

**afsp**

association française  
de  
science politique

## **Lettre de l'Observatoire des métiers académiques de la science politique**

**L'ÉVALUATION SCIENTIFIQUE ET  
LA CERTIFICATION DU « PUBLIANT » EN SHS**

par Jean Pierre Gaudin

**décembre 2008**

**n °6**

## L'ÉVALUATION SCIENTIFIQUE

### ET LA CERTIFICATION DU « PUBLIANT » EN SHS

Jean Pierre Gaudin

#### Introduction

On sait les débats intenses qui traversent aujourd'hui la question de l'évaluation des politiques publiques. Après les tentatives avortées du début des années 1990, des impulsions gouvernementales nouvelles voudraient faire de l'évaluation une évidence contemporaine. Mais les controverses, dans le même temps, se multiplient, tant à propos du bien fondé de certaines démarches évaluatives que des critères et surtout du choix des indicateurs de mesure. Longtemps remises aux bons soins des spécialistes du sujet, ces discussions sur les méthodes de l'évaluation commencent donc d'entrer dans l'espace public. Si les élites en perdent ainsi un peu le monopôle, les contenus des débats ne sont toutefois pas forcément plus éclairants. D'autant que presque chaque politique publique paraît avoir des spécificités redoutables, défiant les indicateurs comptables généraux de type bilan coût/avantage.

De ce point de vue, l'activité de recherche, envisagée dans le cadre des politiques scientifiques, apparaît comme un cas presque limite des difficultés méthodologiques qui se posent. La recherche sera-t-elle mieux mesurable par équipe de travail ou bien par individu ? A l'aune de quoi, précisément, va-t-on en apprécier les « résultats » ? Et comment pourra-t-on combiner des mesures quantitatives, susceptibles de fonder des *indicateurs*, et les appréciations qualitatives qui renvoient au jugement « en substance », sur l'originalité de la démarche, la prise de risques scientifiques, etc ?

Pour réfléchir à la manière dont on commence de répondre à ces questions, on prendra l'exemple d'un nouveau dispositif central d'évaluation de la production scientifique mis en place en France, l'AERES. Son domaine de compétences se veut très général. Il couvre en effet toutes les disciplines scientifiques et l'ensemble des unités de recherche labellisées en France par le ministère de l'Enseignement supérieur mais aussi par les grands organismes (CNRS, INSERM, INRIA, etc). Sur ce très large champ d'intervention, il a semblé nécessaire de se doter d'un outil de mesure de l'activité scientifique des équipes de chercheurs, pour éclairer de manière quantitative leurs « résultats », voire leurs performances. Ce n'était évidemment pas la première fois, car le ministère de l'Enseignement supérieur et plusieurs organismes de recherche avaient déjà expérimenté par le passé des démarches partielles ou sectorielles de ce type. Des « taux de chercheurs publiants » ont déjà été calculés par équipe ou laboratoire de recherche, pour caractériser des dynamiques collectives et des situations individuelles contrastées. Derrière la notion de *publiant*, on voit poindre une certaine mesure des *résultats* de recherche, une des façons de situer l'excellence, une des quantifications de la *performance*. Et la tendance est d'autant plus soutenue qu'elle doit s'appliquer non plus aux seuls dossiers individuels des chercheurs et enseignants mais à des collectifs nombreux (unités de recherche, grands centres de recherche, établissements universitaires même). Nous entrons dans une sorte d'âge industriel de l'évaluation scientifique.

Au sein du dispositif d'évaluation ici étudié, la notion d'*activité* des équipes de recherche a paru caractérisable de deux manières jugées complémentaires : qualitativement, par le biais de l'expertise des pairs universitaires ; et aussi quantitativement, au travers d'indicateurs de « production » scientifique. Mais l'idée générale d'activité a vite dû être décomposée à son tour selon deux grandes facettes : une activité directe de recherche, attestée au premier chef par des publications reconnues ; et des activités connexes mais indispensables (de gestion, d'animation et de valorisation de la recherche), passant par des modalités de certification qui sont autres.

Ainsi, le raisonnement évaluatif a progressivement mobilisé, et naturalisé dans ce domaine, les notions de mesure, d'activité et d'indicateurs. C'est là, il est vrai, une démarche classique de l'évaluation des politiques publiques. Mais en fonction de quels ajustements nécessaires cette démarche peut-elle s'appliquer à cette activité si particulière qu'est la production de recherches (Paradeise, 2005) ? Bibliométrie ou encore docimologie généralisée : le problème posé n'est pas que celui du dénombrement et de son adaptation à la réalité de la recherche ou à la grande diversité des disciplines. C'est aussi, comme on le verra, celui de l'environnement d'ensemble dans lequel s'opère aujourd'hui l'évaluation scientifique. Plusieurs facteurs contextuels interviennent, en effet, dans les perceptions actuelles du développement des mesures bibliométriques, et ce en SHS particulièrement. Risque qu'un modèle de mesure de l'activité de recherche né dans les sciences exactes ne soit généralisé sans guère de précaution et qu'il renforce donc d'autres effets de domination

sur l'ensemble des disciplines. Crainte, aussi, que l'objectivation (via des indicateurs) des activités dans le monde universitaire ne rejoigne des projets de réforme connus qui dessinent des hiérarchies nouvelles pour les enseignants chercheurs. Regret, enfin, du poids croissant des standards internationaux anglo-américains dans les bibliométries informatisées, face à nos spécificités linguistiques et culturelles fragiles. En surface, on pourrait ne voir dans tout cela que la frilosité de quelques disciplines devant l'émergence des critères quantitatifs d'évaluation. L'enjeu est plus substantiel, en vérité, car ces réserves soulignent des problèmes épistémologiques qui traversent classiquement toutes les méthodes d'allocation de la valeur dans le domaine scientifique.

Mais on fait l'hypothèse que les controverses actuelles sur la bibliométrie s'inscrivent plus largement dans une transition d'ensemble. Celle entre une évaluation scientifique « à dire d'expert », d'une part, où se mêlent très globalement une compétence reconnue par les pairs, une expérience importante et l'autorité des titres, et d'autre part un modèle moins *démiurgique* de jugement, où l'objectivation des critères est privilégiée, voire définie pas à pas, et où les protocoles d'évaluation se veulent très détaillés et explicites. C'est là une démarche de type « procédural », qui participe aujourd'hui des nouveaux répertoires pragmatiques de l'action publique.

C'est dans pareil contexte d'ensemble qu'on retracera comment la caractérisation du « publiant » s'est ici opérée selon trois temps successifs. Au point de départ, il y a d'abord la recherche d'un critère d'activité de publication qui puisse avoir une portée générale pour les SHS. Ensuite, sous l'effet de critiques croisées, on note des tentatives de diversification et de raffinement des mesures, quasiment discipline par discipline. Et pour finir, on peut observer le retour consensuel à un indice quantitatif général de la valeur des publications, mais pondéré cette fois très explicitement par des arbitrages de nature qualitative. Cette séquence, si elle traduit des hésitations ou des conflits, n'est pas erratique pour autant. On montrera dans cet article qu'elle s'inscrit en effet dans des processus d'apprentissage collectif de démarches nouvelles de jugement scientifique, au sein de ces équipes d'évaluateurs et (plus largement) dans les différentes communautés disciplinaires. Ce qui conduira à s'interroger sur la portée d'une rationalité procédurale dans l'évaluation des politiques publiques.

## 1 – Vers une mesure unifiée des publications scientifiques

### a- Construire des règles générales d'évaluation scientifique ...

Dans une première note d'orientation (qu'on datera N1, sur un intervalle total d'observation de deux mois environ) produite par les coordonnateurs de ce dispositif national d'évaluation, il a été retenu d'emblée un double critère, à la fois quantitatif et qualitatif : il porte sur le total de ce qui est publié par une équipe de recherche sur le moyen terme des contrats (4 ans) ; et il distingue en même temps ce qui est reconnu comme plus ou moins légitime par les communautés scientifiques.

D'où plusieurs « genres » de publications explicitement référencés dans cette note : articles dans des revues, ouvrages personnels, participation à des ouvrages collectifs, communications publiées, ou encore élaboration de bases de données ; et une première estimation des quantités minimum requises pour chacun de ces genres. Ces catégories sont sans surprise ; elles se retrouvent déjà dans nombre de dispositifs particuliers d'évaluation scientifique préalables (CNRS, CNU, MSTP).

Pour ce qui concerne la valeur des articles scientifiques, la démarche adoptée renvoie à la méthode classique du classement des revues en plusieurs niveaux, où « l'excellence » serait caractérisée par trois critères majeurs : existence d'un comité de rédaction actif, recours à des *referees* extérieurs, impact international de la revue (diffusion et/ou citations). Pour les communications publiées, cette première note d'orientation renvoie également à des distinctions habituelles ...mais un peu floues si on les examine en détail : colloque « international » ou non, colloque avec publication ou non (cf infra).

On note également que, dans ce texte initial, ont été ajoutés à la mesure préalable des quantités de publications d'autres indices potentiels de la qualité des activités scientifiques. Les indices ont trait à la fois au « rayonnement » des publications réalisées mais également à l'évaluation d'autres activités du chercheur que celles de l'écriture. Au titre du rayonnement, sont évoqués ce qu'on appelle communément le *facteur d'impact* des publications (nombre de citations dans d'autres revues ; facteur « H » rapportant les citations au nombre des supports de publications, etc) et encore le degré de prise de risque dans des sujets nouveaux ou des démarches interdisciplinaires. Au titre des autres activités, sont mentionnées la prise en compte des fonctions d'animation de réseaux scientifiques, ainsi que les activités de « réponse » à la demande sociale (conseil, participation au débat public, etc) ou encore celles, longtemps négligées, de gestion de la recherche (participation aux conseils, direction d'équipes ou de laboratoires).

b- ... Sans oublier la diversité des cultures disciplinaires

Après des premiers énoncés, destinés à construire une base de travail commune aux experts autour de règles quantitatives sur les publiants, le texte d'orientation initial introduisait cependant plusieurs critères de modulation, laissés à l'appréciation des experts :

- la prise en compte de l'âge, ou plutôt le degré d'avancement de chaque chercheur dans la carrière, ainsi que l'engagement individuel dans des tâches collectives ;
- et la prise en compte « des pratiques de chaque communauté scientifique ».

C'était là un énoncé plutôt lapidaire, pour une modulation pouvant relativiser fortement l'approche quantitative générale (Hantrais, 2006). Formulé *in fine*, il autorisait à moduler la comptabilité des publications selon chaque discipline, sans en préciser les règles à ce stade. L'appréciation se trouvait alors remise implicitement par le texte d'orientation N1 à chaque expert, en fonction des secteurs disciplinaires.

c- L'implicite d'un *socle commun* aux SHS

Dans cette note N1, apparaissent par conséquent deux objectifs à la fois : celui de tenir compte de la spécificité des SHS (voire de chaque discipline) par rapport à l'ensemble du champ scientifique et celui de pouvoir prendre, ainsi qu'il est écrit, la mesure « du développement de la recherche *en général* ».

Pas question, en effet, de n'adopter que des unités de mesure standards, à la différence par exemple du mode de faire de l'INSERM qui applique exactement les mêmes indicateurs d'activité scientifique à l'ensemble de ses unités, même celles de sciences sociales (marginales en nombre, il est vrai). Toutefois, l'idée prévaut dans la orientation N1 qu'au delà des modulations applicables en fonction des cultures disciplinaires, un même socle de règles peut servir de base aux évaluations en SHS. Ces règles générales, exprimées en termes de quantités pluriannuelles de publications « légitimes » par équipes, pourraient venir encadrer les appréciations trop subjectives (ou partisans) des *diriges d'expert* et réduire ainsi les marges de l'erreur manifeste. Du moins était-ce là un objectif encore implicite à ce stade de rédaction, car le raisonnement n'est pas détaillé dans la note d'orientation initiale. Comme si les limites du *dire d'expert* étaient évidentes pour chacun et du même ordre dans chaque discipline. Et comme si l'articulation des deux méthodes, l'expertise subjective et la mesure quantitative, allaient de soi. Or cette combinaison du qualitatif et du quantitatif, tant en termes méthodologiques purs qu'en fonction des routines de raisonnement, pose des problèmes parfois difficiles. Et leurs dosages sont loin d'être les mêmes dans toutes les disciplines des SHS.

## 2 – Une difficile prise en compte d'exigences contradictoires

a- L'hétérogénéité des jugements en valeur au sein des SHS

Les précautions prises en termes de différenciation des cultures de jugement étaient par conséquent peu développées mais bien présentes dès le texte d'orientation N1 sur la caractérisation des *publiants*.

Pourtant, des réactions trop rapides par Internet (et donc non nuancées par des échanges de parole) et, en contrepartie, la présentation un peu abrupte de cette note d'orientation ont provoqué des réactions multiples et parfois vives dans l'équipe des chargés de mission SHS de ce dispositif d'évaluation scientifique. Mais au delà des données circonstancielles, il est apparu au cours de cette controverse que la difficulté à se rallier pour l'essentiel à un socle quantitatif d'allocation des valeurs venait non seulement de l'histoire de chaque discipline, de son institutionnalisation ou de son rapport à la commande de recherche, mais aussi (plus épistémologiquement) de la manière dont s'y construisent les objets de recherche et dont se définit la connaissance légitime.

Dans ces conditions, les réactions à la note N1 et les contre-propositions émises sont presque toutes allées dans le sens d'une demande d'explicitation accrue et d'un détail plus grand des critères de différenciation par discipline. Avec le souci corrélatif, de ce fait, que ces différences soient bien objectivées dans les futurs textes méthodologiques de référence qui seraient affichés par le nouveau dispositif d'évaluation.

Certaines de ces suggestions relevaient plutôt de l'histoire et de la construction normative de chaque discipline (par exemple, conception différente du travail d'*editing* pour les ouvrages collectifs ; ou usage d'une langue véhiculaire de nature variable par discipline, naturelle ou technique, nationale ou internationale). Car ces facteurs ont un large impact sur les types de publications, leur quantité et leur diffusion. D'autres modulations à formuler tenaient plutôt à la spécificité relative des matériaux traités et élaborés (place des rééditions « critiques » ; ou importance des bases de données). D'autres, encore, se rattachaient plus profondément au degré de différenciation épistémologique, y compris

en SHS, que l'on peut établir, au moins typologiquement, entre des sciences de la preuve et des sciences du « récit » (Passeron, 1991).

C'est la combinaison de toutes ces variables de nature différente qui rend le champ des SHS particulièrement hétérogène. Cela justifiait d'une certaine manière les demandes de différenciation écrite des critères de valeur, au delà de l'énoncé laconique de la première rédaction N1 qui paraissait ne pas apporter assez de garanties. Le paradoxe étant que l'idée initiale de créer une norme quantitative commune (complétée ensuite par les ajustements qualitatifs des experts) a induit, en réaction, des demandes de différenciation très détaillée et qu'on puisse préciser par écrit pour chaque discipline ou presque. La souplesse implicite engendrait donc paradoxalement la complication et le risque de rigidité.

Ces dangers ont toutefois été rapidement ressentis de plusieurs côtés dans l'équipe SHS, que ce soit à travers les réactions écrites au texte initial ou bien lors de réunions de travail sur ce sujet, qui ont eu lieu au bout de quelques semaines. Cela a en définitive induit une demande un peu contraire de la précédente, c'est-à-dire celle de réunifier ou de coordonner d'une manière ou d'une autre des critères d'évaluation détaillés discipline par discipline, qui apparaissaient désormais menacés d'éclatement et de balkanisation.

b- Les « passerelles » entre les évaluations disciplinaires, ou de l'utilité d'équivalents généraux

Lors de ces discussions au sein de l'équipe SHS relatives à l'évaluation quantitative des publications et à la caractérisation des publiants, on a vu par conséquent deux suggestions opposées se présenter successivement puis conjointement. Faire place, d'une part, à la diversité dans l'appréciation de la valeur des productions scientifiques légitimes en SHS ; et créer, d'autre part, un système de passerelles entre évaluations disciplinaires, sous forme de « paiement » des productions de recherche en nombre de *points* ou d'*étoiles*. Ce qui permet au final d'établir des *scores* avec une même unité de mesure comparative.

Cette dernière démarche a en fait été déjà expérimentée dans certaines sections du Comité national de la recherche scientifique au CNRS. Elle y sert principalement à objectiver la mesure des activités de publication et de communication dans les colloques ; et ainsi à faire *contrepoint* aux évaluations qualitatives de l'excellence et du rayonnement des productions scientifiques des équipes (ou des chercheurs pris individuellement). Au CNRS en l'occurrence, chaque discipline SHS construit pour l'instant son raisonnement et sa doctrine concernant cette notation en points ou étoiles. Les économistes, par exemple, ont particulièrement développé cette mesure en termes d'étoiles. Cette section du comité national précise bien que les nombres d'étoiles, correspondant aux divers rangs de publications, ne s'additionnent pas vraiment ; mais certaines associations d'écoles de commerce commencent déjà pourtant à les comptabiliser de manière cumulative pour établir leurs classements comparatifs entre établissements au plan national ...

Dans le cas du dispositif national d'évaluation ici analysé, l'enjeu d'une mesure par points ou étoiles devient plus large. Il s'agit non seulement de classer à l'intérieur de chaque discipline ou sous-discipline mais aussi de ménager des passerelles et des équivalents entre disciplines des SHS. Toutes proportions gardées, on se trouve ici dans une perspective voisine de celle d'un Etat cherchant, comme dans le passé, à unifier différentes monnaies régionales autour d'un étalon national, ou bien de l'Europe créant récemment l'euro. La difficulté est de trouver les bons taux de conversion, tenant compte des « pouvoirs d'achat » respectifs ou, plus largement, des valeurs d'échange préalables

Ce type de démarche, proposé un temps pour l'évaluation des SHS, a conduit en l'occurrence les divers chargés de mission disciplinaires à suggérer dans leurs contre propositions écrites (série d textes N2) une grande diversité de paiements (ou nombre d'étoiles) pour un même type de publication (article dans revue à comité ou ouvrage, par exemple). Mais aussi à donner des fourchettes d'évaluation parfois larges pour chacun d'entre eux, avec des arbitrages laissés en définitive aux experts... Ainsi, l'espoir mis dans cette méthode restait bien de créer des passerelles entre les unités de compte et d'objectiver un peu mieux les critères d'évaluation et les *paiements* qui y correspondent. Mais cela comportait aussi deux conséquences : ces règles écrites poussaient à des argumentations parfois très circonstanciées et à calculs un peu fastidieux ; par ailleurs, des marges d'appréciation importantes étaient réintroduites par les fourchettes, ce qui pouvait laisser place en définitive à autant d'arbitraire qu'auparavant...

c- Un contexte sous-jacent d'inquiétude et de soupçons

Développer des règles d'évaluation très générales, mais ménager la possibilité de les décliner bien différemment d'une discipline à l'autre. Ces deux exigences peuvent ne pas paraître pas incompatibles a priori, mais en pratique elles sont apparues traversées d'exigences contradictoires. Si, en effet, un tel paradoxe est gérable dans son principe, la démarche correspondante est apparue rapidement source de polémiques entre collègues de disciplines différentes, suscitant des débats sur un ton parfois vif et même des procès d'intention au sein du collectif d'évaluation. Cependant, la cause n'en est pas que de nature docimologique, car elle tient aussi pour une large part en France au contexte universitaire d'ensemble de ce débat de méthode particulier

On sait le malaise actuel dans l'université française, avant comme après la loi LRU, qui n'épargne pas les enseignants ni les chercheurs. Il y a la grande question du repositionnement de l'institution, de ses finalités et de ses moyens. Il y a également la redéfinition éventuelle des profils professionnels et des recrutements (Musselin, 2006). Il faut aussi tenir compte des facteurs d'inquiétude concernant la reconnaissance du travail scientifique en SHS, qui s'ajoutent à la classique revendication identitaire de chacun (« notre discipline est à défendre », « on ignore nos spécificités », « nous n'avons pas les mêmes critères d'utilité ou d'excellence que d'autres »). De surcroît, le contexte actuel est marqué par un rapide replis du français comme langue internationale et par le développement de classements comparatifs internationaux, toutes disciplines confondues, où comptent avant tout les publications en anglais et les revues plutôt consacrées à la « science normale ». S'y ajoute encore des interrogations relatives à un projet de réforme des profils professionnels des enseignants (déjà introduite par le rapport Belloc, 2003). Une nouvelle hiérarchie des activités, donnant la priorité au rayonnement international et aux facteurs d'impact pourrait en découler, qui se fonderait sur une évaluation des productions scientifiques essentiellement quantifiée et en termes de facteurs d'impact.

Au total, les classiques susceptibilités disciplinaires ont donc été fortement accrues lors de cette controverse par les méfiances dues à la conjoncture. Cela a conduit, dans les contre propositions N2, à défendre un temps des clés d'évaluation modulées selon chaque discipline, sans que le recours à des unités de compte communes (étoiles ou autres points) n'en rendent pour autant les transpositions bien lisibles. De plus, la formulation écrite et très détaillée des nombreux cas de figure envisageables (par exemple : ouvrages à *editing* explicite ou non, colloque à organisation internationale, ou à communications internationales, en langue anglaise ou non, etc) s'avérait finalement lourde et compliquée, sans apporter beaucoup plus de garanties objectives en contrepartie.

Pourquoi, dès lors, ne pas revenir finalement à un énoncé concis des règles les plus générales et renvoyer, très explicitement cette fois, pour leur mise en œuvre à l'évaluation des experts et des diverses communautés scientifiques ? C'est cette solution qui, au sein de l'équipe du dispositif, sera en définitive consensuellement retenue. Cette approche, on le verra, ne peut pour autant s'analyser comme un simple retour aux évaluations substantielles *à dire d'expert* du passé car elle met en œuvre une démarche de nature nettement procédurale. Ces souhaits de critères très détaillés et de protocoles d'évaluation différenciés indiquent en même temps que le modèle de l'expertise substantielle et globale devient délaissé. Le souci d'argumentation du jugement apparaît plus largement partagé dans ces milieux de travail. Et la volonté d'exposer complètement les critères et les processus d'allocation de valeur indique une attente nouvelle mise dans le déploiement de procédures explicites.

### 3 – La pondération du chiffre et de l'expertise qualitative

En apparence seulement, la solution retenue au final dans le dispositif national d'évaluation (N3) semble combiner de manière classique l'expertise substantielle et quelques indicateurs chiffrés.

#### a- Des règles docimologiques a minima ...

Le texte de référence indique : « l'activité d'une unité de recherche est fonction de la qualité de la production scientifique et du rayonnement, auquel s'ajoute la prise en considération de l'organisation et de la répartition des travaux entre ses membres. L'un des indicateurs est une estimation de la proportion des chercheurs et enseignants-chercheurs « publiants ». Est considéré ainsi celui qui, dans le cadre d'un contrat quadriennal, satisfait à un nombre minimal de « publications, nombre pondéré selon la situation dans la carrière et l'engagement dans des tâches d'intérêt collectif pour la recherche ».

Suit alors un tableau général qui croise différents secteurs disciplinaires et les deux statuts de chercheur et d'enseignant-chercheur. Dans les cases ainsi définies, sont portés des quantités minimales requises de « productions de rang A » sur quatre ans.

Mais à cela est immédiatement ajouté que « le poids relatif des différents types de publications est propre à chaque discipline. Les appréciations de chaque support de publication scientifique et des formes de contribution à la recherche sont différenciées en fonction des pratiques de chaque communauté scientifique ».

Le même texte (N3) précise, ensuite, qu'est considéré comme « une production scientifique de rang A :

- une publication dans une revue internationale avec comité de lecture considérée comme de TRES bon niveau par la communauté,
- un chapitre d'ouvrage ou un ouvrage de recherche reconnu internationalement, ou de portée nationale en SHS),

- pour le secteur STIC ou SHS, un article long dans un congrès international à comité de sélection considéré comme sélectif par la communauté,
- pour le secteur SHS, la constitution de bases de données accessibles ou de corpus de référence, les éditions critiques
- un brevet déposé à l'international. »

Et, in fine, une modulation complémentaire très large est encore ajoutée par ce document de référence :

« Pour tenir compte de la diversité des publications dans le secteur SHS, la liste ci-dessous donne à titre indicatif des équivalences possibles :

- a. Un article dans une revue à comité de lecture (ou considérée comme de très bon niveau par la communauté) est évalué selon la difficulté d'accès à la publication et le nombre de co-auteurs. Ce type de revues implique une procédure explicite de sélection et de révision des articles; n'entreront dans ce décompte que les revues de portée internationale ou nationale ; une liste établie avec le concours de la communauté scientifique, comportant, lorsqu'un consensus peut être construit, une classification en trois catégories (internationales, nationales et locales ou professionnelles) sera affichée et révisable. Les articles publiés dans des revues locales ou professionnelles sont comptabilisés tantôt au titre de la recherche, tantôt au titre de la diffusion de la culture scientifique ou de la valorisation selon les disciplines.
- b. Un bon ouvrage scientifique de recherche ou d'érudition est une publication de rang A (les éditions critiques, les manuels, traités, précis, essais, compilations sont appréciés au cas par cas). La direction d'ouvrages collectifs est comptabilisée dans les disciplines où elle est une forme de production essentielle.
- c. Une communication à un colloque ou un congrès ne peut pas être comptée deux fois si elle est publiée par ailleurs. »

b- ... avec, par conséquent, un large renvoi à l'expertise qualitative

Même pratiquée à plusieurs voix (groupes d'évaluateurs, expertises multiples), l'expertise qualitative présente plusieurs dangers que l'on connaît bien. Celui des jugements partiels et des prises de position marquées par les conflits d'école ou l'affrontement des courants d'analyse. Celui, aussi, des monopoles sur certaines filières détenus par quelques mandarins universitaires ou des positions insitutionnelles qui font autorité presque à elles seules. Ou encore celui, qui se situe bien au delà de ces situations abusives, d'un cloisonnement fort entre les jugements en valeur : chaque expert peut bâtir une doctrine assez variable d'une question scientifique à l'autre, voire d'une conjoncture à l'autre.

A l'inverse, le renvoi substantiel au jugement des experts tout au long du processus d'évaluation valorise évidemment leur rôle et les motive plus que l'application comptable de classements et de « tarifications » par étoiles. L'importance donnée à l'expertise qualitative ne peut que convenir aux experts les plus expérimentés et à ceux qui sont devenus des spécialistes de l'évaluation. Certains en tirent une fierté et une gratification symbolique, qui auréole le rôle classique de « tuteur » (sinon de contrôleur) qui s'attache encore fortement à l'évaluation. La valorisation de « l'expérience » plait donc nécessairement à ceux qui se vivent comme l'élite de ce métier émergent. Chaque chargé de mission du dispositif national d'évaluation peut, en revendiquant les spécificités de « sa » discipline, être mieux perçu par sa communauté scientifique. D'autant que la défense d'une culture commune fait ciment et alimente des reconnaissances croisées. Elle permet aussi aux experts évaluateurs de ne pas se couper de leur « milieu » professionnel, de ne pas apparaître comme de purs agents de l'administration et des bureaucraties évaluatrices. Et au final, de se positionner en médiateurs entre ce milieu et les règles générales et impersonnelles.

Mais de manière plus significative, sans doute, une pluralité des critères, qualitatifs et quantitatifs, conduit à une confrontation entre des valeurs « incommensurables ». La cohabitation ainsi organisée du quantitatif et du qualitatif amène donc à une discussion explicite, à la fois entre experts en charge du processus et avec la communauté de référence (Verdier et Lanciano, 2005). Car on n'oubliera pas qu'au final, dans le guide méthodologique N3 sur les publiants, le renoncement à des unités de compte du type « étoiles », s'accompagne en contrepartie d'une référence répétée aux critères de jugement, non pas d'un seul expert, voire d'un duo ou d'un jury d'ensemble, mais bien de chaque communauté scientifique ou discipline en entier. Face à l'évaluateur individuel, ce n'est pas une docimologie *a priori* qui est censée faire garde-fou mais le consensus du groupe des pairs. Reste à savoir comment cet accord est construit, et à travers quels processus ou apprentissages collectifs.

#### 4 – Indicateurs et nouvelles approches « procédurales » de l'évaluation scientifique en France

La relativisation des critères quantitatifs concernant les standards de publication ne peut donc être comprise comme un pur retour en arrière ou une ratification des modalités « aristocratiques » de la gestion des rapports professionnels dans les communautés universitaires. D'autant que cette lecture n'éclairerait qu'une part des enjeux actuels de l'évaluation scientifique. En effet, l'évaluation des politiques publiques, y compris celles d'enseignement et de recherche, s'inscrit maintenant dans un mouvement d'ensemble où des liens nouveaux se tissent entre pratiques de gouvernement et production des connaissances pour l'action. La réflexivité nouvelle de l'action publique cherche de plus en plus dans des comités de sages, des évaluations ou des démarches prospectives ses outils symboliques et pratiques. Or ces connaissances pour l'action publique alimentent le développement de procédures de discussion et de coordination nouvelles entre acteurs (d'Iribarne, 2007). On peut dès lors considérer la méthode adoptée en l'espèce pour l'évaluation des publications scientifiques moins comme un retour à des modes de faire préalables que comme l'expérimentation d'une approche procédurale de l'évaluation scientifique.

##### a- Procédures et *rationalité procédurale*

Le dispositif national d'évaluation de la recherche et de l'enseignement supérieur dont il est ici question s'affiche à la fois comme distinct des modalités de recrutement ou de gestion des carrières des grands organismes de recherche et comme étant indépendant des ministères. Il a été constitué sous forme « d'autorité indépendante » (ou d'agence de régulation, comme disent les anglo-saxons). Il échappe, en principe, aux règles de la hiérarchie administrative et de la comptabilité publique. Il est doté d'un exécutif autonome et collégial (conseil nommé toutefois par décret).

Les autorités indépendantes (ou agences de régulation) qui se multiplient maintenant en France ont des niveaux de compétences et des répertoires d'intervention assez diversifiés. Au minimum, c'est un rôle d'*observatoire* des pratiques du milieu considéré, permettant ainsi d'établir des mesures et surtout des comparaisons (y compris internationales). Et, le plus souvent, elles ont en charge la production et l'organisation de normes collectives, qui peuvent aboutir à des recommandations ou, après consultations avec une diversité d'acteurs, à ce qu'on appelle des « codes de bonne conduite ». Enfin, pour partie d'entre elles (mais pas dans le domaine de l'évaluation scientifique), elles assurent le suivi de l'application de ces normes collectives, débouchant sur un véritable travail de « gendarme » : injonctions, voire sanctions ou pénalités, mais sous le contrôle d'autres autorités ou des juges (Gaudin, 2004).

La création de ces autorités ou agences se rattache à une approche anglo-américaine des pouvoirs administratifs et de la formation de la règle générale. Mais les agences se développent maintenant très rapidement en France (notamment sous l'influence de l'Union européenne). On aurait tort d'y voir uniquement une solution de type organisationnel (souplesse, réactivité, etc), ou un simple outil du New public management. Car est en jeu, beaucoup plus largement, une approche reconsidérée de la construction et de la gestion de l'intérêt collectif. Avec ces dispositifs de régulation, la décision publique est moins envisagée comme visant l'optimum (c'est-à-dire un but très général, souvent presque idéal), mais comme recherchant plus pragmatiquement ce qui peut être « satisfaisant », et ce en fonction des conditions du moment, des ajustements possibles et des compromis réalisables (Conan, 1998). C'est ce qu'on nomme (après Herbert Simon) « une rationalité procédurale ». Cette conception de l'agir collectif, reprise depuis amplement dans les travaux de J. Rawls, de J. Habermas et des pragmatistes américains, correspond à tout un ensemble de pratiques administratives et politiques dans les pays autres que ceux marqués par la tradition du droit public romain germanique.

En l'occurrence, procédural ne signifie donc pas *qui s'attache aux règles du procès* (comme dans le vocabulaire classique des juristes). Il s'agit plus globalement d'un état d'esprit général, d'une démarche *pragmatiste* qui procède par étapes et expérimentations successives (Laredo et Mustar, 2001). Cette sorte de véritable « carte mentale » est donc caractérisée par le respect de protocoles et de processus, lesquels sont censés être les garants du résultat final. Pareil état d'esprit apparaît également dans le dispositif d'évaluation de la recherche que l'on observe ici. Les guides de méthode, les *déroulés*, les protocoles y tiennent une grande place, au delà de la simple formation d'une culture collective. Mais comment, au sein de ces procédures censées garantir le jugement en valeur, va-t-on pratiquement conduire ces processus (Boltanski et Thévenot, 2001) ? Comment réguler la combinaison entre données chiffrées et jugements qualitatifs dans l'évaluation scientifique ? Et comment intégrer les différentes évaluations dans une approche globale ?

##### b- Vers une « méthode ouverte de coordination » des expertises ?

La méthode adoptée montre que cette démarche d'évaluation de la recherche s'appuie sur deux caractéristiques clés. D'une part, un équilibre détaillé entre critères qualitatifs et quantitatifs au long des étapes de l'évaluation, qui n'interdit pas cependant des comparaisons entre situations (texte N3). D'autre part, un étayage systématique du jugement individuel des experts sur des jurys distribués par sites ou par établissements et sur des consensus méthodologiques propres à chaque communauté disciplinaire. C'est là un pari de type procédural, qui en rejoint aujourd'hui de nombreux autres, dans ce domaine comme dans d'autres politiques publiques



On se souvient, en effet, des prises de position du Conseil scientifique de l'évaluation qui avait été mis en place dès le début des années 1990 pour réfléchir aux choix méthodologiques requis par l'évaluation des politiques publiques alors lancée par l'Etat. Il avait considéré d'emblée que les formes de jugement en valeur qui sous-tendent l'évaluation de l'action publique ne pouvaient sérieusement se référer à des optimums abstraits (à la différence de la rationalité substantielle de la décision, telle qu'elle est formalisée dans les modèles des économistes). Et il conseillait déjà d'envisager pour ces évaluations une rationalité de type procédural, c'est-à-dire basée sur des ajustements progressifs de critères et des démarches construites par étapes successives (B. Perret, 2001).

De son côté, l'Union européenne avait considéré qu'une régulation moderne de l'action publique devait souvent passer par la création d'agences autonomes « à l'anglo-saxonne » produisant et pilotant des règles décentralisées, ainsi que par la systématisation de démarches de « coordination ouverte » entre acteurs autonomes. Il faut dire que cette dernière proposition pouvait servir ses initiatives dans les nombreux domaines où elle ne dispose pas de compétence communautaire explicite (tel l'éducation, ou les droits sociaux, par exemple). Elle a donc cherché, surtout à partir de 1990, à multiplier des opérations visant à faire converger des types de choix publics nationaux, même sans pouvoir édicter de directive ou de règlements (politiques de santé, ou d'éducation, par exemple). Une telle démarche, dite officiellement « méthode ouverte de coordination » (MOC) s'appuie généralement sur un triple dispositif d'ajustement entre politiques nationales, en mettant en œuvre à la fois des étalonnages comparatifs (benchmarking) des actions publiques, des classements de « performance » (ranking) et des codes de bonnes pratiques (best practices). Ces méthodes, issues de l'organisation industrielle japonaise, inspirent aujourd'hui le New public management. On reconnaît là aussi certains éléments communs aux démarches des agences et autorités de régulation qui travaillent en particulier par comparaisons, préconisations et règles de bonne conduite.

Il ressort également de plusieurs comptes-rendus relatifs à l'évaluation scientifique pratiquée dans des pays anglo-saxons (colloque « Enseignement supérieur et recherche : des évaluations à la décision », 2007) que le *quality assessment* y repose le plus souvent sur plusieurs caractéristiques procédurales, que l'Union européenne cherche d'ailleurs à diffuser via une évaluation européenne des évaluateurs nationaux... On y retrouve en effet une *mise à plat* très détaillée du déroulement des évaluations et des procédures de certification, une pondération des critères de publication négociée selon la diversité des disciplines, et une publication systématique des rapports d'évaluation. Cette approche de l'évaluation se veut périodique, explicite et dialoguée entre évaluateurs internes et externes. Pour essayer d'échapper au formalisme, elle fait aussi le pari de l'évolutivité et veut privilégier les apprentissages collectifs.

Au niveau français, on en perçoit déjà l'écho direct. Car les nouveaux protocoles d'évaluation scientifique, les guides de référence pour la constitution des jurys, mais aussi les déroulés des comités de visite des unités de recherche sont officiellement présentés comme évolutifs. L'objectif affiché est de les apprendre à travers des allers et retours entre normes générales et expériences de terrain. Apprentissages internes, au sein du groupe : les différents experts disciplinaires sont ainsi amenés à confronter leurs modes de faire et à rechercher des ajustements possibles entre cultures disciplinaires. Apprentissages externes : les démarches d'évaluation du nouveau dispositif commencent d'être discutées en parallèle avec des membres des diverses communautés scientifiques et comparées à d'autres modes de faire développés en France et à l'étranger.

Dans cet article, l'accent a été mis sur un seul point, la définition du *publiant* dans la caractérisation de l'activité scientifique, afin de pouvoir en approfondir l'analyse. Soulignons toutefois que ce n'est là qu'une seule des multiples procédures actuellement mises en place par le nouveau dispositif d'évaluation, et pas forcément la plus détaillée. Ainsi, la note de méthode actuelle sur les critères d'évaluation dans ce dispositif ne comporte pas moins de huit rubriques thématiques différentes, qui se décomposent elles-mêmes en vingt et un critères principaux. Quant au protocole de constitution des *comités de visite d'évaluation* à former pour chacun des sites scientifiques à examiner, il constitue une sorte de « jeu de l'oie ». C'est un parcours processuel qui, pour chaque comité, passe par douze étapes différentes de contact, de négociation et de transmission d'informations, à suivre successivement. Le rapport d'évaluation, de son côté, rédigé par un président de jury différent à chaque fois, répond également à un cahier de charge méthodologique et rédactionnel très disert.

Ce filet procédural fort serré peut être perçu comme une démarche de standardisation et de contrôle tatillon du travail d'expertise, le dépossédant largement de son initiative et de sa marge de manœuvre. Ce risque n'est pas absent, loin de là, mais il ne correspond sans doute pas véritablement à la stratégie poursuivie. Car il s'agit plutôt d'un pari procédural, où le respect du processus suivi est jugé essentiel car il sera garant de la qualité de la démarche et de son résultat. « Feuille de route » : on sait combien cette formule est devenue un cliché contemporain. Son usage journalistique et ses connotations très militaires ne doivent pas faire oublier sa signification profonde, de type procédural. C'est le cheminement qui doit être modélisé, car c'est lui qui importe en définitive. Ce paradigme de l'*implementation* imprègne les approches pragmatistes et explique l'attention apportée aux *rutines*, comprises non comme des répétitions paresseuses mais comme les apprentissages cognitifs nécessaires aux organisations. Pareille conception des choses sous-tend fortement les évolutions actuelles de l'action publique et de son évaluation.

## Conclusion

*Déroulés* ou *feuilles de route* de type procédural, critères ajustés à travers des allers et retours, publicisation systématique des conclusions. Toutes ces propositions participent du développement de nouvelles « sciences de l'action publique » (Ihl, 2006), c'est à dire de savoirs pour l'action définis sous la double influence de l'europanisation des politiques publiques et du décentrement de l'Etat. Dans cette perspective, le dispositif français d'évaluation scientifique ici envisagé expérimente de nouvelles coordinations entre expertise savante et procédures codifiées. La transition avec les modèles classiques de l'évaluation ne peut être cependant que progressive et assez lente. En effet, quoi qu'on en dise, la culture administrative française reste encore marquée par les codes de la règle générale et par une normativité plutôt descendante. Par ailleurs, l'évaluation scientifique en SHS, n'a pas une pratique générale des indicateurs quantitatifs. Leur maniement et leur combinaison avec des appréciations qualitatives sont encore à expérimenter et à mieux étalonner. On comprend que cette expérimentation requière encore du temps et des apprentissages, au sein des collectifs d'évaluateurs comme dans les milieux académiques de référence. Car les réticences et les incertitudes actuelles peuvent être comprises comme les premières étapes d'un apprentissage organisationnel au sein de l'agence. Et de ce point de vue, même au risque d'impatiser les experts, la tendance est probablement à multiplier de plus en plus les guides méthodologiques d'évaluation, les carnets de visite, les feuilles de route. Il est trop tôt pour dire si ce pari procédural va rapidement se traduire en France dans des pratiques et des croyances légitimes. Mais on peut en tout état de cause en esquisser la double portée potentielle. Avec ces protocoles évaluatifs explicites l'arbitraire individuel ou la bataille d'écoles scientifiques s'avèreront plus difficiles à garder souterrains, sinon à pratiquer. En contrepartie, la mise en scène poussée de comités de visite et de déroulés d'évaluation pousse à une semi professionnalisation et conduit à des risques renouvelés de bureaucratisation des démarches et des contrôles.

Pour l'essentiel, le dispositif d'évaluation scientifique ici considéré rend plus généralement compte d'une évolution qui participe d'une mise en question de la *République des savants*. Une transition semble notamment s'esquisser entre le modèle de « l'expertise substantielle », tâche encore ponctuelle confiée à des pairs, et une évaluation plus « procédurale » et continue, susceptible d'une certaine professionnalisation. Pour s'affirmer, cette évolution suppose cependant l'émergence de nouvelles formes de légitimation du jugement en valeur. En effet, l'expertise substantielle fondait son autorité sur les positions acquises par les pairs dans la communauté scientifique et/ou sur une élection référée à des valeurs et des programmes collectifs. De son côté, l'évaluation procédurale va rechercher des éléments de conviction et des garanties de qualité du côté du respect des processus explicites et des formes de publicisation. Elle fait le pari que son sérieux se mesurera à l'affichage de critères, de classements, de protocoles. Et que sa crédibilité reposera sur l'ouverture contradictoire des démarches. La forme importe-t-elle dès lors plus que le fond ? Plus exactement, c'est le postulat que le fond dépendra beaucoup d'une forme qui saura rendre explicite et mettre en scène des pesées multiples.

## 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
- 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
- Laredo, Ph, Mustar, Ph, Research and Innovation policies in a nex global Economy, Elgar publishing, 2001
- 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
- Perret, B, L'évaluation des politiques publiques, La Découverte, 2001
- Verdier, E, Lanciano-Morandat, C, Les politiques nationales de RD : pour une analyse comparée en termes de conventions d'action publique, in Cahiers de recherche sociologique, n°40, 2005