

L'Évaluation en sciences humaines et sociales : Comment mesurer ce qui compte ?

Marie-Claude Maurel (Directrice d'étude à l'EHESS)

*« Mesurer ce qui compte plutôt que compter ce qui peut être facilement mesuré »
(Rapport AUBR, Mackiewicz, 2009)¹*

Résumé

De nouvelles procédures d'évaluation ajustables aux SHS ont émergé ces dernières années. Les principes et les objectifs de l'évaluation doivent être clairement identifiés si l'on veut asseoir la légitimité de la démarche. Que veut-on évaluer ? Qui veut-on évaluer ? Qui est en mesure d'évaluer ? Autant de questions qui invitent à relier les dispositifs et les procédures afin de maîtriser « l'inflation évaluative ».

Comment mesurer ce qui compte, tel est l'enjeu de l'évaluation, que celle-ci s'applique à l'activité scientifique du chercheur individuel ou à celle du collectif de recherche. Cette contribution a pour objet de développer quelques éléments de réflexion sur des procédures d'évaluation en constante expansion dans le système d'enseignement supérieur et de recherche. Elle se veut ni un plaidoyer pour les outils d'évaluation existants ni une défense des indicateurs bibliométriques. L'intention est tout à fait autre. Dans l'esprit qui a animé les travaux du Conseil pour le développement des humanités et des sciences sociales (CDHSS)², l'accent est mis sur les finalités et les modalités de l'évaluation, sur le « bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs », pour reprendre le titre du rapport de l'Académie des sciences et la proposition qu'il contient³.

A quelles fins évaluer ?

Une exigence de politique publique

L'évaluation est devenue une exigence permettant de mesurer les effets d'une politique publique afin de mettre les moyens consentis en regard des attentes de la société. Le financement de la recherche a besoin d'être fondé sur des systèmes d'évaluation et des indicateurs de performance de manière à justifier la meilleure allocation possible des ressources. S'agissant de l'activité de production scientifique, l'amélioration de la performance répond à un enjeu d'importance stratégique pour le développement économique social et culturel. La qualité du système d'enseignement supérieur et de recherche est en effet reconnue comme un facteur déterminant de la compétitivité globale, au plan

¹ Citation extraite du rapport « *Assessing Europe's University Based-Research* » préparé par le groupe d'experts nommé par la DG recherche de la Commission européenne, 2009.

² Pour des sciences humaines et sociales au cœur du système d'enseignement supérieur et de recherche, Rapport remis à Valérie Pécresse, Ministre de l'Enseignement supérieur et de la Recherche, octobre 2010. Voir en particulier le chapitre 5, L'évaluation en sciences humaines et sociales (77-94).

³ Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs, 17 janvier 2011, <http://www.academie-sciences.fr/activite/rapport/avis170111.pdf>

international. C'est dans ce contexte que les universités, les organismes de recherche et leurs personnels sont soumis à une évaluation multidimensionnelle (multicritères) impliquant la prise en considération de l'ensemble des missions qui leur sont confiées. L'objectif est de promouvoir la qualité de l'activité des enseignants-chercheurs et des chercheurs et de baser sa reconnaissance sur des principes qui s'imposent à tous. La sélection de l'excellence individuelle et/ou collective constitue un volet spécifique de cette politique qui participe à l'objectif d'ensemble. Qui fait ou qui produit de la recherche en SHS en France (les « producteurs » au sens défini par l'AERES), pour quels résultats et à quelles conditions ?

Les responsables de la politique d'enseignement supérieur et de recherche n'accordent-ils pas une confiance imméritée à cette évaluation nouvelle manière, qui trouve sa source d'inspiration dans le « *management* par la qualité » ? L'adoption de nouveaux outils et de nouvelles méthodes de l'évaluation ne risque-t-elle pas de se combiner à un esprit de système – qui finirait au demeurant par rendre l'évaluation inopérante ?

Une fonction inhérente à l'activité scientifique

L'acte d'évaluer est une réalité ancienne. En tant que fonction inhérente à l'activité scientifique, il participe à la validation des savoirs, en ce qu'il contribue à la définition des critères relatifs aux formes et au degré de véracité de ceux-ci. L'acte d'évaluer se réfère à des principes d'action et/ou à des objectifs qui doivent être énoncés à l'avance, et par rapport auxquels sont mesurés les écarts de réalisation des actions. Le respect de ces principes permet de fonder la légitimité de l'acte d'évaluer. Les objectifs de l'évaluation doivent être définis et appréciés en relation avec de multiples facteurs. Les dispositifs d'évaluation sont à mettre en cohérence avec les objectifs qui ont été assignés aux chercheurs, tels que produire des connaissances nouvelles, développer des pratiques de recherche interdisciplinaires et intégrées, etc. Enfin, il s'agit de relier les dispositifs et les procédures mises en œuvre aux objectifs et de manière plus générale aux attentes.

Comment évaluer ?

Dans ce qui suit, on gardera présent à l'esprit les raisons pour lesquelles les activités de recherche et de formation à la recherche sont soumises à évaluation, pour cibler la réflexion sur les procédures adoptées pour ce faire. C'est la question du « comment évaluer ». Cette question se pose avec d'autant plus d'acuité lorsque sont appliquées de nouvelles modalités d'évaluation et introduits des outils d'évaluation rénovés. Plusieurs risques sont mis en avant : la lourdeur du processus et son caractère répétitif ; la crainte qu'il ne soit pas conduit de manière impartiale ; le danger de déséquilibrer les procédures de recherche, un point particulièrement sensible en sciences humaines et sociales.

Pour comprendre ces craintes qui affectent la communauté universitaire, il convient de poser la question de l'évaluation dans sa globalité, c'est-à-dire en prenant en compte non seulement le statut, la composition, les missions et les modalités de fonctionnement des instances d'évaluation, en particulier en sciences humaines et sociales, mais également les pratiques et les critères propres aux divers champs disciplinaires⁴.

⁴ L'analyse des craintes des enseignants-chercheurs concernant les procédures nouvelles d'évaluation ont été recensées dans le rapport ci-dessus mentionné, p.80.

A quelles conditions l'évaluation peut-elle être adéquate et efficace ?

Le respect de quelques principes devrait permettre de mieux répondre aux critères d'adéquation et d'efficacité. Ces principes consistent à :

- impliquer les chercheurs et les enseignants-chercheurs dans la préparation des modalités d'évaluation (par un véritable processus de négociation sociale)
- exiger la transparence et l'objectivité à la fois des procédures d'évaluation et des résultats ;
- assurer la définition explicite des indicateurs et des critères et introduire une pondération de ceux-ci ;
- garantir une stricte séparation des pouvoirs entre évaluateurs et décideurs ;
- assurer la pertinence et l'efficacité du processus.

Comment ces principes sont-ils mis en œuvre ?

Les régimes d'évaluation⁵

Au cours de la décennie écoulée de nouveaux dispositifs et de nouvelles procédures ont été mis en place dont les effets sur l'activité pédagogique et scientifique des enseignants-chercheurs ne sont pas sans incidence. Des analyses récentes parlent de « régimes d'évaluation » et distinguent une évaluation collégiale qui est le fait de jurys, de commissions, d'une part, et un nouveau régime qui a recours à des experts spécialisés dans l'évaluation, d'autre part. Les règles du premier régime, de tradition ancienne, sont bien connues des chercheurs qui soumettent leurs travaux à l'appréciation de jurys. Ce mode d'évaluation requiert la construction endogène de référents communs définis par les pairs siégeant dans l'instance d'évaluation (sections du CNU, du Comité national de la recherche scientifique, etc.). Le second mode, d'apparition plus récente, se présente comme un régime d'objectivité, fondé sur des outils d'évaluation (grilles, indicateurs bibliométriques, système de notation). L'exemple donné à l'appui de ce nouveau régime est l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES). Créée en 2007, l'AERES est un dispositif transversal d'évaluation prenant appui sur un corpus d'outils standardisés applicable à tous les secteurs de la recherche. L'agence évalue les unités de recherche, les formations et les établissements⁶. Si les principes énoncés par l'AERES et qui ont trait à l'impartialité, la fiabilité, la transparence des évaluations représentent une avancée, il est clair que chacun de ces régimes a ses qualités et ses défauts (ou plutôt ses limites).

Ceux qui caractérisent l'évaluation par les pairs sont souvent dénoncés. Si les pairs sont théoriquement bien placés pour apprécier l'originalité de la pensée scientifique, le caractère innovant du projet, le rayonnement national et international, l'impartialité de leur jugement peut être compromise par la fréquence excessive de ce type d'exercice, la subjectivité, les conflits d'intérêt ou encore les effets de groupe. Sous l'influence de quelques leaders certains champs scientifiques peuvent se réguler quasiment en vase clos.

L'évaluation fondée sur l'outil bibliométrie a aussi des limites. Les écueils procèdent du choix opportuniste des références, et du support de publication. Les contextes disciplinaires et les usages les spécifiant sont bien différenciés. Enfin, le choix des indicateurs et celui des bases de données

⁵ On doit le développement de cette notion à Nicolas Dodier, « Penser un régime d'évaluation de la recherche scientifique », 2009, <http://evaluation.hypotheses.org/452>

⁶ L'AERES met en place des comités de visite ad hoc des unités de recherche qui se livrent à un examen des activités à partir de grilles, de vérification d'indices, de critères de bonne gouvernance, etc.

pour les construire constituent un biais non négligeable. Le bon usage des indicateurs a besoin d'être défini, encadré et respecté.

Les indicateurs issus de la bibliométrie

Sur le plan méthodologique des avancées importantes ont été réalisées pour utiliser voire élaborer des indicateurs bibliométriques. Elles sont le fait de diverses instances nationales (comité national, AERES, ANR, etc.), européennes (ESF, ERC, etc.) et internationales (par exemple aux Etats-Unis, en Australie).

Beaucoup reste à faire pour améliorer les bases de données existantes, notamment afin qu'elles prennent mieux en compte les travaux en sciences humaines et sociales. A cet égard, il conviendrait de prendre en considération le spectre complet des recherches (fondamentales et appliquées), ainsi que les contextes régionaux, nationaux, et les publications dans les langues autres que l'anglais. La bibliométrie internationale reste fortement dépendante des indices de citations existants (Web of science, (Thomson Reuters), Scopus (Elsevier), Google Scholar, le logiciel Publish or Perish d'Ann-Wil Harzing qui fonctionne à partir de Google Scholar.

Que mesurent les indicateurs ? Peuvent-ils être considérés comme des guides de la qualité des travaux de recherche ? Peuvent-ils être considérés comme robustes ? On observera que le nombre de citations dont bénéficie un article ne garantit pas *ipso facto* sa qualité. Il peut être lié à un effet de notoriété ou de mode concernant un objet d'étude. Il dépend aussi de la taille de la communauté scientifique concernée.

Les outils connexes du type classement des revues représentent une autre source de catégorisation des publications. Les caractéristiques des diverses listes de classement font l'objet d'autres contributions dans le cadre des Actes de ce colloque, auxquelles le lecteur se rapportera. Bornons-nous ici à proposer de pondérer la logique bibliographique par une logique scientifique, à conseiller de ne pas surestimer les textes banals publiés dans de bonnes revues aux dépens de textes intéressants publiés dans des supports moins conformes. La qualité plus scientifique d'un article n'est pas fonction de la qualité du support et un *impact factor* élevé d'une revue ne devrait pas avoir d'effet sur tel ou tel article.

Quels usages fait-on de ces outils bibliométriques ? Remarquons qu'ils sont habituellement utilisés pour établir une catégorisation hiérarchisée de la qualité scientifique, mais aussi comme outil de calcul en vue d'établir des classements, enfin comme instrument de comptage, par exemple pour établir le taux des « producteurs » d'une institution.

Quand et comment utiliser la bibliométrie ? L'outil peut être recommandé pour évaluer l'impact de la production scientifique d'un chercheur confirmé (un senior ayant une relativement longue carrière), mais il n'est pas approprié dans le cas d'un chercheur junior. L'outil bibliométrique peut être utile pour effectuer un tour d'horizon de candidats à l'échelle d'un jury mono disciplinaire mais dans le cas d'un jury pluridisciplinaire il ne doit l'être que si les experts sont avertis des différences entre les disciplines. Dans tous les cas de figure, la règle doit être de ne comparer que ce qui est comparable (le chercheur doit pouvoir être positionné par rapport à la distribution statistique des chercheurs de sa communauté).

Enfin, l'évaluation bibliométrique ne doit pas conduire à négliger l'originalité et l'innovation qui restent plus difficilement mesurables.

Comment évaluer l'excellence ?

Élément de labellisation ou de reconnaissance, l'excellence scientifique ne se décrète pas, elle doit être identifiée. La question de l'excellence académique s'inscrit dans la problématique plus générale de l'évaluation : comment juger de l'intérêt et de la qualité d'une recherche scientifique et selon quels critères en rendre compte⁷ ? Définie comme une appréciation de la qualité de la recherche, l'excellence scientifique ne peut être tenue pour une valeur absolue en soi mais comme une notion relative et donc l'expression d'un classement (*ranking*). Pour revêtir du sens, le classement des « meilleurs » (chercheurs ou projets) doit être établi à l'issue d'une procédure d'évaluation dont les fondements méthodologiques et les règles sont fixés à l'avance, et respectés par les experts. La reconnaissance de l'excellence académique est l'aboutissement d'une compétition inscrite dans un cadre normatif. Les performances d'un chercheur ou d'une équipe ne peuvent être mesurées exclusivement à partir d'indicateurs bibliométriques, même si ceux-ci peuvent apporter des éléments d'appréciation. L'identification de l'excellence relève d'une construction nécessairement collective et relationnelle dans le cadre d'un mode d'évaluation par les pairs.

Pour illustrer notre propos, la procédure de sélection-évaluation de l'IUF qui prend appui sur une vingtaine d'années d'expérience et autant de promotions de lauréats peut être analysée. La mission d'évaluer, de comparer et de traduire en classement les candidats est confiée à deux jurys (junior et senior) indépendants et souverains, nommés par arrêté ministériel sur proposition d'une commission de désignation. Les lauréats sont sélectionnés au terme d'une évaluation exigeante par ces jurys qui comportent au moins deux cinquièmes d'universitaires étrangers. La pertinence et la faisabilité du projet présenté par le candidat interviennent dans l'appréciation émise par les jurys selon des critères comparables à ceux utilisés par les instances d'évaluation telles que les comités de l'ERC. Les dossiers des candidats sont évalués par deux rapporteurs, spécialistes de la discipline, les avis font ensuite l'objet d'une discussion en sous-jurys et d'une délibération en séance plénière pour établir le classement final. Les jurys sont pluridisciplinaires, il n'y a pas de fonctionnement par sous-groupe étroitement disciplinaire, mais seulement deux sous-jurys (sciences de la matière, de la nature et de la vie, d'une part, sciences humaines et sociales, d'autre part). La difficulté réside dans la diversité des modes d'évaluation de la production scientifique selon les disciplines. La grille d'évaluation, inspirée de celle de l'ERC, est commune mais il est recommandé aux jurés de tenir compte des standards de la discipline. Il est très ambitieux de vouloir juger de l'excellence indépendamment des cadres disciplinaires. Le croisement des regards disciplinaires pratiqué au sein des jurys présente l'avantage de permettre de mieux évaluer l'intérêt des recherches interdisciplinaires et de celles qui ouvrent de nouveaux fronts de la connaissance. Un tel mode de sélection repose d'abord sur la qualité des pairs formant le jury. Le choix se fonde sur un certain nombre de critères : parcours académique de haut niveau, indépendance d'esprit, aptitude à évaluer l'originalité et l'innovation, sens de l'équité... Le bon fonctionnement du jury requiert de chacun de ses membres le respect des règles informelles d'évaluation, de la souveraineté de chaque discipline, ainsi que d'une capacité à délibérer.

⁷ Voir l'ouvrage de MICHÈLE LAMONT, *How Professors Think: Inside the Curious World of Academic Judgment*, Harvard University Press, 2009, ainsi que les actes du Symposium/Thinking Academic Evaluation after Michèle Lamont's *How Professors Think* - Edited by Anna Carola Freschi and Marco Santoro, *SOCIOLOGICA*, 3/2010

Conclusion : Comment améliorer l'évaluation de la recherche en sciences humaines et sociales ?

Soulignons la complexité du processus d'évaluation, sa lourdeur (la fréquence et le caractère répétitif des évaluations). L'évaluation se combine, en France, à une rationalité bureaucratique centralisée risquant de donner à l'évaluation une emprise excessive qui finira d'ailleurs par la vider de son contenu. Il n'est pas de problème qui échappe à l'évaluation et l'autonomie des universités peut renforcer ce mécanisme par le développement d'évaluations internes à chacune d'elles : comités d'évaluation pour les promotions, comités « qualités » qui se mettent en place, audits internes formés par les universités et par les PRES, appels d'offre des investissements d'avenir.

L'adhésion de principe à l'évaluation ne devrait pas conduire à ignorer cette inflation évaluative dont on sait qu'elle conduira nécessairement à une routinisation des procédures et à une augmentation continue des « coûts de transaction ». On risque de répondre à chaque problème par une nouvelle évaluation.

L'agence d'évaluation repose sur le principe d'une séparation du jugement et de la décision, de l'expertise et de la politique. C'est excellent. Mais dans ce cas, il faut souhaiter que les décideurs ne se défaussent pas en créant de nouvelles évaluations en aval ou de manière parallèle. Que les experts expertisent et que les décideurs décident sans se cacher derrière de nouvelles évaluations par les universités, les PRES, les régions, les comités de programmes... Réfléchissons sur la dérive évaluative, non pas pour nous opposer à l'évaluation, mais pour en défendre le principe et l'efficacité.

Pour ce qui est de l'évaluation elle-même, le grand risque reste celui d'un usage mécanique et non raisonné de la bibliométrie : trop serrée, elle exclut, trop large, elle se vide de sens, et surtout, elle formate les modes d'expression des chercheurs les plus jeunes. Les travaux de longue durée (l'écriture d'un livre), la réflexion conceptuelle, l'imagination créatrice, sont autant de risques que les jeunes chercheurs sont réticents à courir en raison des pratiques de l'évaluation. Au final, il convient de plaider en faveur d'un mode d'évaluation hybride intégrant les bénéfices du jugement par les pairs et l'usage réfléchi et critique des outils d'évaluation permettant d'établir des notations objectives.