaître comme d'éventuels « détours » ou « accidents ») ;

- leur stratégie de publication, non seulement prospective (comme c'est déjà souvent le cas dans les évaluations) mais aussi rétrospective.

Ces questions seraient à nouveau modulées en fonction des disciplines et des types d'évaluation (la question des enjeux est sans doute moins évidente pour un chercheur qui entame sa thèse que pour un chercheur post-doc).

Ces 4 propositions entendent répondre à certains objectifs de ce qui nous a paru constituer une « bonne » évaluation. Nous sommes bien conscients que leur mise en œuvre pourrait se heurter à des obstacles pratiques sérieux (par exemple : réticences du FNRS à l'égard de formulaires de candidatures et de critères d'évaluation distincts d'une commission à l'autre, difficultés de trouver des experts qui acceptent de trouver le temps de lire les travaux des candidats, refus des instances dont dépendent les commissions (FNRS, universités...) d'appliquer des critères d'évaluation distincts de ceux qui président à leur propre évaluation/ranking...). La conviction qui a présidé à notre réflexion est toutefois que les conditions d'une bonne évaluation devraient prévaloir sur ce type d'obstacles, sous peine que, pour des raisons pratiques, on finisse par évaluer tout autre chose que ce qu'on voulait évaluer...

ATELIER 2: QUEL DIALOGUE ENTRE ÉVALUATEURS ET ÉVALUÉS ?

ANIMÉ PAR MICHEL CRUCIFIX (FNRS-UCL) ET BENOÎT
TIMMERMANS (FNRS-ULB)

Un dialogue préalable avec les représentants des évalués

Les membres de l'atelier recommandent qu'un dialogue soit mené préalablement au processus d'évaluation, entre les personnes organisant l'évaluation et celles qui seront évaluées ou sont susceptibles de l'être. Ce dialogue devrait avoir pour objet non seulement la procédure d'évaluation elle-même, mais aussi ses objectifs, son rythme, les décisions auxquelles elle peut donner lieu, et ce qui doit ou ne doit pas faire l'objet d'évaluation. Une réflexion devrait aussi être menée sur les règles éthiques d'une évaluation (par exemple le type d'arguments ou d'informations qui peuvent intervenir dans une discussion,

la manière dont les résultats d'une évaluation sont communiqués, la manière dont sont discutées les éventuelles réponses des personnes évaluées, etc.) Cette réflexion devait se faire sur base d'une représentation suffisamment large permettant de prendre en compte l'intérêt général tel qu'il est perçu par les chercheurs, ainsi que le point de vue syndical. Il importe aussi que les représentants des évalués prennent à cœur leur rôle de « représentants », c'est-à-dire tiennent informées les personnes qu'ils représentent et continuent de les consulter.

L'évaluation comme service à la recherche

Dans certains cas l'évaluation devrait pouvoir être envisagée non comme un moyen de sélection, de promotion, voire de sanction, mais comme un outil d'amélioration mis à la disposition du chercheur. C'est l'idée de l'audit, qui fait intervenir des évaluateurs externes plutôt que des instances hiérarchiques ayant une capacité de décision. La proposition n'est pas, ici, de multiplier les procédures d'audit imposées, avec tout le poids administratif que cela suppose. Il s'agirait plutôt d'offrir un service de type « coaching » ou accompagnement aux chercheurs ou équipes de recherche souhaitant y recourir, sans qu'aucune pression ne soit associée à une telle offre.

Les rencontres de débriefing

Il arrive souvent que des candidats ou projets, très bien classés dans une évaluation, ne soient finalement pas retenus par l'institution qui doit prendre en dernier ressort la décision. Il paraît souhaitable, au moins lorsque les candidats ou projets sont liés à l'institution en question, que celle-ci organise une rencontre afin d'expliquer sa décision.

La question du droit de réponse

Tous les participants de l'atelier s'accordent pour dire, qu'une réponse obligatoire ou automatique de la part des évalués, par exemple à un pré-rapport de synthèse d'évaluation, ne paraît pas souhaitable en raison, non seulement du surcroît d'administration que cela représenterait, mais aussi des effets de « banalisation » des réponses que cela pourrait produire. En revanche, il paraît utile de proposer un *droit* de réponse dans le cas de certaines

procédures à enjeu particulièrement important (par exemple la nomination d'un chercheur pour une durée indéterminée), compte tenu notamment de la possibilité d'un désaccord avec les arguments avancés par une commission d'évaluation ou l'un de ses experts, ou encore de la nécessité ressentie par un chercheur ou une équipe d'expliquer plus précisément les conditions dans lesquelles s'exerce sa recherche. L'atelier n'a pas discuté de la question de savoir si ce droit de réponse devrait intervenir en cours de procédure (avec les risques d'alourdissement administratif associés) ou après coup (avec le risque de ralentissement de la carrière de certains chercheurs).

Expliciter les rôles

La question a été posée de savoir s'il ne faudrait pas préciser en quoi consiste le mandat des experts et des membres d'une commission scientifique. Peut-on clairement identifier leurs rôles ? Par exemple, on ne demandera pas à un expert international de compter les publications ou les conférences, mais plutôt d'évaluer l'originalité, l'intérêt ou éventuellement la faisabilité d'un projet de recherche. Certains mandats plus précis peuvent également être définis, comme lire et commenter des articles. Toutefois il ne faudrait pas que cette clarification des mandats réduise les personnes chargées d'une évaluation à n'être que des exécutants. À chaque étape de l'évaluation les intervenants doivent pouvoir assumer la part subjective des jugements qu'ils sont amenés à poser.

La question de l'anonymat

Une autre question soulevée est celle de l'anonymat. Comment s'assurer qu'un expert sollicité pour évaluer une personne ou un projet est compétent dans le domaine précis qui est à évaluer ? Une possibilité pourrait-être de laisser le choix à l'expert de garder ou non son anonymat. Toutefois cette proposition ne fait pas consensus au sein de l'atelier, dans la mesure où le contexte de confidentialité peut favoriser une certaine sincérité des jugements et arguments.

ATELIER 3: QUELLE PLACE POUR LES ÉVALUÉS (PROJETS OU PERSONNES) EN SITUATION DÉFAVORABLE ?

ANIMÉ PAR BARBARA CLERBAUX (ULB) ET EDWIN ZACCAI (ULB)

Limiter la concentration des moyens : les différents instruments

Pour l'ensemble du groupe de travail de l'atelier, il a semblé important d'éviter la concentration à outrance des moyens, afin de favoriser une diversité dans la recherche. En effet, dans des cas de petites unités de recherche, le financement d'un projet peut être crucial, la survie de leur fonctionnement peut en dépendre. Le groupe de travail propose de favoriser et de développer davantage des instruments de financement différents, accessibles pour des profils de groupes de recherche variés. Dans certains cas, il est aussi utile de limiter les cumuls de financements ou les applications successives à des programmes de financement. Il est à noter que l'idée n'est pas ici de faire un « saupoudrage » du financement sur l'ensemble des acteurs, mais plutôt de soutenir une diversité d'équipes, petites ou grandes, jeunes ou plus matures, installées ou novatrices.

Évaluation de profils/projets éventuellement atypiques

Lors de l'évaluation d'une personne ou d'un projet, le groupe de travail de l'atelier propose de davantage tenir compte du temps effectif que le(a) candidat(e)/groupe peut consacrer à la recherche, et ce pour diverses raisons : choix d'un temps de travail partiel, activités en parallèle au sein de la société, mission de vulgarisation, etc. Dans le même contexte, le profil des chercheurs inclut les trois missions bien connues (recherche, enseignement, et soutien à la collectivité), l'évaluation doit tenir compte de chacun de ces aspects, et en particulier du temps que le chercheur peut consacrer à la recherche. Une diversité de critères en fonction des instruments est à définir, afin d'éviter un alignement systématique de l'évaluation sur les critères de financement d'un seul type d'instrument (ERC par exemple). Également le contexte général de recherche, notamment des moyens et de la taille du groupe dans lequel la recherche s'opère doit être pris en compte. Ceci ne doit pas nécessairement se faire au niveau de l'évaluateur, mais peut avoir lieu au niveau de la politique de l'université : stratégie pour gérer les priorités. Finalement, le groupe

de travail de l'atelier propose d'ajouter aux rapports d'évaluateurs externes recueillis en commission, une forme de droit de réponse de ceux qui sont évalués. Certains pays (Pays-Bas) utilisent déjà un système de ce type. Pour certains postes ou certains gros projets, une interview des candidats/portes-paroles est souhaitable.

Ouverture à la société

Le groupe de travail de l'atelier propose de favoriser dans certains cas des profils ouverts sur la société. D'une part, il est important de ne pas valoriser uniquement la recherche à portée internationale, mais également la recherche locale, étudiant des problèmes spécifiques à un lieu ou à un groupe de personnes. D'autre part, certains chercheurs et enseignants tels que des architectes, avocats, médecins cliniciens... (non limitatif) se nourrissent dans leur recherche de leur pratique, ce qui enrichit aussi leur enseignement. Pourtant ces profils ont été défavorisés dans certaines facultés, dans la mesure où ces personnes seront moins complètement investies dans la recherche. Finalement le groupe de travail de l'atelier remarque que si les missions des chercheurs/enseignants en ce qui concerne la recherche et l'enseignement sont bien définies, ce n'est pas le cas de leur mission par rapport à la société, et souhaite que celles-ci soient mieux définies.

Carrière

Les chances d'insertion de docteurs dans la carrière universitaire sont devenues faibles dans certains domaines au moins. En conséquence, il serait pertinent pour les universités de mieux réfléchir et de mettre en évidence les apports d'un grade de docteur. Certains parlent ainsi de formation « par la recherche » plutôt que de formation « à la recherche ». En France il semble y avoir plus de valorisation de ce titre.

Genre

Le groupe de travail de l'atelier propose de former/d'informer systématiquement les membres des commissions avec des données statistiques et autres à la question du genre dans les carrières. La Charte européenne du chercheur demande le meilleur environnement possible pour ceux-ci. Certains critères

pratiques pourraient être plus adaptés aux chercheuses. Viser une représentation des deux genres dans les commissions d'évaluation et de classement est indispensable. À cette fin, l'utilisation de quotas dans les commissions semble une solution intéressante.

ATELIER 4: COMMENT ÉVALUER L'INTERDISCIPLINARITÉ ? ANIMÉ PAR MAREK HUDON (ULB) ET ANTOINE NONCLERCQ (ULB)

Temps nécessaire

De nombreux chercheurs estiment que l'interdisciplinarité nécessite plus de temps, notamment afin de faire émerger de la connaissance, du savoir à la frontière entre plusieurs domaines de connaissance. Un même objet peut ainsi être compris et analysé différemment selon la manière dont on le voit d'une discipline à l'autre. Ceci semble un élément important dans l'évaluation de cette interdisciplinarité, car de nombreux évaluateurs ne perçoivent pas comment les différentes équipes interagissent et peuvent interagir.

Distinguer multi- et interdisciplinarité

Nous nous sommes aussi rendu compte que les projets de recherche se rapprochent plus souvent de la multi ou pluridisciplinarité alors que leurs promoteurs pensent faire de l'interdisciplinarité. C'est ainsi que plusieurs disciplines analysent un même objet, mais ce n'est pas pour autant qu'il y a quelque chose de commun à celles-ci qui émerge. Cette émergence d'un objet commun, un article par exemple, prend aussi beaucoup de temps alors qu'on sait que l'évaluation est souvent disciplinaire, facultaire, avec des grilles d'analyse différentes. Quand un objectif commun se dessine, on a rapidement tendance à compartimenter les étapes nécessaires pour l'atteindre en fonction des disciplines, plutôt que de promouvoir l'approche interdisciplinaire.

Compétences des évaluateurs

Dans les propositions, nous nous sommes rendu compte que, souvent, une connaissance disciplinaire approfondie permet de mieux faire émerger de

l'interdisciplinarité. Cependant, les évaluateurs doivent combiner connaissance disciplinaire avec une autre compétence liée à la transversalité des disciplines considérées. Il semble essentiel que les évaluateurs aient un profil interdisciplinaire. Dans certains cas, cela requiert de demander aux personnes qui évaluent les projets interdisciplinaires à quel point ils pensent être compétents dans l'un ou l'autre élément disciplinaire ou méthodologique du dossier. Ainsi, alors que le FNRS demande à quel point on est lié à l'une discipline ou l'autre discipline, on pourrait imaginer renvoyer cela à l'expert – si cela ne se fait pas déjà – et cela permettrait d'aider à clarifier les compétences de chacun.

Outils de développement de l'interdisciplinarité

Un dernier élément dans les propositions émanant de l'atelier est de mettre en place des endroits spécifiques qui permettent de développer cette interdisciplinarité. Nous nous sommes rendu compte qu'on met fréquemment l'accent sur l'approche, mais sa mise en œuvre est remplie d'obstacles. Une aide à la structuration de projets particulièrement interdisciplinaires ou de centres de recherche interdisciplinaires pourrait leur permettre de se développer plus facilement. Des formations spécifiques à l'approche interdisciplinaire existent aussi dans certaines institutions.

COLLECTION « MÉMOIRES DE LA CLASSE DES SCIENCES »

Denis Diagre, *Le Jardin botanique de Bruxelles (1826–1912). Reflet de la Belgique, enfant de l’Afrique* (2012)

Georges Bernier, *Darwin, un pionnier de la physiologie végétale. L’apport de son fils Francis* (2013)

Jacques Reisse, *Alfred Russel Wallace, plus darwiniste que Darwin mais politiquement moins correct* (2013)

Francis Robaszynski, Francis Amédéo, Christian Devalque et Bertrand Matrion, *Le Turonien des massifs d’Uchaux et de la Cèze* (2014)

Pierre Verhas, *L’histoire de l’Observatoire royal de Belgique* (2014)

L’Homme, un animal comme un autre ? Actes de colloque édités par Jacques Reisse et Marc Richelle (2014)

De Mons vers le Nouveau Monde. Lettres de Jean-Charles Houzeau en Jamaïque (1868–1876), Hossam Elkhadem et Marie-Thérèse Isaac (ed.) (2015)

Frédéric Boulvain et Jean-Louis Pingot, *Genèse du sous-sol de la Wallonie* (2015, 2^e éd. revue et augmentée)

Émile Biémont, *Le règne du temps : des cadrans solaires aux horloges atomiques* (2016)

L’évaluation de la recherche en question(s). Actes de colloque (2016)

COLLECTION « MÉMOIRES » DE L’ACADÉMIE ROYALE DE BELGIQUE

Les Minorités, un défi pour les États. Actes de colloque (2012)

L’idéologie du progrès dans la tourmente du postmodernisme. Actes de colloque (2012)

Musique et sciences de l’esprit. Actes de colloque (2012)

Catherine Jacques, *Les féministes belges et les luttes pour l’égalité politique et économique (1914–1968)* (2013)

Athéisme voilé/dévoilé aux temps modernes. Actes de colloque (2013)

Stéphanie Claisse, *Du Soldat Inconnu aux monuments commémoratifs belges de la Guerre 14–18* (2013)

La démocratie, enrayée ? Actes de colloque (2013)

Catherine Thomas, *Le visage humain de l’administration. Les grands commis du gouvernement central des Pays-Bas espagnols (1598–1700)* (2014)

La bataille de Charleroi, 100 ans après. Actes de colloque (2014)

Pierre Assenmaker, *De la victoire au pouvoir. Développement et manifestations de l’idéologie impériatoriale à l’époque de Marius et de Sylla* (2014)

Le Quatrième partage de la Pologne. Actes de colloque (2015)

La liberté d’expression. Menacée ou menaçante ? Actes de colloque (2015)

Robert Wangermée et Valérie Dufour (dir.), *Modernité musicale au XX^e siècle et musicologie critique. Hommage à Célestin Deliège*. Actes de colloque (2015)

Jean-Louis Migeot, *Des chiffres et des notes. Mathématique et solfège, physique et musique : une introduction* (2015)

Hugues Bersini, *Quand l’informatique réinvente la sociologie* (2015)

Jean-Louis Kupper, *Notger de Liège (972–1008)* (2016)

Jean-Charles Speeckaert, *Dominique de Lesseps. Un diplomate français à Bruxelles au temps du renversement des alliances (1752–1765)* (2016)

COLLECTION « L'ACADÉMIE EN POCHE »

1. Véronique Dehant, *Habiter sur Mars ?* (2012)
2. Xavier Luffin, *Religion et littérature arabe contemporaine* (2012)
3. François De Smet, *Vers une laïcité dynamique. Réflexion sur la nature de la pensée religieuse* (2012)
4. Richard Miller, *Liberté et libéralisme ? Introduction philosophique à l'humanisme libéral* (2012)
5. Ivan P. Kamenarovic, *Agir selon le non-agir. L'action dans la représentation idéale du Sage chinois* (2012)
6. Jean Mawhin, *Les histoires belges d'Henri Poincaré* (2012)
7. Jacques Siroul, *La musique du son, ce précieux présent* (2012)
8. Baudouin Decharneux, *La religion existe-t-elle ?* (2012)
9. Jean-Marie Rens, *'Messagesquise' de Pierre Boulez* (2012)
10. Jean de Codd, *Faut-il s'inspirer de la justice américaine ?* (2013)
11. Bruno Colmant, *Voyage au bout d'une nuit monétaire* (2012)
12. Philippe Manigart et Delphine Resteigne, *Sortir du rang. La gestion de la diversité à l'Armée belge* (2013)
13. Hervé Hasquin, *Les pays d'islam et la Franc-maçonnerie* (2013)
14. Monique Weis, *Marie Stuart, l'immortalité d'un mythe* (2013)
15. Xavier Luffin, *Printemps arabe et littérature. De la réalité à la fiction, de la fiction à la réalité* (2013)
16. Myriam Rimmelink, *Éthique et biobanque. Mettre en banque le vivant* (2013)
17. Marie-Aude Baronian, *Cinéma et mémoire. Sur Atom Egoyan* (2013)
18. Frédéric Boulvain et Jacqueline Vander Auwera, *Voyage au centre de la Terre* (2013)
19. Daniel Salvatore Schiffer, *Métaphysique du dandysme* (2013)
20. Philippe de Woot, *Repenser l'entreprise. Compétitivité, technologie et société* (2013)
21. Jacques Scheuer, *L'Inde, entre hindouisme et bouddhisme. Quinze siècle d'échanges* (2013)
22. John F. May, *Agir sur les évolutions démographiques* (2013)
23. Yaël Nazé, *À la recherche d'autres mondes. Les exoplanètes* (2013)
24. Jean Winand, *Les hiéroglyphes égyptiens. Aux origines d'une écriture* (2013)
25. Frans C. Lemaire, *Dimitri Chostakovitch. Les rébellions d'un compositeur soviétique* (2013)
26. Baudouin Decharneux, *Lire la Bible et le Coran* (2013)
27. Bruno Colmant, *Capitalisme européen : l'ombre de Jean Calvin* (2013)
28. Françoise Meunier, *Quel avenir pour la recherche clinique en cancérologie ?* (2014)
En anglais : Françoise Meunier, *What is the future of cancer research ?* (2014)
29. Jean Winand, *Décoder les hiéroglyphes. De l'Antiquité tardive à l'Expédition d'Égypte* (2014)
30. Jacques Joset, *Louis-Ferdinand Céline : mort et vif... !* (2014)
31. Jean-Baptiste Baronian, *La littérature fantasmagorique belge. Une affaire d'insurgés* (2014)
32. Valérie André, *La rousseur infamante. Histoire littéraire d'un préjugé* (2014)
33. Jean-Pierre Contzen, *Les menaces venant de l'espace* (2014)
34. François Mairesse, *Le culte des Musées* (2014)
35. Guy Haarscher, *La Cour suprême des États-Unis. Les droits de l'Homme en question* (2014)
36. Catherine de Montlibert, *L'émancipation des serfs de Russie. L'année 1861 dans la Russie impériale* (2014)
37. Jean-Pol Poncelet, *Une énergie dérangeante. Nucléaire : une controverse durable ?* (2014)
38. Philippe Samyn, *La ville verticale* (2014)
En anglais : Philippe Samyn, *The Vertical City* (2014)
39. François De Smet, *Une nation nommée Narcisse* (2014)
40. Jean-Pol Schroeder, *Le jazz comme modèle de société. Livre-disque, avec la participation du Steve Houben trio* (2014)
41. Jean-Pierre Hansen, *Une quête de Graal* (2014)
42. Hervé Hasquin, *Déconstruire la Belgique ? Pour lui assurer un avenir ?* (2014)
43. Philippe de Woot, *L'innovation, moteur de l'économie* (2014)

44. Bruno Colmant, *Crises économiques et dette publique* (2014)
45. Samuele Furfari, *L'énergie, de la guerre vers la paix et la stabilité* (2014)
46. Samuel Wajc, *Que faire de la mer Morte ?* (2014)
47. Gilbert Hottois, *Le transhumanisme est-il un humanisme ?* (2014)
48. Benoit Frydman, *Petit manuel pratique de droit global* (2014)
49. Alain Eraly, *Quand les mots construisent la réalité* (2014)
50. Xavier Dieux, *Le marché bien tempéré* (2014)
51. Marc Wilmet, *Petite histoire de l'orthographe française* (2015)
52. Amand A. Lucas, *Les savants d'Hitler et la bombe atomique* (2015)
53. Jean-Marie André, *Fleuve jaune, papillons amoureux et musique classique de la Chine du XX^e siècle* (2015)
54. Françoise Lauwaert, *Puissance et pouvoir de l'écriture chinoise* (2015)
55. Jean-Pol Poncelet, *À toute ardeur ! Science et technique sur le chemin de l'énergie* (2015)
56. Jacques Pélerin, *Wallonie, réindustrialisation et innovation. « Sortir par le haut ? »* (2015)
57. Jacques Joset, *Louis-Ferdinand Céline: la manie de la perfection... !* (2015)
58. Daniel Salvatore Schiffer, *Le clair-obscur de la conscience* (2015)
59. Jean-Marie Frère, *La résistance des bactéries aux antibiotiques* (2015)
60. François de Callataÿ, *Cléopâtre, usages et mésusages de son image* (2015)
61. Anne Staquet, *Descartes avance-t-il masqué ?* (2015)
62. Guillaume Wunsch, Michel Mouchart et Federica Russo, *Les limites de la connaissance en sciences sociales. L'explication mise en cause* (2015)
63. Vincent De Coorebyter, *Deux figures de l'individualisme* (2015)
64. Daniel Droixhe, *Fer ou ciguë ? Récits sur le cancer du sein au 18^e siècle* (2015)
65. Véronique Dehant, *Habiter sur une lune du système solaire ?* (2015)
66. Pierre Somville, *Pour une esthétique du coeur* (2015)
67. Jean-Pierre Schaecken, *Pic pétrolier, pic gazier sans cesse reportés* (2015)
68. Jean Creplet et János Frühling, *Penser les soins de santé* (2015)
69. Frédéric Boulvain et Francis Tourneur, *Pierres et marbres en Wallonie* (2016)
70. Marc Wilmet, *Il y a grammaire et grammaire* (2016)
71. Pierre Petit, Patrice Lumumba. *La fabrication d'un héros national et panafricain* (2016)
72. Viviane Pierrard, *Les colères du Soleil* (2016)
73. Philippe de Schoutheete, *La création de L'Euro* (2016)
74. Pierre Somville, *Brasillach écrivain, mal-aimé des Lettres françaises* (2016)
75. Roland Souchez, *Glaces polaires et évolution de l'atmosphère* (2016)
76. Hervé Hasquin, *Le soi-disant « Gladio belge »* (2016)
77. Francis Delpérée, *J'écris ton nom, Constitution* (2016)
78. Christophe Van Staen, *La Chine au prisme des Lumières françaises* (2016)



L'évaluation de la recherche scientifique constitue un sujet d'une grande importance. D'abord, pour les chercheurs et professeurs dans le monde académique. Ensuite, pour les institutions qui les emploient et les financent. Enfin, pour la société qui entretient des attentes envers les connaissances scientifiques et leurs applications.

Quels modes d'évaluation sont aujourd'hui pertinents et appropriés à la diversité des disciplines, des pratiques et des objectifs de recherche? Cette question se pose avec acuité dans le cadre de bouleversements induits par l'informatisation, l'internationalisation, les contraintes financières des pouvoirs publics, et plus largement par les évolutions contemporaines des sociétés et de l'économie.

Cet ouvrage fait le point sur ce sujet avec des chercheurs belges et internationaux, réunis à l'occasion d'un colloque « Penser la science ».

Avec les contributions de :

Hugues Bersini, Barbara Clerbaux, Gabriel Colletis, Alain Eraly, Bernard Fusulier, Yves Gingras, Véronique Halloin, Marek Hudon, LAC (L'Atelier des Chercheurs), Bruno Leclercq, Pierre Marage, Dominique Pestre, Maryse Salles, Isabelle Stengers, Benoît Timmermans, Christian Vandermorten, Edwin Zaccai.



www.academie-editions.be

ISSN 00365-0936

ISBN 978-2-8031-0561-8

MEM-8-IV S-7 2113

