

Revue européenne
des sciences sociales

European Journal of Social Sciences

Revue européenne des sciences sociales

European Journal of Social Sciences

XLVI-141 | 2008

Processus d'évaluation des sciences sociales :
acteurs et valeurs

Modalités et enjeux de l'évaluation des SHS : le tournant actuel

Jean-Pierre Gaudin



Édition électronique

URL : <http://journals.openedition.org/ress/150>

DOI : 10.4000/ress.150

ISSN : 1663-4446

Éditeur

Librairie Droz

Édition imprimée

Date de publication : 1 mai 2008

Pagination : 63-72

ISBN : 978-2-600-01220-1

ISSN : 0048-8046

Référence électronique

Jean-Pierre Gaudin, « Modalités et enjeux de l'évaluation des SHS : le tournant actuel », *Revue européenne des sciences sociales* [En ligne], XLVI-141 | 2008, mis en ligne le 01 mai 2011, consulté le 05 mai 2019. URL : <http://journals.openedition.org/ress/150> ; DOI : 10.4000/ress.150

Jean-Pierre GAUDIN

MODALITÉS ET ENJEUX DE L'ÉVALUATION DES SHS : LE TOURNANT ACTUEL

INTRODUCTION

Qui donc atteste de la valeur scientifique des énoncés en sciences humaines et sociales, et comment ? A cette question, les réponses apportées se sont profondément modifiées en quelques décennies.

Le « fond de décor » nous est familier et a longtemps été assez stable. A mesure que l'épistémè scientifique s'affirmait dans le monde occidental, des sociétés savantes se sont peu à peu constituées et diversifiées. Ces sociétés ont structuré peu à peu des réseaux de reconnaissance croisée : elles se sont équipées de revues avec des comités de sélection des articles, elles ont lancé des concours d'idées, soutenu des expériences ou des expéditions, et elles ont distribué solennellement des récompenses. C'est très largement ce modèle du *club* intellectuel, plutôt élitiste et fermé, que reprendra ensuite l'université moderne du XIX^e siècle. Les jurys de thèse sont devenus des dispositifs de certification mais aussi de (pré)recrutement et donc, pour l'essentiel, d'agrégation à des petits mondes académiques. Les revues scientifiques à comité sont toujours des espaces de validation entre collègues ou aspirants collègues. Ces différents types d'outils d'évaluation (mais aussi de reconnaissance réciproque) restent aujourd'hui encore prévalents dans l'organisation des disciplines universitaires, même s'ils sont à présent concurrencés par de nouveaux dispositifs d'attestation de la valeur sociale accordée aux activités scientifiques. Même dans le pilotage des programmes incitatifs de recherche, qui mobilisent aujourd'hui des moyens considérables (appels d'offres de fonds nationaux, tel le Fonds National de la Recherche Suisse, agences de recherche, telle l'Agence Nationale de la Recherche française, programmes européens, tel le Programme Cadre de Recherche Développement) et qui déstabilisent en partie le fonctionnement académique classique, le recours à des jurys ad hoc d'évaluation prévaut toujours.

Toutefois, le changement en cours, qu'on l'appécie ou qu'on le critique, doit être reconnu comme profond. Les règles du jeu sont en train de se transformer parce que les financements de l'enseignement supérieur et de la recherche sont en train d'être réorganisés et partiellement réorientés. Dans le modèle de l'université issu du XIX^e siècle, l'Etat, (même s'il était plus présent dans les financements de l'université en Allemagne, en France ou en Italie que dans les pays anglo-saxons à universités privées), a délégué la majeure partie de la gestion et de l'application des règles générales aux enseignants et aux chercheurs eux-mêmes (recrutement des collègues, certification des activités, diffusion des travaux). Cette délégation qu'on peut donc appeler « corporative », c'est-à-dire remise à un corps

professionnel particulier, s'observe encore partout. Mais elle tend maintenant à être redoublée et en partie concurrencée par des programmes incitatifs à fort pouvoir financier qui se dotent de dispositifs spécifiques d'évaluation.

Dès lors, la délégation corporative de la certification opérée au bénéfice du système universitaire, s'il elle n'apparaît pas nécessairement diminuée en soi, se révèle relativisée par la montée en puissance d'autres dispositifs d'allocation de valeur. Le jugement par les pairs directs, au sein de chaque discipline et/ou établissement universitaire, se combine dorénavant avec d'autres formes d'évaluation faisant intervenir des logiques d'organisation (développement de grands organismes de recherche, gérant de nombreuses équipes et des carrières spécifiques, programmes nationaux et internationaux de recherche incitative) mais aussi avec des références à l'*utilité sociale* globale de la recherche. Ces deux facteurs impliquent la présence de nouveaux types de participants, parfois extérieurs au monde académique, dans les jurys et commissions de jugement de la valeur scientifique. Autre élément d'inflexion, on entend évaluer dorénavant non seulement les activités de publication des chercheurs et enseignants mais aussi l'encadrement des doctorants, la performance des parcours pédagogiques, les propositions de recherche aux appels d'offres, les apports des travaux sur contrats, etc, etc.

S'agit-il au total d'un changement de régime de l'évaluation? On vit aujourd'hui pour le moins un moment particulier, où l'évaluation scientifique des SHS est mise en situation de combiner le jugement académique classique, censée définir l'excellence, et l'évaluation de politique publique, supposée mesurer la pertinence et l'efficacité. Cela va impliquer l'invention de nouvelles formes de coordinations entre les jugements mais aussi entre les organisations. Pour certains, il ne s'agit là que d'un passage à un stade quasi industriel: il faut accéder à une autre échelle et une plus grande intensité dans l'évaluation scientifique. Mais on verra que c'est la création d'espaces nouveaux de débat et de délibération autour de l'évaluation des productions scientifiques qui est réellement en jeu. Vaut-il vers une pluralisation des profils d'experts? Comment agencer une coopération entre normes de nature différente? Et qui évaluera les évaluateurs? C'est donc le gouvernement de la société scientifique qui est concerné. L'évaluation scientifique apparaît de ce point de vue comme étant aussi une question politique.

I – L'ÉVALUATION COMME FORME DE JUGEMENT ET COMME MODE D'ORGANISATION

A. De l'évaluation artisanale aux démarches plus quantitatives

L'université est souvent considérée hors du monde, elle serait encore repliée dans sa tour d'ivoire. Cette impression d'enfermement ou de repli sur soi est facilement confirmée par les formes classiques de l'évaluation académique. Le jugement porté sur les activités y reste centré sur le milieu universitaire et sa reproduction, d'autant que l'attribution de valeur donnée aux publications et aux travaux est contrôlée par les pairs ou plutôt par la hiérarchie des pairs (hiérarchie définie par les titres, la réputation, l'ancienneté, les positions institutionnelles acquises). L'expertise savante se structure donc autour de mini collectifs, faits de collègues constitués en *commissions de spécialistes* ou en comités de rédaction.

La confrontation des systèmes de valeurs et des jeux d'intérêts s'opère alors en fonction de jugements formulés « en conscience » ou en conviction et par le biais de critères peu formulés de manière explicite mais qui renvoient globalement aux cadres intellectuels de chaque discipline et à une accumulation raisonnée de connaissances. Mais ce pouvoir des pairs ou *mandarins*, à l'intérieur de leurs mondes disciplinaires respectifs, est aussi la contrepartie des libertés académiques reconnues et le produit d'une délégation « corporative », on l'a déjà indiqué, de la gestion universitaire qui s'inscrit dans une histoire très longue (Musselin, 2006).

Pourtant, des mutations très significatives ont été engagées dès après la seconde guerre mondiale et ont modifié le rôle des protocoles endogènes d'évaluation. Dans les pays qui se sont dotés de grands organismes de recherche, de programmes incitatifs ou de dispositifs nationaux de régulation des carrières académiques, ces jeux micro-locaux de validation des personnes et des activités ont été partiellement relativisés par de nouvelles échelles de négociation et par un recrutement élargi des détenteurs du pouvoir d'expertise.

Jurys constitués à l'échelon national même, élection (au moins partielle) des membres de ces commissions d'évaluation, relativisation de l'expertise des mandarins par introduction d'autres catégories de personnel universitaire, ces trois caractéristiques d'ouverture potentielle du jeu classique se retrouvent par exemple dans les dispositifs contemporains du Conseil national des universités et du Comité national de la recherche scientifique en France. De purement endogamique, l'évaluation devient alors d'assise plus large. Les recrutements, carrières individuelles, les activités collectives deviennent observables sur une scène élargie, d'ampleur nationale. C'est en quelque sorte une évaluation « délocalisée », c'est-à-dire non localiste.

En dépit d'une coloration parlementaire ou du moins *représentative* donnée à ce type d'organismes élus partiellement, les règles prévalentes ne se trouvent pas entièrement transformées par rapport au modèle historique. Les jeux d'influence mandarinaux perdurent. Et les décisions prises en aval des évaluations, par exemple en France, se révèlent être fonction d'équilibres parfois subtils entre trois pôles d'acteurs au moins : le ministère ou les directions des organismes de recherche, les commissions nationales et chaque établissement universitaire ou patron d'une unité de recherche. Dans d'autres pays, où le système universitaire n'a jamais été décentralisé ou bien a été récemment décentralisé, les dispositifs locaux d'évaluation scientifique perdurent. Mais ils échappent en partie aux clubs des pairs professoraux pour être de plus en plus intégrés dans les outils de gouvernance des présidents d'universités qui affirment ainsi leur autonomie.

Un pas supplémentaire a cependant été franchi plus généralement en termes d'exogamie de l'évaluation scientifique par la multiplication de programmes « prioritaires » de recherche sur fonds publics. Ils s'accompagnent de dispositifs de pilotage et d'évaluation faisant, à côté du recours classique aux universitaires, une place élargie à des acteurs socio-économiques et politiques. Le cas est particulièrement manifeste pour les programmes dits de *recherche développement* ou de *recherche innovation*, où l'interface avec les entreprises est privilégiée sans que la mobilisation des sciences humaines et sociales soit pour autant exclue par principe. Le PCRD européen a été bâti dans cet esprit. Les appels d'offres prennent en charge des « questions de société » plutôt que des champs de recherche fondamentale. Les formulaires de réponse appellent à préciser quels sont les

apports attendus pour l'action. Les modalités de sélection des projets passent par des étapes multiples où les expertises de base sont croisées avec des critères de choix qui relèvent plutôt de la pertinence sociale immédiate ou de l'opportunité politique (répartition géographique des équipes, relais dans le monde décisionnel, etc).

Ces différentes évolutions vers des évaluations moins endogamiques et localisées ne rendent cependant pas compte de toute la réalité actuelle, car les dispositifs d'organisation de l'allocation de valeur scientifique restent encore fortement micro-corporatifs. Néanmoins, elles sont actuellement renforcées par la multiplication observable dans de nombreux pays d'agences de programmes et d'autorités indépendantes (Social Science Research Council aux Etats Unis, Economic and Social Research Council au Royaume Uni, Fonds National de la Recherche Suisse) qui contribuent à délocaliser les scènes d'évaluation et à les ouvrir à des experts internationaux ou étrangers au monde universitaire. En France, le CNER (Comité national d'évaluation de la recherche) créé en 1989 participait de ce même mouvement. Il inaugure en effet une ouverture de l'expertise au sein de son organisation collégiale (sur 10 membres, 4 *seulement* sont issus du milieu académique, 4 choisis « en fonction de leur compétence dans les domaines économique, social, culturel, scientifique et technique », 2 sont membres de hautes juridictions). Il peut de surcroît, préciser ses statuts, faire appel à des experts français ou étrangers. Et surtout, il est chargé explicitement de définir des « méthodes objectives d'évaluation » permettant d'apprécier « le bien fondé des orientations et l'adéquation des moyens affectés aux programmes ainsi que l'efficacité des coopérations mises en œuvre avec les entreprises » (statuts). Les inflexions sont ici notables, les sciences sociales n'étant pas considérées alors comme une exception aux évolutions qui parquent déjà les sciences dures : exogamie et recours à des experts profanes par rapport au monde académique, objectivation croissante des méthodes d'évaluation, référence aux attentes des entreprises...

Mais, tout autant que la composition et la distribution des organes d'évaluation, ce sont l'échelle et surtout la nature des jugements en valeur qui se transforment dans ce contexte nouveau des programmes incitatifs, des agences de moyens et des concurrences internationales en termes de société de la connaissance (Paradeise, 2005).

B. Les indicateurs de l'évaluation et la question des sciences de gouvernement

L'évaluation pose intrinsèquement la question de la mesure et de la comparaison (cf texte d'introduction à ce dossier), car il n'est pas d'argumentation possible du jugement sans étalonnage et appréciations comparatives. Mais dans les SHS, que peut-on mesurer ? Des découvertes, des brevets, ou plutôt des contributions à des débats, des constructions de données, des interprétations ?

A partir du noyau originel (rapport sur la thèse, évaluations des *referees* et du comité de rédaction des revues), le champ couvert par l'évaluation scientifique en SHS s'est en réalité progressivement diversifié. Mais plus encore, cette évaluation actuelle tend à vouloir ajouter des indicateurs de priorité et de choix publics à la mesure de l'activité scientifique elle-même.

Dans les instances d'évaluation des carrières, l'éventail des critères s'élargit. Le nombre brut des publications réalisées est maintenant systématiquement croisé avec la réputation des supports d'édition et de diffusion. Il est aussi décomposé plus méthodiquement selon les genres d'ouvrages ou les types de fonctionnement des revues (Hantrais, 2006). A cette première forme d'appréciation s'est ajouté plus récemment le taux de « publiants » par équipe ou centre de recherche. Ainsi l'activité professionnelle est-elle non plus considérée individuellement mais par rapport à une organisation collective du travail scientifique. Troisième facette de cette mutation, la plus actuelle, c'est la mesure de la *visibilité* du chercheur, de l'équipe, voire de l'université toute entière (cf. l'article de D. Pumain dans ce numéro). Ceci peut être caractérisé par la mesure de la fréquence des citations observable dans l'espace des revues scientifiques et des ouvrages, voire dans celui des supports grand public pour certains cas... Ceci aboutit à la définition d'un taux d'*impact* qui permet à son tour des classements. On passe ainsi progressivement d'une mesure de données brutes à des indices puis à la production d'*indicateurs*, qui, comme le rend explicite l'origine du terme, sert à construire des tableaux de bord pour l'action et des instruments modernes de pilotage gouvernemental (Ihl et alii, 2005)

Ces types d'instruments autorisent une certaine commensurabilité directe des productions, certes, mais ils comportent aussi des biais d'autant plus redoutables que l'effet médiatique, et donc politique, des classements auxquels on aboutit est important. Il n'est qu'à voir les débats autour du récent classement mondial des universités, dit de *Shanghai* (bien qu'il reprenne pour l'essentiel des critères américains d'activité universitaire).

La bibliométrie est, de manière plus générale, un outil qui vise à dé-contextualiser en partie l'activité scientifique (c'est-à-dire à la sortir de son milieu direct de production) pour l'inscrire dans une commensurabilité plus universelle (...qui peut correspondre, comme celle de Shanghai, à la culture dominante). Si elle est utilisée de manière mécanique, une bibliométrie, même sophistiquée comme celle basée sur les indices H corrigés, peut cependant devenir très réductrice, surtout pour les SHS. Certains considèrent cependant que la bibliométrie reste pertinente, à condition d'être en quelque sorte re-contextualisée, c'est-à-dire en apportant des correctifs ou des pondérations aux facteurs bruts d'impact. Ainsi, faudrait-il tenir le plus grand compte de l'audience relative des langues de publication, en particulier, mais aussi de la diversité de structurations disciplinaires qui sont historiquement très variées, voire du rapport plus ou moins direct des sujets de recherche aux attentes sociales et à leur traitement dans l'actualité. Dans une perspective qui considère les mesures bibliométriques comme devenues inévitables dans le contexte actuel, ces correctifs ou ajustements pourraient, même si c'est un paradoxe, aider à remettre la bibliométrie « en contexte ».

Mais à ce type de correctif, il faudrait encore ajouter une nécessaire prise en compte des biais qui tiennent au comportement stratégique des acteurs concernés par les démarches bibliométriques. Les grandes revues scientifiques (spécialement anglo-saxonnes), devenues un élément clé du dispositif de mesures bibliométriques, ont des positions et des parts de marché à défendre. Leurs stratégies jouent un rôle dans la sélection des articles, interférant avec le point de vue des *referees* et des comités scientifiques. Les effets en sont maintenant connus : fermeture relative aux nouveaux auteurs et aux outsiders, culture des « petits

mondes », faveur donnée à certaines modes ou écoles de pensée (A. d'Iribarne, in R. Barré et alii, 2007). Par ailleurs, on ne saurait sous-estimer la capacité d'anticipation de chercheurs ou d'équipes de recherche par rapport au rôle des indicateurs dans l'allocation des moyens et la gestion des carrières. Ce qui parfois les conduit à des tactiques d'instrumentation de ces types de mesure. D'où l'inflation des réseaux de citations croisées ou de co-édition systématique (le phénomène des « les auteurs en meute »). Aujourd'hui, ils peuvent être assez bien repérés, y compris automatiquement, et donc relativisés. Mais le *saussissonage* des résultats de recherches pour multiplier mécaniquement les publications et les effets de *labellisation* fréquents en sciences sociales (usage de formules attractives et de titres chocs) sont plus difficiles à contrer dans les mesures bibliométriques.

La prudence devrait donc être de rigueur devant ces indicateurs, d'autant qu'ils pourraient devenir des instruments généraux du gouvernement universitaire, naturalisés à partir des disciplines dites exactes, où existent des espaces de référence intellectuels homogènes (Passeron, 1991), avec généralement un *main stream* et la référence possible à une science *normale*.

Reste qu'il faut être conscient de l'ampleur de la pression sociale en faveur de ces *thermomètres* de l'activité scientifique. L'échelle des programmes incitatifs nationaux ou le poids financier du « Programme cadre de recherche-développement » de l'Union européenne en sont des exemples forts, probablement prospectifs. Les grandes agences nationales de recherche évoluent d'ailleurs dans le même sens. De ce fait, l'évaluation de l'enseignement universitaire et celle de la recherche ne sont plus seulement des activités artisanales de jugement par les pairs, ce sont aussi des *politiques publiques*, avec leurs deux grandes caractéristiques : l'importance croissante des logiques d'organisation ; et le rôle des financements publics rapportés à des enjeux économiques (Gaudin, 2004). Ces deux facteurs vont mécaniquement dans le sens d'un pilotage des politiques universitaires par indicateurs et d'une appréciation quantitative, et à court terme, des valeurs ajoutées par les recherches.

Mais l'analyse du tournant actuel ne peut s'arrêter là. Car si l'évaluation scientifique est à présent une politique publique dans tous les pays développés ou émergents, elle devient en même temps un espace potentiel de débat public. En somme, les mutations actuelles ne peuvent se résumer à un passage au stade industriel de l'évaluation scientifique et à la montée en puissance des mesures quantitatives de l'activité universitaire. Le changement actuel du régime de l'évaluation est aussi à considérer comme une inscription nouvelle du jugement en valeur dans l'espace public du débat.

II – L'ÉVALUATEUR FACE AU DÉBAT PUBLIC

A. Les corporations universitaires et leurs « chapelles »

Un fort soupçon pèse couramment sur l'impartialité du jugement des pairs dans les sciences sociales. C'est la suspicion que suscitent en général tous les milieux fermés, aux règles peu écrites et donc opaques vues de l'extérieur. Mais c'est plus spécialement l'impression que donnent les mondes académiques, traversés de courants divisés en multiples « chapelles », dit-on souvent. Et ce

serait, plus encore, le cas des sciences humaines et sociales qui souffriraient d'un manque d'unification théorique et de l'influence des controverses idéologiques. Par parenthèse, on pourrait démontrer aisément que les SHS n'en n'ont pas l'apanage... Plus largement, le jugement dans les corporations universitaires est marqué par des confrontations nombreuses, qui imbriquent très étroitement différentes dimensions, celle des analyses ou des démonstrations et celle des luttes de puissances, de territoires et de places.

Sur les conflits qui traversent l'évaluation scientifique à dominante endogène et sur les confrontations de jugements auxquels elle donne lieu, deux grands types de regards coexistent. D'un côté, le point de vue dit *stratégiste* ou réaliste, parce que centré sur les confrontations, observe des conflits entre jugements sur une même production scientifique. Il prend acte de la différence des systèmes de valeurs possibles et s'attache aux processus de négociation entre positions (ou intérêts). Il étudie l'agencement des compromis trouvés et des rapports de pouvoir dont ils rendent compte. C'est ainsi, pour reprendre un exemple célèbre, que le débat théorique et méthodologique entre la sociologie de Durkheim et la géographie de Vidal de la Blache au tournant du XX^e siècle a pu être référé à une concurrence entre ces deux disciplines en vue de créer une *science sociale globale* pour contrôler ainsi une large part du nouvel espace universitaire alors offert par la République (Karady, 1979).

Mais d'un autre côté, on peut choisir un point de vue presque opposé : non plus la perspective du conflit mais plutôt la volonté de construire un espace commun à des interlocuteurs multiples. En termes d'évaluation scientifique, vouloir fabriquer du *commun* impliquera de privilégier la commensurabilité des énoncés et des apports de résultats (cf article de P. Livet). Dans cette perspective, les apprentissages cognitifs et les valeurs partageables jouent un grand rôle. Mais la commensurabilité des énoncés, qui autorise une évaluation comparée, suppose un certain renoncement aux intérêts partiels de chacun.

Cette dernière approche tranche évidemment avec le cynisme relatif des points de vue classiques en science politique qui se veulent « réalistes », elle renouvelle les perspectives et dynamise les analyses de l'évaluation. Toutefois, on ne saurait oublier les conditions de possibilité qui s'imposent à elle et qui en restreignent beaucoup les applications. Ce point crucial ressort nettement des deux perspectives d'analyse qui se sont récemment attachées à l'examen des formes délibératives.

Ainsi, dans la théorie démocratique d'Habermas, la co-construction des normes va nécessairement de pair avec la pleine reconnaissance de l'Autre (Droit et démocratie, 1997). Et cela ne peut se réaliser que sous la double condition du déploiement d'argumentations de type véritablement rationnel et du respect de la *bonne foi* dans l'échange des points de vue entre acteurs sociaux. On voit dès lors quelles sont les limites de possibilité d'une délibération de ce type, qu'on l'applique à l'évaluation scientifique comme à d'autres domaines.

Par ailleurs, dans l'analyse des *Cités* argumentatives (c'est-à-dire des références globales) qui structurent les différentes justifications des acteurs sociaux, Boltanski et Thévenot ont étudié comment au cours des débats s'opère la *montée en généralité* des argumentations et leur confrontation préalable à des choix collectifs. Ils montrent (De la justification, 1991), comment ces arguments se ramènent en réalité à quelques grands types stables et limités. Mais ils notent

également comment les ajustements entre les diverses justifications en présence sont fonction des circonstances de la discussion et que les consensus construits restent toujours fragmentés et ponctuels.

Chacune à sa manière, par conséquent, ces deux analyses soulignent la fragilité et l'instabilité des coordinations fondées uniquement sur des procédures argumentatives. D'où la nécessité, nous semble-t-il, de ne pas en rester au niveau des seules interactions entre acteurs et de revenir à une prise en compte plus détaillée des contextes organisationnels dans les évaluations scientifiques.

B. Une évaluation distribuée et publicisée

L'évaluation scientifique doit-elle rester principalement le fait de clubs de proximité (jurys universitaires) constitués de pairs notabilisés? Répondre par la négative ne signifie pas automatiquement les remplacer par des acteurs économiques et sociaux, transformés ainsi en nouveaux arbitres des financements de recherche. Car si certains sujets ou thèmes de recherche peuvent requérir leur présence dans des jurys (au risque de fortes dissonances dans les systèmes de valeurs et donc dans les priorités), la question est d'ordre plus général. Elle concerne toutes les modalités d'élargissement des protocoles du jugement scientifique et de l'allocation de valeurs, au-delà des jurys universitaires. Il s'agit en effet aussi des procédures d'appels d'offres, celles des programmes des ministères techniques, celles des agences nationales de recherche et enfin celles des programmes finalisés des grands organismes de recherche eux-mêmes. Ces appels d'offres se dotent en effet chacun de conseils scientifiques et de jurys ad hoc, constitués de compétences individuelles puisées dans des mondes différenciés. L'élément de référence n'est donc plus l'établissement universitaire ou le centre de recherche: l'évaluation, dans ce cas, n'est plus auto-centrée, d'autant qu'un réseau d'experts, extérieurs au conseil scientifique ou au jury, est mobilisé en amont, (tout en laissant une certaine latitude au jury opérant la synthèse des points de vue extérieurs). En France, l'ANR, comme beaucoup d'autres programmes, fonctionne sur ce principe, où la pluralité systématique des expertises n'apporte toutefois pas nécessairement une garantie de pertinence ou de qualité des experts mobilisés. Et la Mission d'évaluation scientifique et pédagogique, qui existait jusqu'en 2007 au sein du ministère de l'Enseignement supérieur et de la recherche, a adopté cette même démarche pluraliste pour évaluer les projets et préparer la certification des masters et des écoles doctorales (dans le cadre de contrats quadriennaux d'objectifs et de moyens signés entre les établissements universitaires et le ministère).

Mais l'évaluation *représentative*, basée sur le jugement de pairs pour partie élus (de type CNU ou Comité national en France), doit-elle être laissée de côté pour autant? A sa création, elle a eu pour ambition de combiner le jugement des pairs et la représentation des courants syndicaux dans le monde universitaire. Elle a permis une réelle ouverture par rapport aux évaluations localisées et des formes de compte-rendu aux mandants. Elle est aujourd'hui menacée parce qu'on lui reproche de relayer les corporatismes et de cultiver l'entre-soi. Plutôt que de la marginaliser, il serait plus judicieux de vouloir combiner ses atouts avec une évaluation basée sur des comités de visite plus systématiques. Cela conduirait à une évaluation scientifique à la fois publicisée et distribuée.

C'est ce à quoi peut contribuer, par exemple, la nouvelle Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES), créée récemment en France comme autorité indépendante. L'évaluation n'y est pas seulement pluraliste, elle est confiée et distribuée à de multiples jurys d'évaluation extérieurs à l'Agence, constitués à chaque fois sous l'autorité d'un président pleinement responsable de son rapport d'expertise. Dans ce cadre, l'évaluation des centres de recherche et des filières de master n'est donc plus simplement externalisée ou sous-traitée, elle devient plutôt décentralisée et distribuée à travers l'espace scientifique. Car ces jurys sont constitués par ensembles géographiques et disciplinaires dans l'ensemble du pays. On peut d'ailleurs considérer qu'une telle évaluation distribuée prépare à des systèmes d'évaluation par établissement qui se développeront probablement dans le futur, en cohérence avec l'autonomie croissante des universités.

Autre inflexion importante dans les démarches nouvelles d'évaluation scientifique, c'est la volonté de publiciser les jugements et avis. Auparavant, l'opacité de l'expertise et l'anonymat de l'expert (destiné, disait-on, à le protéger) prévalaient dans beaucoup de pays. Sur demande plus ou moins expresse, l'évaluation pouvait être communiquée aux évalués. Pour l'essentiel cependant, des avis synthétiques et anonymes parvenaient seuls aux établissements. Les jugements scientifiques n'étaient ni aisément ni largement diffusés (surtout dans les pays de droit romain germanique).

L'accès de tous aux évaluations, par diffusion sur Internet comme l'organise en particulier l'AERES, donne une autre signification à ces évaluations. Il suggère un nouvel espace public d'information, voire de débat. Mais il met aussi en position renouvelée l'évaluateur ou le collectif d'évaluation. Le juge se soumet-il ainsi à son tour au jugement des autres ? L'évaluateur devient en tout cas redevable dans ce cas d'un certain contrôle du public. Et la divergence des points de vue au sein des jurys d'évaluation peut éventuellement prendre place dans cette diffusion élargie.

Tout ceci a pour corollaire le déploiement de protocoles d'évaluation très développés et de cheminements méthodologiques toujours plus détaillés. L'évaluation devient alors une démarche *processuelle*, où les conditions de forme constituent des éléments de la qualité et de l'acceptation de la procédure. Toutefois, dans ce cas de figure, il ne s'agit pas de la quête d'une évaluation proprement dialogique, d'un *bien commun* passant par le renoncement partiel aux intérêts privés (telle qu'évoquée par P. Livet), mais, de manière plus pragmatique, de l'agencement d'espaces de débat relatifs à l'évaluation scientifique, de formes nouvelles, mieux distribuées et publicisées, du jugement en valeur.

De cette manière, l'évaluation scientifique (comme organisation et comme politique publique) pourra prendre de la distance avec les connotations de contrôle, voire de tutelle, qui s'attachent souvent à elle, en particulier dans les pays où elle est d'abord un instrument préalable à la répartition de moyens aux universités et aux centres de recherche. Le jugement en valeur n'est plus directement une sanction ou un classement, il est d'abord, à travers le débat et des allers-retours entre les parties prenantes, l'élément d'un processus d'apprentissage collectif. Application négociée, par exemple, de nouvelles normes d'excellence et de pertinence, ou bien ajustement des indicateurs généraux d'évaluation (bibliométriques notamment) à des situations disciplinaires particulières ou à des formes

spécifiques d'organisation du travail scientifique. On retrouve là l'esprit de ce que, dans d'autres domaines de l'action publique, on appelle une démarche *d'évaluation-formation* (Conan, 1998). Ceci revient à considérer les collectifs de travail comme des organisations apprenantes et les apprentissages organisationnels comme le cœur des dynamiques réflexives contemporaines. Du moins est-ce peut être une perspective envisageable quand l'évaluation ainsi coordonnée ne se dégrade pas en dispositif normatif rampant (on pense ici à la démarche promue par la Commission européenne dite « Méthode ouverte de coordination » (avec sa trilogie circulaire : *benchmarking, ranking, best practices*) qui débouche sur des réglementations rampantes.

CONCLUSION

Actuellement, le tournant « organisationnel » de l'évaluation scientifique n'ouvre pas de perspective proprement délibérative. Le développement accéléré d'instruments et de procédures évaluatives appliquées au monde universitaire participe en effet d'abord de la sophistication accrue de ces « sciences de gouvernement » contemporaines qui accompagnent le contrôle des financements de politiques publiques et la promotion intensive de la *société de la connaissance* (en tant qu'impératif économique et culturel contemporain). Cependant, il apparaît que, si l'évaluation scientifique est devenue un outil de politique publique, c'est aussi un outil politique tout court, au sens où en complément des évaluations faites par des pairs élus, il peut faire place à un débat plus explicite et distribué sur les jugements experts en permettant des coordinations ouvertes entre évaluateurs et évalués.

Institut d'Etudes Politiques, Aix-en-Provence

BIBLIOGRAPHIE

- Boltanski, L, Thévenot, L, *De la justification*, Gallimard, 2001
 Conan, M, *L'évaluation formative*, Editions de l'Aube, 1998
 Gaudin, JP, *L'action publique*, Dalloz et Presses de Sciences Po, 2004
 Habermas, J, *Droit et démocratie*, Gallimard, 1997
 Ihl, O, Pollet G, Kaluzinski, M, *Les sciences de gouvernement*, Editions Economica, 2005
 Iribarne, A. d', in Barré, R, Theys, J, *Le management de la recherche*, de Boeck, 2007
 Hantrais, L, *Pour une meilleure évaluation de la recherche en sciences humaines et sociales*, Comité d'évaluation de la recherche, Documentation française, 2006
 Karady, V, Les Durkheimiens, *Revue française de sociologie*, XX, 1, 1979
 Musselin, Ch, *Le marché des universitaires*, Presses de Sciences Po, 2006
 Passeron, JC, *Le raisonnement sociologique*, Albin Michel, 1991
 Paradeise, C, *Piloter la réforme de la recherche publique*, Futuribles, 2005