

Vie de la recherche

L'évaluation de la recherche en question : pratiques, dérives et alternatives

Marine Lugen¹, Quentin Hiernaux²

¹ Doctorante en environnement, Université libre de Bruxelles, Centre d'études du développement durable, Bruxelles, Belgique

² Doctorant en philosophie, Université libre de Bruxelles, Centre de recherche en philosophie, Bruxelles, Belgique

Mots-clés :
 évaluation ; recherche ;
 indicateurs ;
 universités ; expertise

Résumé – L'évaluation produit des effets, désirés ou non, sur la recherche, ses orientations et ses résultats. Dans le cadre des séminaires Ilya Prigogine « Penser la science » à l'Université libre de Bruxelles, deux journées ont été consacrées à ce thème. La première a réalisé un état des lieux de la question, en étudiant les effets standardisants des indicateurs et des classements et en s'interrogeant sur le statut des chercheurs dans un contexte de compétition et sous-financement. La seconde journée a envisagé des propositions alternatives : valoriser les formes d'intelligence collective, adapter les indicateurs aux disciplines, repenser l'utilité et la responsabilité face aux résultats, récompenser la prise de risque, retourner à des démarches qualitatives... Loin de critiquer la notion d'évaluation *per se*, ces pistes appellent à des formes d'évaluation plus démocratiques, transparentes et contradictoires.

Keywords:
 evaluation; research;
 indicators;
 universities; expertise

Abstract – **Research evaluation called into question: practices, distortions and alternatives.** The evaluation of research, driven by the development of computer tools and the desire to "objectify" researchers and research projects, produces effects, desired or undesired, on research, its directions and results. As part of the Ilya Prigogine Seminars "Penser la Science" at Université libre de Bruxelles, two days were devoted to this topic. The first day focussed on assessing the situation by deconstructing the evaluation of research and related concepts. Participants discussed indicators, rankings and effects of standardization on individuals, research themes and institutions. Comments were made on the hegemony of certain disciplines, journals and Anglo-Saxon culture. The status of research actors and their working conditions in a highly competitive and underfunded context were also considered. The second day examined alternative proposals regarding the current situation. Several possibilities were discussed: enhancing forms of collective intelligence and various knowledge transfer mechanisms; adapting indicators to individual disciplines and proposing specific tools for interdisciplinarity; rethinking the usefulness of and accountability towards research results; rewarding risk-taking and originality; returning to qualitative rather than quantitative approaches... Far from criticizing the notion of evaluation *per se*, these proposals called for more democracy, transparency and contradictory forms of evaluation.

Le colloque intitulé « L'évaluation de la recherche en question » a été organisé à l'Université libre de Bruxelles (ULB) dans le cadre des séminaires Ilya Prigogine « Penser la science » qui visent à promouvoir échanges et

débats appelés à nourrir une culture des sciences¹. Le thème de l'évaluation fut suffisamment riche pour couvrir deux journées. La première, le 9 mai 2015, portait sur les instruments et les acteurs de la recherche et de son

Auteur correspondant : . Lugen, marine.lugen@ulb.ac.be

¹ Le comité scientifique et organisateur du colloque incluait : Hughes Bersini, Barbara Clerbaux, Marek Hudon, Bruno Leclercq, Benoît Timmermans et Edwin Zaccai. Le colloque a également bénéficié de la participation d'Alain Eraly (sociologue, ULB) comme président de session. Le programme détaillé est disponible sur internet : <http://penserlascience.ulb.ac.be/seminaires/2015/article/seminaire-penser-la-science-2015>. Les actes issus de ce colloque sont publiés par l'Académie royale de Belgique : Zaccai E., Timmermans B., Clerbaux B., Leclercq B., Bersini H. (Eds), 2016. *L'évaluation de la recherche en question(s)*, Bruxelles, Académie royale de Belgique, http://penserlascience.ulb.ac.be/IMG/pdf/penserlascience_2015_actes_complets.pdf.

évaluation. La seconde journée, le 16 octobre, a proposé des alternatives à certaines des pratiques existantes².

L'évaluation de la recherche : pratiques actuelles et dérives

Dominique Pestre (historien des sciences et des rapports entre savoirs et sociétés, École des hautes études en sciences sociales) a inauguré le colloque en analysant les formes plus larges dans lesquelles s'insère, ou dont émerge, l'évaluation. Forme de gouvernement des institutions et des individus, elle s'inscrit aujourd'hui dans une logique de gouvernementalité libérale qui se centre sur la réinstallation de l'individu en charge de lui-même et la mise en place de dispositifs lui permettant de s'améliorer et de mieux se gérer. Cette logique découle du modèle de management d'entreprise. La fixation d'objectifs de bon sens, évidents et incontestables (excellence, médecine des preuves, amélioration de soi...) précède la définition d'une métrique à partir de laquelle les meilleurs individus et pratiques sont distingués et *de facto* encouragés. La clé du modèle est ainsi la définition d'un *modus operandi* pragmatique, permettant de faire circuler l'information objective nécessaire pour progresser et se mesurer aux autres. Il ne s'agit jamais de discuter des objectifs, de s'attarder sur un débat d'idées, d'ouvrir la porte aux intellectuels critiques ; le conflit et le désaccord sont absents du modèle. La création de la métrique conduit à l'établissement d'un *benchmark* (standard) qui renvoie au paradoxe de ce système libéral : pour permettre aux individus et institutions de s'affirmer, il faut d'abord être extrêmement interventionniste. Les personnes sont partagées entre les meilleures et les plus mauvaises et sont normalisées. Les palmarès connaissent une ascension irrésistible entretenue par la dynamique propre de l'espace public qui se charge de leur diffusion et circulation. Il devient extrêmement difficile d'y échapper et de ne pas en tenir compte. Une tyrannie de la

pseudo-transparence se met en place. On constate alors un effacement progressif de séparations qui font la liberté des Modernes selon Benjamin Constant, soit une autonomie partielle du politique par rapport à l'économique et de l'espace personnel par rapport à la culture d'entreprise.

Poursuivant la déconstruction de l'évaluation, Bruno Leclercq (philosophe, Université de Liège [ULg]) a mis en exergue quelques-uns de ses concepts-clés. Le premier est la prégnance des indicateurs de performance, au moyen desquels on veut objectiver la recherche à des fins de comparaison (*benchmarking*) et de classement (*ranking*). Les indicateurs offrent des données standardisées, de préférence chiffrées, qui quantifient la qualité à de multiples niveaux (tant les individus que les équipes ou les institutions) et sous tous ses aspects (enseignement et recherche). Loin d'être accessoires, ils ont des effets profonds sur la sélection des chercheurs et des projets, la répartition des financements, etc.

La reddition des comptes est le second concept qui relève d'une logique de marché. Il s'agit d'une exigence d'imputabilité et d'*accountability*, c'est-à-dire la fixation d'objectifs chiffrés et la vérification régulière qu'ils sont atteints. On retrouve ici la nécessité de disposer d'une information standardisée (ou objective, *cf. supra*) permettant de tirer le meilleur bénéfice possible des services en jugeant leur performance. C'est le credo de la « nouvelle gestion publique ». L'information standardisée permet d'effectuer un choix rationnel entre un panel d'objets possibles, mais aussi de justifier d'une transparence liée à ce choix.

Cette notion de transparence est capitale : c'est l'idée de gouverner par le nombre et par les faits (l'objectif) au nom de la chasse à l'arbitraire (le subjectif), idée que B. Leclercq et Catherine Fallon (sociologue, ULg) identifient sous le terme de « paradigme de la transparence gestionnaire et de sa logique extensionnelle ». Cette logique entend s'opposer à une certaine opacité de contextes intentionnels tels que celui de la donation de sens, de la prise de position ou de la décision politique. Elle procède en instituant des mises en équivalence par des calculs et des chiffres qui imposent un choix indépendamment du sens donné à ce choix. Si le monde académique est traditionnellement porté par une logique de réputation, au départ intentionnelle, celle-ci devient de plus en plus extensionnelle et par conséquent chiffrée : elle se forme aujourd'hui par le biais d'indicateurs de performance standardisés permettant de comparer l'ensemble des chercheurs, y compris à distance. L'intégration du modèle s'explique par un certain consensus autour de ces critères, qui sont confortables, mais aussi par un processus de renforcement continu du système : une fois en place, les indicateurs sont extrêmement normatifs et deviennent incontournables.

² À noter que les pratiques en Belgique s'écartent pour certains aspects de la recherche en France. Si la part cumulée des dépenses publiques et privées pour la financer y est relativement similaire, avec dans les deux cas une proportion largement majoritaire des fonds publics, ces derniers sont plutôt centralisés en France et régionalisés en Belgique. On constate en outre, dans les dix dernières années, davantage d'efforts pour accroître le budget en France, particulièrement via le développement du financement sur projets. Celui-ci implique une diminution de la stabilité des investissements dans les laboratoires et une priorité accordée aux contrats à court ou moyen terme. Dans les deux pays, notamment chez les jeunes chercheurs, l'accès à des postes permanents est extrêmement rare et tournerait autour de 10 % pour les titulaires d'un doctorat.

L'usage des indicateurs

Reprenant la distinction opérée par Kant entre jugements déterminants et jugements réfléchissants³, B. Leclercq a estimé que la prévalence des indicateurs tend à conduire à des évaluations de type déterminant. Ils pourraient être souhaitables pour autant (1) qu'ils couvrent bien toutes les dimensions de la qualité et (2) qu'ils permettent vraiment de se faire une idée précise plutôt que partielle et partielle de ces dimensions.

Yves Gingras (historien et sociologue des sciences, Université du Québec à Montréal) est revenu sur leur usage dans l'évaluation et a montré que les conditions précitées ne sont pas remplies. La course aux indicateurs est assez récente en Europe. Un repère est 2003, année de la publication du premier classement de Shanghai entre universités. Or, évaluation et classement ne sont pas les mêmes choses. Le problème n'est pas l'évaluation *per se*, mais le remplacement d'une démarche qualitative par une approche quantitative qui ne respecte pas toujours les critères requis pour un bon indicateur (adéquation à l'objet, homogénéité des mesures, variations conformes à l'inertie de l'objet, relation monotone croissante entre le concept mesuré et l'indicateur choisi). Par exemple, la création du « facteur d'impact » en 1975 fut la première tentative d'exprimer sous une forme numérique l'importance ou l'influence des revues. Pour un journal donné, ce facteur est calculé en divisant le nombre de citations de ses articles publiés par le nombre total de ses articles publiés durant une période de deux (ou cinq) ans. Une première faiblesse apparaît tout de suite : les temporalités des disciplines étant différentes, la comparabilité n'est pas pertinente.

Remplacer l'évaluation par des classements produit des effets sur la recherche, notamment de standardisation. Au niveau des individus, cela induit une modification des comportements : usage stratégique des citations, choix des objets de recherche (surtout dans les sciences sociales), abandon des objets locaux. Selon Y. Gingras, les classements sont non seulement manipulables⁴, mais également techniquement faux. Par exemple, les importantes variations d'une année à l'autre du classement des universités ne respectent pas le critère de conformité à l'inertie de l'objet évalué.

Certaines nuances ont été apportées par Hughes Bersini (ingénieur physicien, ULB), qui a mené une réflexion sur le h-index (indice h, en français) s'appliquant surtout à des personnes. Il s'agit selon lui d'un compromis ingénieux entre le nombre de publications et

³ Les jugements déterminants consistent à appliquer à des situations nouvelles des catégories morales dont on disposait préalablement, tandis que les jugements réfléchissants s'interrogent, face à la singularité de ce qu'on doit juger, sur les concepts mêmes qui devraient guider le jugement.

leur impact référentiel. Le h-index les combine par l'identification d'un nombre unique X, correspondant à X publications ayant été citées plus de X fois. Publier et être référencé est en effet jugé fondamental dans la science pour un certain nombre de raisons (faire connaître ses travaux, les faire évaluer et avaliser par les pairs, rendre compte de la progression douce de la connaissance...). Les classements, facilités par un nombre unique tel que celui proposé par le h-index, permettent de repérer et d'identifier des experts et d'orienter les carrières. Toujours est-il que certaines failles ont encore été évoquées, applicables à ce cas précis : mieux adapté à certaines disciplines que d'autres, le h-index est aussi sensible à l'âge et à la banque de données, à des effets sur les contenus, et ne favorise pas la diversité ni la prise de risques.

Quelques exemples disciplinaires

Marek Hudon (sciences de gestion, ULB) et Barbara Clerbaux (physicienne, ULB) ont comparé les pratiques d'évaluation de leurs disciplines. Ils notent que la transparence et le retour à l'intéressé (soit l'évalué) ont émergé comme critères dans les comités. Un autre élément est la tendance à la standardisation par un alignement sur des critères internationaux. En comparant sciences de gestion et physique, plusieurs constats se recourent : le nombre de publications, la qualité et la renommée des revues, les collaborations internationales et les facteurs d'impact sont repris comme indicateurs dominants. Les parents pauvres tendent à être les mêmes : les projets interdisciplinaires, publiquement plébiscités mais relativement exclus en pratique, et la question des transferts de savoir, notamment les processus de vulgarisation et l'enseignement qui ne sont pas valorisés, ce qui pourrait à long terme aggraver le fossé entre sciences et société. Une tendance inquiétante est relevée dans la discipline économique uniquement : le fait que des poids numériques très précis sont attribués à l'avance à certaines revues, impliquant une classification quasi établie dans les commissions avant évaluation.

Christian Vandermotten (géographe, ULB) nous a proposé une réflexion sur l'évolution et la diversité des revues disciplinaires suite au *ranking* dans le cas de la

⁴ À cet égard, notons que les classements font référence aux principes de Berlin, définis en 2006 par l'autoproclamé « *International Ranking Expert Group* », composé précisément de ceux qui établissent ces classements. Il s'agit de : définir clairement les objectifs ; être transparent dans la méthodologie ; choisir des indicateurs adéquats et valides ; identifier clairement la pondération ; reconnaître la diversité des institutions. D'après Y. Gingras, en pratique, aucun de ces principes n'est appliqué et tous les systèmes existants seraient considérés comme non valides si on appliquait ces critères à chaque indicateur. En outre, aucun ne tient compte de la spécificité des disciplines.

géographie. Ce choix est intéressant parce que c'est une discipline mal positionnée dans les classements et très hétérogène, qui subit ainsi des évaluations où, souvent, les experts ne sont pas géographes. En outre, la géographie humaine a un fort ancrage dans les problématiques sociétales locales alors que les critères d'évaluation sont souvent généralistes et non adaptés à chaque langue.

Son étude a comparé la situation actuelle et celle antérieure à la Seconde Guerre mondiale. Il constate une explosion du nombre de revues avec une hégémonie de l'anglais, induisant des distorsions dans leurs h-index selon leur langue et leur lieu d'édition ainsi que des disparités dans la proportion de revues non anglophones reprises dans certaines banques de données. Par exemple, le *Social Science Index* reprend 69 % des 723 revues publiées en anglais et recensées par l'Union géographique internationale, 50 % parmi les 14 publiées en allemand, 12 % parmi les 57 publiées en français, 4 % parmi les 240 publiées en chinois et 3 % parmi les 108 publiées en espagnol. Les titres sont en outre de plus en plus spécialisés. Les pratiques disciplinaires globalisantes ont ainsi tendance à s'effacer au profit de recherches très thématiques. Enfin, C. Vandermotten a constaté une prise de contrôle des grands groupes d'édition sur les revues historiques des sociétés de géographie quand elles sont publiées en anglais.

Sur le plan épistémologique et social, il s'ensuit la diffusion d'une vision hégémonique standardisée, d'une manière de publier et de véhiculer des concepts dominée par la conception anglo-saxonne. D'autres conséquences sont la dévalorisation des livres au profit des articles et le recours presque obligé à la citation de quelques « grands noms » de son domaine, évoqués non comme références sur un point précis, mais comme sortes d'argument d'autorité ou de « preuve » de sérieux scientifique. Enfin, les classements bibliométriques reconnus négligent le critère de l'utilité sociale de la production scientifique et des revues. À l'ULB, par exemple, le message véhiculé est de valoriser l'utilité sociale et l'enseignement. Pourtant, dans les processus de sélection, ces critères sont inexistant ou peu pris en compte.

Les acteurs de la recherche

Bernard Fusulier (sociologue, Université catholique de Louvain) a rendu compte des résultats d'une recherche⁵

⁵ En 2010, B. Fusulier lance un programme de recherche sur la question du genre et des carrières scientifiques. D'abord non financé, ce programme reçoit ensuite l'appui du Fonds de la recherche scientifique (FNRS) belge et prend une dimension européenne. Le premier volet visait les post-doctorants et chargés de recherche du FNRS qui se trouvent souvent dans un cycle de vie déterminant – tant sur le plan professionnel que privé.

sur les conditions du travail scientifique à travers l'expérience de chercheuses et chercheurs.

L'université y apparaît comme un lieu de conflits et de lutte pour l'accès à la reconnaissance et aux positions valorisées selon des critères et des règles propres, dont la définition devient nécessairement un enjeu stratégique. C'est aussi une « *greedy institution* » au sens de Coser⁶, qui exige de ses membres une implication totale et volontaire dans l'université. Cette dynamique nécessite une grande disponibilité de la part du chercheur et un investissement sans partage, ce qui a des impacts sur sa vie personnelle également. Cela crée en outre une forme de division sexuelle du travail, due à une inquiétude renforcée chez les femmes de l'interférence existant entre profession et vie familiale. S'y ajoute la dimension de l'excellence dont l'équation se résume à produire beaucoup, être compétitif, rendre des comptes et avoir une mobilité internationale. La mesure de ces critères, dans un contexte de précarité professionnelle, s'accompagne d'une concurrence croissante.

Les résultats de l'étude sont parlants. Si l'on observe un rapport très positif au travail qui, tant du point de vue des contenus de recherche que de son environnement, demeure hautement apprécié⁷, le statut, par contre, récolte de nombreuses critiques : l'incertitude quant à la possibilité de poursuivre le travail scientifique est énorme, en raison notamment de la concurrence accrue et du sous-financement de la recherche. Bien que la grande majorité des personnes consultées (90 %) espère un poste fixe, la probabilité d'obtenir un tel poste est très faible et fait naître de fortes incertitudes, avec des différences, d'une part, entre les femmes et les hommes, et, d'autre part, entre une situation avec ou sans enfant (les mères étant les plus pessimistes). En résultent une tension travail/famille et un sentiment d'empiétement dans les deux sens. Globalement, même si une carrière scientifique favorise chez les chargés de recherches un engagement couplé à une grande liberté de pensée, d'action et d'organisation, cette situation oblige en même temps qu'elle libère : il n'y a pas de contraintes imposées par rapport aux performances suggérées, mais les chercheurs ont le sentiment qu'ils risquent de se mettre « hors course » s'ils ne s'y soumettent pas.

⁶ Coser définit les institutions sociales « voraces » comme des institutions « qui réclament l'adhésion complète de leurs membres et qui s'efforcent d'absorber l'intégralité de leur personne » (Coser L.A., 1974. *Greedy institutions. Patterns of undivided commitment*, New York, Free Press).

⁷ Le rapport au salaire est plutôt positif, malgré certaines critiques portant sur des comparaisons avec des disciplines équivalentes dans le secteur privé. Ce n'est toutefois pas un point de préoccupation majeure.

Cette tension a été relayée par le collectif de l'Atelier des Chercheurs⁸. D'après eux, la quête de l'excellence a conduit à l'élaboration de dispositifs d'évaluation reposant sur des critères flous et produisant des effets pervers. Ils poussent les scientifiques à publier « beaucoup, vite, loin, mal », ce qu'ils ont appelé ironiquement la méthode BVLM. Le collectif a constaté aussi le statut conflictuel et la précarisation des conditions de travail, voire l'éviction finale du monde scientifique. De plus, la fatigue et la perte de sens induites par cette méthode produisent de la souffrance au travail. D'après le collectif, la méthode BVLM, en quantifiant la recherche, fait du mal aux esprits et aux corps et pourrait engendrer un enjeu de santé publique.

Quid du côté des évaluateurs ?

Trois représentants d'institutions ont été invités à prendre la parole et commenter les exposés de la première journée : Jean-Pierre Bourguignon, président du Conseil européen de la recherche (*European Research Council*, ERC), Véronique Halloin, secrétaire générale du FNRS (Fonds de la recherche scientifique belge) et Serge Schiffmann, vice-recteur à la recherche et au développement de l'ULB.

Les points discutés lors de cette table ronde ont été : la diversité de la recherche d'une discipline à l'autre (sur le plan du nombre de personnes impliquées, de la temporalité, des besoins financiers, par exemple) ; la question du sous-financement, qui induit deux problèmes majeurs, des dossiers excellents qui ne peuvent pas être financés et de jeunes chercheurs évincés ; le caractère endogène de l'usage des indicateurs bibliométriques (non prescrit par les institutions) ; la prédominance de l'expertise qualitative au FNRS ; l'importance de l'accès libre aux données (*open access*) ; l'intérêt des règles de non-cumul des demandes pour éviter la concentration des moyens. Les systèmes d'évaluation, on le rappelle aussi, n'ont fait qu'évoluer avec le temps, notamment dans leurs caractéristiques informatisées et internationales. Il existe plusieurs types d'évaluation selon les recherches et un indicateur peut être adéquat pour un objet, mais pas pour un autre.

L'évaluation de la recherche : quelles alternatives ?

Six mois après la première rencontre s'est tenue une seconde journée consacrée aux propositions et alternatives en matière d'évaluation de la recherche. Isabelle

⁸ Formé en 2011, ce collectif de chercheurs de l'ULB vise à mener une réflexion sur les nouvelles formes prises par la recherche et notamment la course à l'« excellence ». Il a créé une Charte de la désexcellence (lac.ulb.ac.be/LAC).

Stengers (chimiste et philosophe des sciences, ULB), après avoir montré le caractère crucial de l'évaluation pour la recherche et son ancrage dans l'histoire et les pratiques de la science moderne, en a pointé les limites et les dérives actuelles.

Pour commencer, le modèle de validation de la science expérimentale a été généralisé à toutes les disciplines, alors que la notion même de « fait » sur lequel il repose est problématique. Toutes les disciplines ne partagent pas nécessairement des valeurs communes. Par exemple, l'économie classique et l'économie marxiste ne peuvent pas s'évaluer mutuellement sur la base de critères communs car elles sont en conflit de fond sur la doctrine. En conséquence, toutes les disciplines ne peuvent pas internationaliser leurs publications à partir du moment où seuls certains critères dominants sont retenus. Cette crise est aujourd'hui accentuée par l'homogénéisation de l'évaluation issue de modèles économiques qu'on lui applique.

Face à la crise de l'évaluation, I. Stengers a insisté sur la régénération de formes d'intelligences collectives que tend précisément à détruire l'économie de la connaissance actuelle en valorisant la compétition et les critères du management au sein même de la recherche. Désormais, il faut faire preuve de flexibilité et d'opportunisme par rapport au marché, quitte à créer ce marché (par exemple avec les *rankings* des universités). Quelles sont alors les alternatives ? Partant du constat qu'à l'intérieur du savoir, il y a différentes façons de procéder, une première idée consisterait à évaluer chaque projet, non selon une vision universaliste, mais selon des critères de pertinence propres à chaque discipline. Chacune doit pouvoir imaginer les critères qui lui importent pour bien évaluer ses recherches.

Les scientifiques et les universités doivent aussi pouvoir évaluer eux-mêmes les effets indésirables de leurs travaux. On ne peut se contenter de renvoyer ces questions au verdict de « la société » qui tranchera, comme si toute recherche était un but suffisant en soi. Malheureusement, les scientifiques qui veulent se montrer critiques vis-à-vis de certains travaux sont trop souvent accusés de politiser la science, de sortir de leur domaine de compétence. Or, le manque de prise de responsabilité de la part des sciences nous a justement conduits à des perspectives d'avenir catastrophiques liées à la crise environnementale. Ainsi, nous devons revaloriser des publications et des projets « situés » pour favoriser l'agir.

Concrètement, on peut imaginer des conventions citoyennes informées qui donneraient un avis sur des projets à soutenir ou non : cela nécessiterait que les scientifiques s'adaptent à d'autres critères d'évaluation. La place des amateurs a aussi été négligée dans la recherche. On oublie que l'université produit certes quelques experts, mais surtout beaucoup d'amateurs qui sont à

même d'émettre des jugements selon leur recul et leur formation, mais aussi en tant que citoyens.

La valorisation doit être différente selon les recherches ; un projet d'innovation (qu'il soit scientifique, social, économique) doit faire appel à des contre-expertises chargées d'évaluer les effets collatéraux en dehors de l'université. Dans le même ordre d'idée, on doit se demander à qui profitent les découvertes, notamment lorsque des applications sortent du laboratoire. Il faut alors prendre en compte tous les acteurs pour qui le projet peut être un bien ou un mal. Nous devons, en tant que chercheurs responsables, savoir aussi envisager les perspectives des recherches en aval.

Pour Maryse Salles (informaticienne, Université de Toulouse) et Gabriel Colletis (économiste, Université de Toulouse), qui ont basé leur réflexion sur un rapport de l'Association française d'économie politique (AFEP⁹), les acteurs de la recherche sont encadrés par des normes (indicateurs, procédures, etc.) qui se présentent comme simplement neutres et/ou techniques. Or leur double hypothèse de travail est que (1) les normes ne sont jamais neutres (mais socialement contextualisées et véhiculant des représentations) et que (2) les principes constituent un niveau intermédiaire entre les représentations (ou *doxai*) et les normes concrètes. Ces principes sont des technologies gouvernementales au sens de Foucault ou des technologies de gestion. Celles-ci permettent de structurer le réel en réduisant la complexité, notamment en induisant de la régulation sociale, des choix ou des automatismes de décision. Le niveau des principes est donc une charnière essentielle pour rendre opérationnelle une *doxa* à travers des normes¹⁰.

Quels sont les éléments pour une pratique alternative de l'évaluation ? Elle passe également par les trois niveaux (*doxai*, principes, normes). Les *doxai* devraient aider à soutenir des finalités de la recherche qui sont utiles à la société, avec une attention particulière aux dimensions émergentes, et qui reconnaissent la richesse de sa diversité. Les finalités de l'évaluation devraient, quant à elles, garantir certaines valeurs (de la démocratie, de l'interdisciplinarité, etc.), identifier des savoirs et compétences pour les promouvoir, conseiller et non sanctionner,

⁹ Colletis G., Lamarche T. (coord.), 2011. Pour une nouvelle démarche d'évaluation des laboratoires de recherche, des enseignants-chercheurs et des chercheurs en économie, *Association française d'économie politique*, http://assoekonomiepolitique.org/pour_une_nouvelle_demarche_devaluation_des_laboratoires_de_recherche_des_enseignants-chercheurs_et_des_rechercheurs_en_economie/.

¹⁰ L'évaluation de la recherche est aujourd'hui soumise à une *doxa* dominante issue du monde de l'entreprise : un modèle prédéfini qui tend vers l'homogénéisation, tout peut être quantifié et comparé, la compétition entre chercheurs est considérée comme un modèle par défaut...

favoriser les synergies et non la compétition, faire progresser la professionnalisation des acteurs de la recherche. L'objectif est de recontextualiser la science et l'évaluation dans le temps et l'espace. L'évaluation doit être considérée non pas comme un simple outil managérial, mais comme politisée à travers des finalités de projets de société réfléchis. De même, elle ne doit pas être strictement quantitative. Les objets à évaluer, tels que les laboratoires, ne sont pas de simples unités de production, mais des systèmes multidimensionnels dans lesquels s'inscrivent des chercheurs avec une vision collective qui ne se réduit pas à leurs intérêts personnels.

Les changements escomptés passent par l'inclusion de tous les acteurs concernés pour modifier les principes actuels de l'évaluation (car s'attaquer aux *doxai* reste souvent rhétorique et s'attaquer à une norme est trop local puisqu'il s'agit seulement d'une application particulière de principes structurels dont elle dépend). Parmi les suggestions : revenir à un langage de la recherche, élargir les conventions qui déterminent ce qu'est une production scientifique valable, rendre plus souple, explicite et publique la quantification. Les évaluateurs devraient être élus par leurs pairs. Le processus d'évaluation doit être démocratique, transparent, inclure la lecture des travaux, un droit de réponse et une attention aux enjeux éthiques de la recherche. En outre, le rapport AFEP conseille de tempérer les indicateurs bibliométriques et de prendre en compte les groupes en difficulté ou débutants et les projets innovants, aux retombées méthodologiques ou fonctionnant en réseau. Les systèmes d'information comme les bases de données doivent inclure ces évolutions souhaitables dans les paramètres de l'évaluation. Il en va de même pour les normes : des listes pourraient reprendre plus exhaustivement les types d'activités valorisables dans le cas de la recherche en les pondérant (un article publié aura plus de poids que la participation à un comité éditorial) ou encore des normes établissant les règles de la transparence ou du choix des évaluateurs.

Rejoignant les autres conférenciers sur l'idée que les critères de l'évaluation de la recherche doivent être plus clairs, notamment en raison de la dérégulation actuelle du métier de chercheur, Pierre Marage (physicien, ULB) a souligné d'autres enjeux. Comme alternative plus globale, il suggère d'évaluer les dossiers en posant un maximum de questions qualitatives et de produire un avis lui-même qualitatif. Le fonctionnement des commissions doit reposer sur un dialogue critique et constructif, et pas seulement sur les comparaisons de données quantitatives. Des entretiens pourraient, par exemple, permettre de prendre en compte des qualités personnelles invisibles sur papier.

Un problème plus spécifique, mais majeur, est lié à la question du genre au sein de la recherche. Il faut d'ailleurs veiller à la résorption des inégalités dont sont victimes les femmes, d'une part, par un travail de statistique

et d'évaluation des effets de genre, et d'autre part, par des formations qui sensibilisent, notamment les commissions, aux effets insidieux et peu décelables du problème. Des quotas dans ces commissions et dans les classements peuvent être utiles.

En ce qui concerne le retour des évaluations, on pourrait améliorer la diffusion des rapports des experts, voire des commissions (même si c'est plus difficile dans le cas de concours ou de classement où les experts doivent pouvoir s'exprimer anonymement). Enfin, gardons à l'esprit qu'à l'université, l'évaluation du travail des chercheurs représente aussi une dimension capitale pour les aider à remettre en question leurs idées mais aussi à mieux orienter leur choix d'objectifs et de stratégies professionnelles.

Conclusion

L'évaluation a toujours été associée à la recherche scientifique. Cependant, entre, d'une part, une logique standardisante, parfois dangereuse et inadaptée à certains champs disciplinaires tout comme à l'interdisciplinarité, et, d'autre part, un manque de financement susceptible de gêner la liberté et la sécurité du chercheur, il est évident que l'évaluation actuelle de la recherche produit des effets non désirés et qu'il y a lieu de réfléchir à des pratiques plus adaptées.

Le colloque a permis d'ébaucher plusieurs pistes intéressantes, en réponse aux critiques mentionnées. Les participants réunis autour de quatre ateliers thématiques¹¹ ont également fourni des propositions concrètes pour réorienter l'évaluation de la recherche.

Pour éviter les pièges, les processus d'évaluation gagneraient à affiner les principes sur lesquels ils reposent : être pluralistes, démocratiques, transparents et réellement contradictoires¹² (cf. le rapport de l'AFEP). Les commissions disciplinaires devraient pouvoir choisir

les informations à demander pour construire leur évaluation sur des indicateurs appropriés, non standardisés et couplés à une méthode qualitative¹³. Cela suppose la lecture des travaux dont on pourrait évaluer l'originalité et la profondeur par des questions. L'importance de certaines productions scientifiques devrait avoir plus de poids que le nombre de citations, et lorsque les enjeux sont importants, le chercheur devrait bénéficier d'un droit de réponse. Des critères comme l'âge, le caractère inédit ou atypique de travaux, la prise de risque... doivent pouvoir être considérés et si besoin valorisés. Les critères d'utilité sociale, les tâches d'enseignement et de vulgarisation, l'utilisation de résultats dans une perspective sociétale pourraient être mieux reconnus dans les missions de la recherche. À cette fin, on pourrait limiter la concentration des moyens pour favoriser la diversité et la survie des petits projets : développer davantage la variété des instruments de financement et éviter les cumuls. De même, les projets interdisciplinaires doivent pouvoir être évalués en vertu de critères spécifiques, et par des experts présentant eux-mêmes un profil interdisciplinaire. Les résultats de recherche peuvent être étendus et soumis à d'autres mondes que le seul monde académique, et valorisés dans de nouveaux contextes, tout comme le métier de chercheur et spécifiquement le titre de docteur. En outre, l'évaluation devrait aussi être possible comme audit, en dehors de la pression de la sélection, pour aider à améliorer des projets.

Enfin, la frénésie des critères dominants, comme la publication à outrance dans des revues réputées, pourrait être confrontée à la mise en place de mouvements alternatifs, tels que la *slow science*. Celle-ci, qui propose de prendre le temps d'observer et d'apprécier l'environnement, de remettre en question la science et son rôle dans la société, d'intégrer des acteurs traditionnellement écartés comme les amateurs, conduit au fond à entamer un dialogue contradictoire entre les disciplines et avec l'espace public. En cela, elle fait peut-être descendre les sciences, et leurs évaluations, de leurs tours d'ivoire...

¹¹ « Quel usage faire des indicateurs ? » ; « Quel dialogue entre évaluateurs et évalués ? » ; « Quelle place pour les évalués (projets ou personnes) en situation défavorable ? » ; « Comment évaluer l'interdisciplinarité ? ».

¹² Concrètement, un dialogue préalable doit pouvoir avoir lieu pour informer et échanger au sujet des critères, règles éthiques et objectifs de l'évaluation. De plus, les rôles des membres et experts des commissions devraient être clarifiés et le chercheur rendu capable de se prononcer sur sa stratégie.

¹³ Les indicateurs bibliométriques pourraient ainsi être retenus au cas par cas. Au-delà d'un seuil minimum de production fixé et communiqué préalablement par les commissions, seule une évaluation qualitative devrait être appliquée.