

Evaluer la sociologie : Vers de nouvelles normes d'évaluation scientifique ?

Par Daniel Benamouzig et Giovanni Prête

Compte rendu de la demi-journée du 20 octobre 2008, organisée par le **Réseau thématique pluridisciplinaire (RTP) « Société en évolution, science sociale en mouvement »** (CNRS, Ministère de l'Enseignement et de la Recherche) et l'Association des Enseignements du Supérieur (ASES).

Le 20 octobre dernier, le Réseau Thématique Pluridisciplinaire (RTP) « Société en évolution, science sociale en mouvement » (CNRS, Ministère de l'Enseignement Supérieur et de la Recherche)¹ et l'Association des Sociologues de l'Enseignement Supérieur (ASES)² organisaient un séminaire commun intitulé « Evaluer la sociologie : Vers de nouvelles normes d'évaluation scientifique ? »³. Alors que, le même jour, plusieurs organisations syndicales appelaient à manifester devant le CNRS pour protester contre la politique de réforme du dispositif de recherche et d'enseignement supérieur engagée par le gouvernement, ce sont une cinquantaine de personnes, chercheurs et enseignant-chercheurs majoritairement, qui se sont réunis pour réfléchir aux conditions d'évaluation scientifique en sociologie. La demi-journée était articulée autour de deux tables rondes, respectivement consacrées aux usages des outils de mesure en matière d'évaluation et à la place des nouveaux dispositifs institutionnels. Les deux sessions étaient animées successivement par Claire Lemerrier (CNRS IHMC) et Philippe Bézès (CNRS CERSA), membres du bureau du RTP.

L'évaluation de la recherche, une nouveauté ?

Une grande partie des discussions de la journée a porté sur des évolutions très actuelles : controverses autour de la création de listes hiérarchisées de revues, mise en place de nouvelles instances d'évaluation (AERES, ANR). Cependant, la première intervention de la journée par Emmanuelle Picard (INRP-ENS) a souligné l'importance de réinscrire les évolutions à l'œuvre dans le temps long. L'historienne de l'éducation a souligné que des systèmes d'évaluation existaient depuis la fin du 19^{ème} siècle (sous l'appellation, alors, de certification professionnelle) au moment où les universitaires ont réussi, dans le système universitaire très centralisé mis en place sous l'Empire, à s'imposer comme les gestionnaires de leur corps. Cette imposition s'est traduite par un fort découpage disciplinaire des pratiques d'évaluation. Ce découpage a non seulement rendu difficile l'organisation d'activités pluridisciplinaires mais a aussi favorisé l'absence d'explicitation des critères

¹ <http://www.cnrs.fr/shs/recherche/RTP-societes-evolution/presentation.htm>

² <http://www.ases-asso.com>

³ <http://www.cnrs.fr/shs/recherche/RTP-societes-evolution/fichiers/Programme20octobre2008.doc>

d'évaluation. Dans cette perspective, on peut analyser les tensions actuelles comme l'opposition entre deux logiques. La première valorise une autogestion par les corps tandis que la seconde pousse au développement d'une évaluation que l'on pourrait qualifier d'explicite, voire de mécanique. Comme l'ont rappelé Didier Torny (INRA RITME) et Christine Musselin (CNRS CSO), une approche historique permet d'introduire un regard critique sur un discours gouvernemental présentant les évolutions en cours comme radicalement nouvelles : l'évaluation des unités, le financement de la recherche par projets se sont développés dès les années 1980 en France. L'explicitation formelle des critères d'évaluation était un aspect important des réformes induites par la loi d'orientation sur la recherche de 1983.

L'idée d'une radicale nouveauté est apparue encore plus questionnable au regard des dynamiques d'appropriation des instruments et de mise en œuvre des évolutions institutionnelles de l'évaluation. L'intervention de Lynda Hantrais (Loughborough University) sur les réformes institutionnelles au Royaume-Uni depuis les années 1980 a largement souligné l'importance de ces processus. Le dispositif britannique diffère sur de nombreux points du dispositif français : la majorité de la recherche a lieu dans les universités, les enseignants chercheurs ne sont pas fonctionnaires, les universités sont autonomes en ce qui concerne le recrutement et la gestion de leur personnel, la recherche est financée par deux principales agences de moyens (HEFC ; ESCR). A partir des années 1980, le financement de la recherche s'est organisé autour d'un processus d'évaluation pluriannuel des départements académiques et des chercheurs. D'une périodicité variable, comprise entre quatre et sept ans, le *Research assessment exercise* (RAE) porte sur l'ensemble du dispositif de recherche. Il se fonde sur plusieurs critères d'évaluation (publications au sens large, reconnaissance par les pairs, travail d'encadrement et d'animation de la recherche etc.). Individuellement, les chercheurs *font en particulier état de quatre productions, qui ne sont pas nécessairement des articles scientifiques*. Le processus vise à attribuer une note aux départements de recherche sur la base de laquelle sont alloués les financements. Depuis la mise en place du RAE de nombreux rapports ont mis en doute la pertinence des évaluations et ont discuté leurs effets – notamment une augmentation de la concentration des ressources financières sur quelques universités et un cloisonnement des disciplines. Ces rapports ont débouché sur des évolutions du dispositif.

Pierre Lascoumes (CNRS CEVIPOF), à partir de son expérience d'évaluateur au sein de la section 40 du Comité National de la Recherche Scientifique a également insisté sur la nature évolutive des pratiques d'évaluation. Il a décrit comment l'explicitation croissante des critères d'évaluation (par ex. l'obtention de l'HDR pour la promotion au grade de DR), l'augmentation générale des exigences faites aux chercheurs (par ex. une attention à la publication dans une diversité de revues de la discipline) et la formalisation du travail d'évaluation (par ex. la lecture collective des dossiers individuels) constituaient trois dynamiques à l'œuvre depuis plusieurs années dans le fonctionnement du comité du CNRS.

Certains intervenants ont néanmoins insisté sur des éléments nouveaux introduits par les réformes récentes. Jean-Louis Briquet (CNRS CERI), membre de l'AERES, ne nie pas qu'il existe une continuité entre l'agence et des structures antérieures comme la MSTP. Mais il a mis en avant le fait que le principe de la « visite sur site », par exemple, et le principe de la séparation de l'évaluation scientifique et des décisions d'allocation des moyens, surtout, représentaient un changement important, qui posait la question centrale de l'utilisation des évaluations produites. En tant que Président d'Université, Laurent Batch (Paris IX Dauphine) a également souligné que

les réformes engagées par le gouvernement constituaient un changement important, dans la mesure où elles mettaient l'évaluation au centre du dispositif de régulation de la recherche universitaire. Rupture ou continuité, les débats de la journée de réflexion ont traduit une inquiétude par rapport à ces réformes et la manière dont il fallait se positionner à leur égard.

Critiquer du dehors ou transformer du dedans : quel rapport aux dispositifs ?

Les intervenants et participants à la journée se sont largement retrouvés pour critiquer les limites de l'évaluation quantitative de l'activité de recherche. Franck Laloé (CNRS ENS- Laboratoire Kastler Brossel) a, en tant que physicien, souligné que des critiques pouvaient être faites à la fois du point de vue des SHS et des sciences dites « dures ». Il a montré comment différents indicateurs bibliométriques (IF, H etc.) avaient tendance à donner une prime à des stratégies de publication de court terme et à privilégier ce qu'il a appelé des « styles » de recherche, aux dépens d'autres, pourtant également indispensables. De manière plus générale, plusieurs dangers ont été attachés au développement de dispositifs quantitatifs d'évaluation. Michel Grossetti (CNRS LISST) a souligné « le risque de bureaucratisation de l'évaluation », les « dérives du *publish or perish* », qui peuvent favoriser « la multiplication des fraudes, les stratégies de publication circulaires, le découpage en tranche des articles ou encore la standardisation des publications et des recherches ». Il a signalé que l'usage des indicateurs bibliométriques pouvait avoir pour effet de fabriquer artificiellement des hiérarchies très tranchées et artificielles entre les chercheurs. D'autres ont souligné que cet effet de hiérarchisation pouvait aussi concerner les revues, au risque d'en « tuer » un nombre important.

Dans cette perspective, le débat s'est en partie construit autour d'une alternative : Faut-il participer aux instances d'évaluation qui se mettent en place et alimenter les instruments qui les renseignent pour les faire évoluer en les intégrant dans une régulation des communautés scientifiques par les pairs, et ce au risque d'éventuelles dérives ? Ou bien faut-il au contraire les critiquer du « dehors », en considérant qu'ils relèvent d'une logique aboutissant nécessairement à un dessaisissement de la régulation scientifique par les pairs et à sa bureaucratisation ? Les orateurs invités, ayant pour la plupart participé – de manière plus ou moins critique – à la mise en place de ces dispositifs formels, ont généralement défendu la première position. La seconde position a été discutée dans le cadre d'échanges nourris avec la salle.

Laurent Batch, en s'appuyant sur un parallèle avec la notation des écoles de management, a ainsi justifié la participation active aux dispositifs formels de l'évaluation « qui émanent des professions ». Il a clairement défendu l'idée que si l'institution académique actuelle ne s'engageait pas dans l'évaluation, « ce serait le marché », c'est à dire d'autres agences de notation, y compris internationales et peu légitimes d'un point de vue académique, qui se chargeraient du ranking.

De manière moins enthousiaste, Fabien Jobard (CNRS CESDIP), ancien secrétaire scientifique de la section 40 du Comité national du CNRS, a pour sa part défendu le principe d'une publicisation des pratiques d'évaluation. Se référant aux procédures en cours de classement des revues considérées comme importantes dans les communautés scientifiques, il a regretté la logique de classement des revues retenues en rangs A et B. Il a également décrit la manière dont sa participation à l'établissement de listes pour l'AERES visait à contrer un classement effectué au niveau européen sur la base de critères de diffusions des revues, afin que l'on ne s'en remette pas seulement aux « indicateurs les plus violents, comme les indices de

citation ». Plusieurs personnes ont également souligné que l'utilisation de moyens formels d'évaluation, qu'ils soient bibliométriques ou non, pouvaient améliorer le fonctionnement des dispositifs d'évaluation moins formalisés, qui ne sont pas exempts d'imperfections : effets d'interconnaissance, difficulté de dépasser les oppositions paradigmatiques pour faire émerger le consensus, ou encore limites de l'expertise des évaluateurs quand ils sortent du cadre étroit de leur discipline...de nombreux mécanismes que la sociologie mertonienne des sciences a mis en évidence depuis plusieurs décennies.

Le compte rendu de Christine Musselin sur sa participation à l'ANR s'est inscrit dans ce positionnement, valorisant une participation critique aux dispositifs d'évaluation. Etablissant un bilan plutôt positif de l'Agence après ses premières années d'existence, elle a indiqué que – de son point de vue – les comités scientifiques disciplinaires (CSD) qui évaluent les réponses aux appels d'offre font un travail sérieux, dans une certaine transparence et avec un souci d'équilibrage disciplinaire. Elle a cependant énoncé certaines limites du dispositif, telles que le manque de communication avec les porteurs de projet ou la pénurie et l'absentéisme des experts, qui encouragent parfois le recours à des gens disponibles plutôt que compétents. Elle a aussi exprimé ses interrogations sur les conséquences à plus long terme de la mise en place de l'ANR, à la fois en termes de gestion des carrières (des chercheurs confirmés mais également de « l'armée de CDD » qui est en train d'être créée et qui n'est « gérée par personne ») et de contenus scientifiques, l'ANR promouvant un modèle de recherche particulier (projets impliquant plusieurs chercheurs et institutions).

Dans la salle, plusieurs interventions sont allées dans le sens d'une participation critique aux dispositifs d'évaluation formels. Louis-André Vallet (CNRS CREST), notamment, a pris la parole pour affirmer que « sauf à considérer que les SHS ont un statut épistémologique particulier, on ne peut pas les faire échapper à des outils de classement formels », comme des outils perfectibles « utilisés de manière responsable et parmi d'autres ». Parallèlement, d'autres intervenants ont adopté une posture beaucoup plus critique, à l'image de Marie-Anne Dujarier (U. Paris III, CNRS-CNAM), qui s'est interrogée sur la compatibilité entre des instruments d'automatisation du jugement et une activité scientifique reposant sur la controverse. Inquiète du devenir des critères d'évaluation implicites, propres au métier de chercheur, elle a souligné que le système mis en place pouvait, par des effets de hiérarchisation, « supprimer une partie de notre métier ». Elle s'est dite favorable à ne « même pas négocier les critères des dispositifs d'évaluation formalisés ». C'est à une même conclusion qu'est arrivé Jan Spurk (U. Paris V GEPECS), très inquiet d'une tendance à l'uniformisation de la production scientifique, liée à la formalisation de l'évaluation.

Au-delà de ces positionnements alternatifs, plusieurs personnes ont rappelé les limites d'une analyse critique des processus en cours se concentrant sur les dispositifs d'évaluation. Dans la salle, certaines interventions ont mis en avant la nécessité de replacer la discussion sur l'évaluation dans une analyse plus large des évolutions économiques du monde de l'édition scientifique. Huguette Rigot (U. Paris VII CHCSC) a souligné que les « dispositifs sociotechniques » servant à produire des indicateurs de classements bibliométriques avaient pour effet de produire un marché de l'édition, qui, en France, a donné un poids prépondérant et discutable au « bouquet Cairn ». Dans le cadre de cette problématique, Frédéric Neyrat (U. Limoges GRESCO) a souligné qu'il aurait été utile d'organiser une réflexion sur

l'organisation économique de l'édition. Elle aurait selon lui permis de discuter des effets de prime à l'ancienneté induits par la mise en place de listes et de classements de revues. De manière générale, il a été souligné que les dispositifs d'évaluation s'inscrivaient dans un cadre institutionnel et organisationnel plus large dans lequel ils étaient appropriés et que c'est sur l'évolution de ce cadre que l'attention, l'analyse et la critique devaient porter. Fabien Jobard a ainsi souligné que c'est bien dans un système qui permet la communication entre chercheurs, la collégialité des décisions et la continuité des institutions que l'utilisation d'outils formels d'évaluation peut avoir une valeur et une utilité. En affirmant cela, il posait les pistes d'un certain nombre de conditions et d'améliorations possibles à la participation aux dispositifs formels d'évaluation de la recherche.

Propositions pour améliorer les dispositifs d'évaluation de la recherche

Différentes voies d'amélioration des dispositifs formels d'évaluation de la recherche ont été évoquées au cours de la journée. D'une manière générale, les intervenants et les échangent ont discuté les possibilités d'articulation entre trois types d'éléments différents, pouvant se combiner de manière plus ou moins intéressante ou problématique. Ces trois types d'éléments sont respectivement : les outils quantitatifs, notamment bibliométriques ; les procédures formelles d'évaluation ; et le temps du jugement et de la délibération entre pairs. Les déséquilibres entre ces trois éléments peuvent induire des modalités d'évaluation peu acceptables. En tout état de cause, dans l'état actuel des pratiques, leur articulation n'étant pas aboutie.

Plusieurs intervenants ont défendu l'idée d'une multiplication des outils et des critères d'évaluation, afin de ne pas exagérer l'importance de certains d'entre eux. Déjà dans les années 1980, a rappelé Emmanuelle Picard, un groupe de travail d'astrophysiciens avait, dans le cadre d'une réflexion sur les critères d'évaluation de l'ensemble des activités de recherches (trois types d'activités ayant été défini : publications, gestion scientifique et expertises), proposé de prendre en compte 14 critères pour juger de la qualité des publications. Il était ainsi proposé de mesurer la manière dont une publication serait citée au bout de cinq ans, ce qui permettait d'identifier les travaux scientifiques ayant un impact durable. Dans cette perspective d'amélioration des évaluations bibliométriques, Franck Laloé a également proposé d'annuler les effets clusters induits par les indices utilisés en divisant très simplement l'indice par le nombre d'auteurs. Jean-Louis Briquet s'est interrogé sur l'idée de multiplication des indicateurs en rapportant comment un mouvement avait été engagé au sein de l'AERES pour promouvoir l'usage de quatre notes, pour empêcher que les décideurs n'utilisent que la seule note globale attribuée par les comités de restitution. Malheureusement, il a indiqué que cette tentative s'était soldée par la création d'une cinquième note dite « globale » synthétisant les quatre autres, illustrant par là la difficulté d'introduire de la complexité dans des institutions qui cherchent à simplifier les processus de décision.

A plusieurs reprises, l'idée d'une révision régulière des procédures des dispositifs d'évaluation a été discutée. Michel Grossetti a souligné que dans le cadre du groupe de travail sur les listes des revues de l'AERES, il avait demandé avec des collègues à ce que les listes soient périodiquement réévaluées, selon des critères débattus et explicites.

Plusieurs intervenants ont en outre souligné qu'il était important de garantir des moyens financiers capables de garantir le fonctionnement et l'indépendance des

structures d'évaluation. Pierre Mercklé (ENS LSH) s'est ainsi inquiété de l'indépendance d'une agence de moyens comme l'ANR qui, faute de moyens, délègue la gestion administrative des appels d'offre à des établissements d'enseignement supérieurs qui sont ainsi, de fait, mieux informés des opportunités de financement. Par rapport à cette question des moyens, les débats ont laissé apparaître une inquiétude quant aux ressources disponibles pour garantir la qualité et l'indépendance des experts évaluateurs, notamment des rapporteurs qui semblent jouer un rôle particulièrement important que ce soit à l'ANR ou dans les comités du CNRS.

Les propositions d'amélioration qui ont pu être faites au cours des échanges traduisaient de fait le souci de garantir la continuité de l'exercice du jugement au sein des dispositifs d'évaluation. De manière plus ou moins radicale, tous les participants ont souligné combien l'évaluation était un processus complexe qu'il serait aberrant de vouloir mécaniser, ou même formaliser complètement. Cette complexité implique l'aménagement d'un temps spécifiquement consacré à l'étude approfondie, au jugement et à la délibération pluraliste portant sur les résultats des activités scientifiques, entendues au sens large. Sans viser une conclusion définitive à un débat passionné se prolongeant par ailleurs, la journée a donné l'occasion d'échanges de vue informés et responsables, traduisant des inquiétudes et des souhaits témoignant de l'intérêt de moments collectifs et pluralistes de réflexivité des sciences sociales sur leurs propres pratiques.